UNIVERSIDAD MAYOR DE SAN SIMÓN

FACULTAD DE CIENCIAS Y TECNOLOGÍA

CARRERA DE INGENIERÍA CIVIL

MATERIAL DE APOYO DIDÁCTICO DE LA ENSEÑANZA Y APRENDIZAJE DE LA ASIGNATURA DE "PROBABILIDAD Y ESTADÍSTICA (CIV – 271)" (TEXTO ALUMNO)

Trabajo Dirigido, Por Adscripción, Presentado Para Optar al Diploma Académico de Licenciatura en Ingeniería Civil

Presentado Por: PAMELA CARRASCO COVARRUBIAS

MARIO ZUÑIGA ZEBALLOS

Tutor: Ing. M.Sc. Wilson Trujillo Aranibar

COCHABAMBA – BOLIVIA

AGOSTO, 2007

DEDICATORIA

A nuestras familias por apoyarnos siempre en los momentos más difíciles.

FICHA RESUMEN

El presente documento ha sido elaborado con el propósito de ofrecer al estudiante un apoyo didáctico de la enseñanza y aprendizaje de la asignatura de Probabilidad y Estadística, perteneciente a la carrera de Ingeniería Civil de la Facultad de Ciencias y Tecnología.

El mencionado documento tiene un contenido que abarca desde los aspectos más básicos de la Estadística descriptiva, en su función de resumir, presentar y comunicar los resultados de cualquier estudio, abarcando un estudio Unidimensional y Bidimensional además de la teoría de regresión y correlación, hasta el estudio de la Probabilidad estudiando lo referido a Variables Aleatorias, Distribuciones de Probabilidad y Estimación Estadística.

La variabilidad que ha generado los nuevos planes de estudio no facilita la selección de unos contenidos que abarque la totalidad de los programas de todas las universidades, sin embargo hay una parte troncal que constituye un porcentaje amplio del conjunto de todos ellos. Esta es la parte que hemos seleccionado para nuestro contenido, de manera que podamos acercarnos lo máximo posible a lo que pudiera ser un libro de texto para la asignatura de Probabilidad y Estadistica que se imparten en la mayoría de las universidades.

El desarrollo de la relación enseñanza-aprendizaje nos condiciona a que teoría y práctica avancen simultáneamente complementándose la una con la otra y apoyándose mutuamente en ejemplos que puedan facilitar al estudiante en su comprensión acercándolo a situaciones más cotidianas de su entorno.

Esperamos que el empeño puesto en este texto permita asimilar de mejor manera la enseñanza de la Probabilidad y Estadística en nuestra carrera, convirtiéndose en un elemento eficaz de ayuda, apoyo y consulta entre los estudiantes.

ÍNDICE

Pág.

CAPÍTULO I

CONCEPTOS Y DEFINICIONES GENERALES

CAPITULO II

ESTADISTICA DESCRIPTIVA UNIDIMENSIONAL

CAPITULO III

ESTADÍSTICA DESCRIPTIVA BIDIMENSIONAL

CAPITULO IV

TEORIA DE LA PROBABILIDAD

CAPÍTULO V

VARIABLE ALEATORIA

CAPÍTULO VI

FUNCIONES DE DISTRIBUCION DE VARIABLES ALEATORIAS

CAPÍTULO VII

TEORIA DE LA ESTIMACIÓN ESTADISTICA

ÍNDICE DE FIGURAS

Pág.

CAPÍTULO I

CONCEPTOS Y DEFINICIONES GENERALES

CAPITULO II

ESTADISTICA DESCRIPTIVA UNIDIMENSIONAL

CAPITULO III

ESTADÍSTICA DESCRIPTIVA BIDIMENSIONAL

CAPITULO IV

TEORÍA DE LA PROBABILIDAD

CAPÍTULO V

VARIABLE ALEATORIA

CAPÍTULO VI

FUNCIONES DE DISTRIBUCION DE VARIABLES ALEATORIAS

CAPÍTULO VII

TEORÍA DE LA ESTIMACIÓN ESTADISTICA

ÍNDICE DE TABLAS

Pág.

CAPÍTULO I

CONCEPTOS Y DEFINICIONES GENERALES

CAPITULO II

ESTADISTICA DESCRIPTIVA UNIDIMENSIONAL

CAPITULO III

ESTADÍSTICA DESCRIPTIVA BIDIMENSIONAL

CAPITULO IV

TEORÍA DE LA PROBABILIDAD

CAPÍTULO V

VARIABLE ALEATORIA

CAPÍTULO VI

FUNCIONES DE DISTRIBUCION DE VARIABLES ALEATORIAS

CAPÍTULO VII

TEORÍA DE LA ESTIMACIÓN ESTADISTICA

CAPÍTULO I

CONCEPTOS Y DEFINICIONES GENERALES

CAPÍTULO I

CONCEPTOS Y DEFINICIONES GENERALES

1.1 INTRODUCCIÓN

En el presente capítulo se aborda la definición de algunos conceptos básicos, y sin embargo pilares para la comprensión intuitiva y real de lo que es la Estadística. De esta forma y para facilitar el estudio se busca establecer una terminología común mediante una serie de convenciones, para hacer más ágil la lectura del texto. El problema consiste en la diversidad de términos usados para nombrar los mismos conceptos, debido a la gran variedad de autores, modas y tendencias, que existen en la bibliografía actual. Cada uno, aportando su cuota de originalidad, pero complicando la simplicidad. Otro problema que se trata de resolver, es el uso de palabras con un significado claro en el leguaje diario, pero con uno diferente en la Estadística.

1.2 CONCEPTOS

La palabra "Estadística", derivada del latín *status*, que significa estado, posición o situación, como muchas palabras tienen varios significados. En el lenguaje común, el término "estadística" hace referencia a una relación de datos numéricos presentada de forma ordenada y sistemática. Esta idea es la consecuencia del concepto popular que existe sobre el término y que cada vez está más extendido debido a la influencia del entorno, ya que hoy en día es casi imposible que cualquier medio de difusión, periódico, radio, televisión, etc., no aborde diariamente cualquier tipo de información estadística sobre accidentes de tráfico, índices de crecimiento de población, turismo, tendencias políticas, etc..

Sólo cuando se adentra en un mundo más específico como es el campo de la investigación de Ciencias como la Ingeniería, Medicina, Biología, Psicología, etc. Se empieza a percibir que la Estadística no sólo es algo más, sino que se convierte en la única herramienta que, hoy por hoy, permite dar luz y obtener resultados, y por tanto beneficios, en cualquier tipo de estudio, cuyos movimientos y relaciones, por su variabilidad intrínseca, no puedan ser abordados desde la perspectiva de las leyes deterministas. Se puede, desde un punto de vista más amplio, definir la Estadística como un conjunto de métodos científicos para la recopilación, representación condensación y análisis de los datos extraídos de un sistema en estudio, con el objeto de hacer estimaciones y sacar conclusiones, necesarias para tomar decisiones

1.3 DEFINICIONES

La Estadística no es una ciencia en sí misma. Se trata de un grupo de métodos con base científica. Los *métodos* son modelos que optimizan matemáticamente los objetivos buscados. De hecho, la Estadística es una rama de las Matemáticas.

Recopilar datos significa obtenerlos efectuando mediciones, muestreos, encuestas, censos, etc.. La *representación* de datos implica mostrarlos en gráficos, en tablas, en forma de texto, o cualquier combinación de éstas. La *condensación* de los datos implica reducir su número a dos o tres valores representativos de todo el grupo, denominados *estadísticos, estadígrafos o números índices,* tales como la media, la mediana, la varianza, costo de vida, etc.. El *análisis* se hace con las herramientas estadísticas, empleando la información obtenida de los datos, para realizar *estimaciones* o inferencias, comprobar hipótesis de trabajo y así tomar las *decisiones* más adecuadas en cada caso particular, basadas en la evidencia científica suministrada por estos análisis.

De la definición anterior, surge que la Estadística puede ser usada en cualquier *sistema* en estudio. En la práctica, esto significa una gran cantidad de posibilidades, pues, donde pueda definirse un sistema, allí podrá emplearse la Estadística.

Definición 1.1 ***Sistema*** *es un conjunto de elementos que se aíslan para su estudio en función de sus interrelaciones*.

Los *sistemas* tienen tres características que permiten identificarlos de entre todos los conjuntos:

- *Objetivos*: Saber lo que se quiere del sistema, o lo que se espera que éste haga.
- *Comportamiento:* Lo que realmente hace el sistema, cómo se comporta.
- *Estructura:* La manera en que se interrelacionan los elementos.

Es importante destacar que el sistema completo bajo estudio estadístico se constituye en la *población* o *marco de referencia.* Toda la información obtenida adquiere sentido si se sabe respecto de qué o de quiénes, es válida. Por eso en el estudio de la Estadística se necesita de un sistema referencial: la *población.*

Definición 1.2 ***Población*** *es el conjunto o grupo de individuos o sucesos, con al menos una característica común, en la cual todos los miembros están considerados* .
Atendiendo a su tamaño, la población puede clasificarse como:

- *Finita: S*í los elementos de ésta son susceptibles de ser contados.
  Ejemplo: “la estatura de todos los alumnos de la Facultad de Ciencias y Tecnología de la UMSS”.

- *Infinita*: Cuando no es posible realizar un conteo.
  Ejemplo: “todos los posibles sucesos (caras o cruces) en el lanzamiento de una moneda en forma sucesiva”

Definición 1.3 ***Muestra*** *es un conjunto de datos obtenidos de una población cualquiera, con el método de recopilación elegido. Se la puede imaginar como un subconjunto del conjunto población.*

Se toman muestras, cuando no se puede o no conviene, tomar la población entera. Si se tiene una población de tamaño infinito, no se podrá nunca tomar todas las muestras posibles, como por ejemplo, las mediciones repetidas de una misma magnitud, que se pueden repetir indefinidamente mientras el ensayo no sea destructivo. Hay ocasiones, en las cuales si bien la población es finita, es tan grande que no resulta práctico tomar todos los casos. Pueden haber razones de tiempo que impidan analizar a toda la población. Si el método de ensayo es destructivo, entonces no hay más remedio que tomar muestras. La idea básica es tomar muestras representativas de la población desconocida, y a través del análisis de las mismas hacer deducciones acerca de esa población.

Tradicionalmente la Estadística se divide en **Estadística descriptiva** y **Estadística inferencial.**

Definición 1.4 ***Estadística descriptiva*** *es la parte de la Estadística que se ocupa de recopilar, representar y condensar los datos obtenidos del sistema en estudio. Encierra cualquier tratamiento de datos numéricos que comprenda generalizaciones, agrupa todas aquellas técnicas asociadas con el tratamiento o procesamiento de conjuntos de datos, su objetivo comprende la caracterización de conjuntos de datos numéricos, la misma pretende poner de manifiesto las propiedades de estos conjuntos lo cual se puede lograr de forma gráfica o analítica.*

Definición 1.5 ***Estadística Inferencial*** *es la parte de la Estadística dedicada a la formulación de supuestos y estimaciones, para hacer predicciones y poder sacar conclusiones de los datos obtenidos con el estudio de las muestras, y así, tomar decisiones con base científica.*

Para el manejo de los conceptos que se manejaran en los capítulos posteriores es preciso conocer previamente ciertos términos que se utilizaran a continuación.

Definición 1.6 ***Variable*** *es una característica de una población que se va a investigar y que puede asumir diferentes categorías.*

Definición 1.7 **C*arácter*** *es toda propiedad, rasgo o cualidad, de un elemento integrante de la población, susceptible de ser medida u observada.*

Definición 1.8 ***Categorías*** *son las diferentes situaciones posibles de un carácter. Las categorías deben ser a la vez exhaustivas y mutuamente excluyentes (cada elemento posee una y sólo una de las categorías posibles).*

Definición 1.9 ***Medir*** *es comparar un carácter con otro de su misma especie, considerado como referencia o patrón. Ya sea usando instrumentos, o bien, por medio de los sentidos.*

Definición 1.10 ***Parámetro*** *es todo carácter que tiene la misma categoría dentro de una población. O sea, no permite diferenciar entre sí a sus elementos componentes.*

De las definiciones anteriores surge que una *muestra* puede ser interpretada como un conjunto de categorías que pueden adoptar un carácter. Y cuando tal conjunto, está formado por todas las categorías posibles del carácter, entonces se trata de la *población.* Mientras que si se tiene una única categoría constante para toda la población, se puede tomar como un *parámetro* o como una muestra de un único elemento.

Los caracteres pueden ser clasificados como se indica en la Figura 1.1:

- Carácter
  - Cuantitativo
    - Continuo
    - Discreto
  - Cualitativo
    - Ordinales
    - Nominales
      - Dicotómicas (binarias)
      - Politónicas

Figura 1.1 Tipos de caracteres.

La diferencia entre caracteres cuantitativos y cualitativos está en la relación que tienen con el patrón o estándar, contra el cual son comparadas al ser medidas. Si esta relación puede ser expresada con números, debido a una proporcionalidad, entonces el carácter es *cuantitativo*. En general, todos los caracteres físicos basadas en el Sistema Internacional de Unidades (metro, kilo, segundo, etc.) son ejemplos clásicos de caracteres de este tipo. En cambio, un carácter es *cualitativo* cuando su relación con el patrón no es una proporción numérica.

Los *caracteres cuantitativos* se clasifican en continuos y discretos. Se diferencian entre sí, porque en la primera la relación numérica con el patrón puede ser cualquiera, mientras que en la segunda hay algunos valores prohibidos. Entonces un carácter *continuo* se expresa mediante números reales. En un carácter continuo hay infinitos puntos posibles dentro de un intervalo cualquiera de la misma, en el dominio de los números reales. En cambio, cuando el carácter tiene algunas categorías que son posibles y otros que no, entonces se trata de una *discreta* (se expresan por lo general con los números enteros positivos). Por ejemplo, las de recuento o enumeración.

Los *caracteres cualitativos* se clasifican en nominales y ordinales. Se diferencian entre sí en que los nominales son las cualidades del objeto de la medición, observables sin emplear instrumentos. En cambio, las *ordinales* implican medir el orden de los resultados obtenidos, para luego clasificarlos. Debe destacarse que un carácter de tipo continuo, puede usarse como carácter ordinal y aún nominal. Depende de la convención utilizada para expresar los resultados. Al precisar la talla de una persona, se está midiendo un carácter de tipo continua, pero si esos datos se usan para clasificarlo en muy alto, alto, normal, bajo o enano, entonces se "ordinaliza" el resultado, expresándolo en forma cualitativa. Es conveniente tener en cuenta, desde el punto de vista de la cantidad de información, que la riqueza contenida en un carácter continuo se va perdiendo al volverlo cualitativo, mediante algún tipo de convención.

Las variables pueden subdividirse de la misma manera en diferentes tipos, de acuerdo a su asociación con alguna de estos caracteres, de ésta forma si una variable está asociada a un carácter cualitativo será una *variable cualitativa* o si está asociada a un carácter cuantitativo será una *variable cuantitativa.*

Las variables pueden asumir datos *Uni-variados*, se obtienen cuando se mide una sola variable. Los datos *Bi-variados* se obtienen cuando se miden dos variables en una sola unidad, y los *Multi-variados* cuando se miden más de dos variables. Las variables además son susceptibles a ser analizadas por herramientas creadas especialmente según su naturaleza.

1.4 METODOLOGÍA DE LA INVESTIGACIÓN EN ESTADÍSTICA (PROTOCOLO DE OBSERVACIÓN)

El protocolo de observación, constituye la herramienta de planificación de procesos de generación de datos en los cuales el investigador se limita simplemente a observar, sin modificar o controlar el objeto de estudio. Diversas son las disciplinas de la ciencia que recurren a este sistema de generación de datos.

1.4.1 Definición de resultados esperados

Considerando que un trabajo de investigación esta orientado en buscar "conocimiento" a propósito de un objeto es de importancia fundamental poder identificar de manera absoluta precisa la *necesidad de información*

En esta fase de la planificación, el investigador, debe cerrar los ojos y soñar que su trabajo ha sido concluido. ¿ Que es lo que se tiene como aporte?, ¿Qué conocimientos ha generado la investigación?. El resultado de estos sueños en muchos casos puede provocar que el investigador abandone el tema de su investigaron. Sin embargo si el sueño ha sido "conciente" el resultado es un listado de los "resultados esperados de la investigación".

La responsabilidad de este punto del protocolo es única y exclusivamente del especialista, el estadístico, en función de su experiencia, podría eventualmente colaborar, sin embargo corresponde al investigador tener muy clara la idea.

Al finalizar esta etapa el investigador dispone de la "primera columna" de la matriz de planificación que será presentada mas adelante.

RESULTADOS ESPERADOS
R_1
R_2
. .
R_i
. .
R_n

Tabla 1.1 Matriz de planificación (resultados esperados)

Se debe notar que el investigador, en función del tema de investigación puede plantear *n* resultados esperados, siendo R_i, el i-esimo.

1.4.2 Definición de la población y las unidades de observación

Una vez identificados los resultados esperados, el paso siguiente consiste en definir cual es el objeto de la investigación en términos de la población de estudio y las unidades de observación.

Las unidades de observación constituyen precisamente los "individuos" que serán el objeto de la medición.

El termino "individuos" es utilizado en su contexto mas genérico : personas, familias, localidades, empresas, etc..

1.4.3 Definición de variables

En este paso del protocolo se plantean dos problemas: ¿Cuál es el numero de variables que se debe observar?, ¿Cuál es la naturaleza del dato generado?.

Para resolver el problema del número de variables a observar, el procedimiento es relativamente sencillo, la idea es llegar a establecer el set mínimo de variables en base a los resultados esperados. Se recomienda usar la siguiente matriz.

RESULTADOS ESPERADOS	VARIABLES
R_1	X_1: Variable 1 X_2: Variable 2
R_2	X_3: Variable 3 X_1: Variable 1 X_4: Variable 4
R_3	X_5: variable 5

Tabla 1.2 Matriz de planificación (resultados esperados, variables)

En la tabla anterior se puede observar que el investigador persigue tres resultados esperados (R_1, R_2, R_3). Para alcanzar el primer resultados esperado se requiere evaluar dos variables X_1 ya participa en R_1, finalmente el R_3 requiere la observación de la variable X_5. Para responder a los tres resultados esperados, es necesario observar cinco variables, La consideración de otras variables no es necesaria y si por el contrario se observan menos variables, el riesgo es no alcanzar uno o más de los resultados esperados.

La naturaleza de los datos es el segundo problema relativo al paso 3 del protocolo. La definición del tipo de dato es fundamental en el paso 5 en el que se definen las herramientas de análisis de datos. Se recomienda en la tabla anterior especificar para cada una de las variables la naturaleza del dato generado cuando se evalúa la variable.

1.4.4 Definición del plan de muestreo

El plan de muestreo se refiere a aquellas operaciones orientadas a seleccionar individuos de la población, que efectivamente serán observados, constituyendo así la muestra.

El principio o regla de “oro” en el problema de la selección de la muestra es conformar un grupo de “individuos” representativos. Es importante disponer de una buena fotografía de la población de manera que el proceso de inferencia (generalización de los resultados de la muestra a la población) sea adecuado.

De manera general los planes de muestreo se clasifican en dos categorías: muestras probabilísticas y no probabilísticas. Los muestreos probabilísticos, responsabilizan al azar el proceso de selección de las unidades de la población para conformar la muestra, por el contrario los muestreos no probabilísticos, responsabilizan el proceso de selección al investigador. Es importante notar que en el caso de los muestreos no probabilisticos, supone que el investigador tenga un profundo conocimiento de la población, caso contrario el proceso de selección no garantiza la representatividad de la muestra.

Los métodos de muestreos normalmente usados son:

Muestreo No probabilistico:

- *Muestreo por conveniencia*, con base en la conveniencia.
- *Muestreo por juicio*, con base en lo que opina un experto.
- *Muestreo bola de nieve*, cuando no existe un registro de esta muestra.

Muestreo probabilistico:

- *Muestro aleatorio simple*, todos tienen las mismas probabilidades dentro de la población.
- *Muestreo estratificado*, existen grupos naturales ya definidos dentro la población.
- *Muestreos por conglomerados*, pueden ser sistemáticos o por áreas.

1.4.5 Determinación del tamaño de la muestra.

De manera general, los resultados esperados de una investigación pueden tener dos propósitos:

- La estimación

- Probar una hipótesis planteada.

En el caso de la *estimación*, se requiere conocer el tamaño de la muestra, de ésta manera lograr un estimador con una precisión deseada o admisible (d_r) a un nivel de confianza o seguridad también definido por el investigador (α). Finalmente participa un tercer elemento que no puede ser fijado si no estimado por el investigador, se trata de la variabilidad asociada la variable que estamos observando (σ). Por tanto en el caso de la estimación.

$$n = f(1-\alpha, d_r, \sigma)$$

En el caso de *probar un test de hipótesis* la noción de precisión ya no tiene sentido, se habla de la potencia del test. De manera simple la potencia del test se define como la capacidad de la herramienta para decir la verdad. En efecto el resultado de una prueba de hipótesis es aceptar o rechazarla. Si la herramienta acepta la hipótesis cuando en la realidad debiera rechazarse se comete un error de tipo II, y si por el contrario la herramienta rechaza una hipótesis cuando en la realidad debiera aceptarse, se comete un error de tipo I. La determinación del tamaño de la muestra, en el caso de un test de hipótesis debe garantizar con un cierto nivel de confianza, que la herramienta no se equivoque.

Para poder determinar el tamaño de la muestra es necesario trabajar por resultado esperado. Previamente a la determinación del tamaño de la muestra por resultado esperado se debe realizar la selección de la herramienta en el paso siguiente.

1.4.6 Análisis de datos

El análisis de datos constituye la etapa final del protocolo de observación, de manera sencilla la intención en esta fase es lograr un aprovechamiento optimo de los datos colectados con el propósito de alcanzar los resultados esperados.

De manera general el problema del análisis de datos plantea la dificultad de seleccionar la herramienta adecuada. La selección de la herramienta de análisis es función de cuatro criterios a considerar: a) los resultados esperados, b) el numero de variables que participan en el resultado esperado, c) la naturaleza del dato y d) las condiciones de aplicación del método.

El proceso de análisis de datos implica una serie de etapas no necesariamente secuenciales:

a) *Análisis exploratorio de datos – EDA*. El propósito de esta fase es controlar la calidad de los datos. Quiere decir identificar aquellos datos influyentes o aberrantes. Es importante ser demasiado cuidadoso con el tratamiento de los datos, en ningún caso se debe proceder de manera sistemática a su eliminación. Un dato influyente debe ser objeto de revisión responsable.

b) *Análisis descriptivo de los datos*. El propósito es resumir los datos. Las herramientas de resumen de datos son diversas: tablas, graficas y parámetros o estadígrafos.

c) *Análisis inferencial de los datos*. El propósito es generalizar los resultados de la muestra a nivel poblacional. Existen dos grandes áreas: el problema de estibación en el cual se busca estimador a nivel poblacional y el problema del test de hipótesis en el que se pretende captar o rechazar una hipótesis planteada en el estudio.

En cada una de estas fases del análisis de datos es importante considerar el número de variables que se consideran en el resultado esperado. Si el resultado esperado hace participar una sola variable se buscan herramientas de control de calidad o herramientas de control de calidad de datos, herramientas de resumen de datos o herramientas de inferencia bidimensionales. Finalmente si el resultado esperado hace participar a mas de dos variables, las herramientas tanto de EDA, descriptivas e inferenciales deben ser multivariables.

Finalmente el protocolo de Observación se resume en la construcción de la matriz siguiente. Es importante notar que el investigador no debe iniciar ningún proceso de investigación si no ha definido de manera detallada la siguiente matriz.

RESULTADOS ESPERADOS	VARIABLES	METODO ESTADISTICO (Descriptivo – Inferencial)	TAMAÑO DE LA MUESTA (Método de Muestreo)
R_1	X_1: Variable 1 X_2: Variable 2	Método 1	Formula 1
R_2	X_3: Variable 3 X_1: Variable 1 X_4: Variable 4	Método 2	Formula 2
R_3	X_5: variable 5	Método 3	Formula 3

Tabla 1.3 Matriz de planificación

En el caso de la determinación del tamaño de la muestra es necesario ser particularmente atento al hecho siguiente: si el resultado es descriptivo no existen formulas, por el contrario si el resultado es de carácter inferencial existe una formula o en algunos casos procedimientos mas complejos de definición del tamaño de la muestra. Finalmente es importante notar que el tamaño de la muestra corresponde al valor más alto determinado por el resultado esperado.

1.4.7 Presentación de datos

Los métodos para la presentación de datos en Estadística son de tres tipos. Casi siempre se los combina con el objeto de lograr una mayor claridad y transparencia en la información que se transmite al lector. Sería muy engorroso leer un informe donde se detallan uno por uno los cientos de valores obtenidos en la recopilación. Por eso, toda esta *nube de datos*, como se la llama en la jerga estadística, debe ser presentada de manera simple y en lo posible amena.

Métodos de representación de datos
- Textual
- Tabular
- Gráfico

Figura 1.2 Tipos de representación de datos.

Cualquiera sea el método de presentación adoptado, es obligatorio hacer mención específica de la fuente desde donde se obtuvieron los datos, de manera tal, que el lector pueda hallarla, en caso de ser necesario.

1.4.7.1 Método textual

Este método de presentación de la información consiste en el empleo de palabras y cifras combinadas en un texto, para informar los datos obtenidos.

Desde el punto de vista del lector, este método es el más engorroso de los tres para poder entender el significado de los valores obtenidos y obtener alguna conclusión.

Desde el punto de vista del redactor, el método textual tiene una ventaja importante con respecto a los otros: se puede influenciar al lector. El autor puede resaltar ciertas cifras de su interés, puede remarcar conceptos apropiados para sus fines y hacer pasar desapercibidos a los otros. Se puede focalizar la atención del lector, de tal manera que pase por alto ciertos datos evitando que saque sus propias conclusiones. Todo lector prevenido debe desconfiar cuando encuentra juicios de valores en un trabajo científico. Se ha preparado un pequeño ejemplo ilustrativo de estos conceptos. Para entenderlo mejor, es conveniente seguir los pasos indicados: primero se debe leer el texto de la Parte A. Luego, en un papel, el lector escribe las cifras que recuerde y las conclusiones que saca. Por último, lee la Parte B y la compara con sus propias conclusiones:

Ejemplo 1.1

Paso 1) Leer una sola vez el párrafo identificado como Parte A.

Parte A: "La actual comisión directiva de nuestro club de canotaje, que hoy finaliza su tarea, tiene el gusto de presentar a vuestra consideración los éxitos y logros alcanzados durante su gestión, tales como el incremento de un 400% en la cantidad de remos extras, y todavía un bote tipo doble más. Se *mejoró* la seguridad y vigilancia de los bienes de nuestro club de dos años de vida, en un
200%. Con gran previsión de futuro y preocupación por el equilibrio ecológico, se *incrementó en un*
500% el capital arbóreo. Los *ingresos* por cuota societaria *aumentaron* en un *50%,* acrecentando a su vez el número de socios. Con esta exitosa gestión financiera de este período, se construyó un nuevo vestuario para mayor comodidad de nuestros asociados mejorando substancialmente el aspecto sanitario..."

Paso 2) Escribir en un papel lo que se recuerda de la primera lectura y opinar sobre la gestión realizada por la comisión directiva saliente **Paso 3)** Leer el Parte B, comparar con la anterior comisión directiva para ver si se mantiene la opinión escrita en el Paso 2.

Parte B: "...Del Resumen y Balance de los dos primeros períodos del club se han extraído las si- guientes cifras:

	1ª Gestión	2ª Gestión
- Cantidad de Botes	20	21
- Cantidad de Remos	42	50
- Árboles plantados	5	25
- Número de serenos	1	3
- Cuota mensual del socio	30$	40$
- Número de socios	110	112
- Superficie construida	600 m^2	606 m^2 (los 6 m^2 se destinaron a vestuario)

Puede notarse del ejemplo 2.1, que mientras las primeras autoridades del club pudieron comprar 20 botes, la segunda solo uno, pero al comprar 8 remos de más, incrementaron en un 400% la cantidad de remos sobrantes. Colocaron 20 plantines de árboles en el terreno y con eso dicen que incrementaron un 500% el "capital arbóreo". El alza de un 50% en los ingresos del club no fue aumentando el número de socios, sino subiendo la cuota. Contrataron 2 serenos, cuando antes con uno alcanzaba. Construyeron 6 m^2, cuando los anteriores lograron centuplicar ese valor. Sin embargo, todas estas cosas no surgen del texto del discurso de la comisión saliente, sino de una tabla comparativa que presenta sólo cifras. Esto ilustra la conveniencia de adjuntar a los textos las tablas correspondientes, para hacer más transparentes los conceptos que se puedan verter. Una tabla no puede influir al lector tanto como lo puede hacer un texto.

1.4.7.2 Método tabular

Este método de presentación de la información consiste en presentar los datos por medio de una tabla o cuadro. En esta, se debe colocar al principio un *título* identificatorio que en forma clara, breve y completa, explique el contenido de la tabla. Luego viene el *cuerpo* y al final se debe colocar la *fuente* de donde se tomaron los valores mostrados.

Una tabla simple consta de un cuerpo dividido en columnas, cada una con su encabezado y una aclaración de las unidades que se están usando. Una tabla de doble entrada consta de un cuerpo dividido en filas y columnas, con sus respectivos encabezados.

Una tabla es el método más imparcial para presentar la información, muestra los datos crudos, dejando al lector la tarea de interpretarlos sin hacer ni sugerencias ni comentarios. Al contrario de un gráfico, no se puede entender la idea general de un simple vistazo salvo a trazos gruesos.

Si los datos que se dispone son numerosos, es indispensable clasificarlos en un cuadro o tabla resumen de las observaciones originales, a las que en adelante se llamará ***tabla de distribución de frecuencias*** o simplemente ***tabla de frecuencias***.

1.4.7.3 Método Gráfico

Se ha visto que la tabla estadística resume los datos que disponemos de una población, de forma que ésta se puede analizar de una manera más sistemática y resumida.

Para darse cuenta *de un sólo vistazo* de las características de la población resulta aún más esclarecedor el uso de gráficos y diagramas.

1.4.7.3.1 Gráficos para variables cualitativas

Los gráficos más usuales para representar variables cualitativas son los siguientes:

a) Diagramas de barras

Siguiendo la Figura 1.3, Se representará en el eje de ordenadas las modalidades y en abscisas las frecuencias absolutas o bien, las frecuencias relativas. Si, mediante el gráfico, se intenta comparar varias poblaciones entre sí, existen otras modalidades, como las mostradas en la Figura 1.4. Cuando los tamaños de las dos poblaciones son diferentes, es conveniente utilizar las frecuencias relativas, ya que en otro caso podrían resultar engañosas.

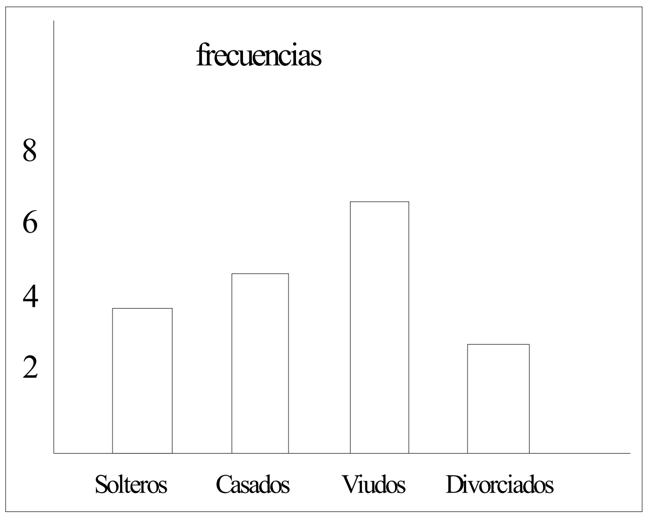


Figura 1.3 Diagrama de barras para una variable cualitativa.

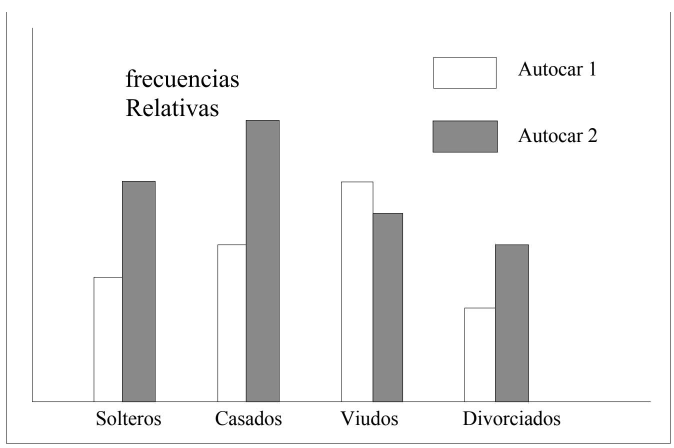


Figura 1.4 Diagramas de barras para comparar una variable cualitativa en diferentes poblaciones. Se ha de tener en cuenta que la altura de cada barra es *proporcional* al número de observaciones (frecuencias relativas).

b) Diagramas de sectores

También llamados *tartas (pie),* se divide un círculo en tantas porciones como clases existan, de modo que a cada clase le corresponde un arco de círculo proporcional a su frecuencia absoluta o relativa (Figura 1.5).

El arco de cada porción se calcula usando la *regla de tres*:

$$\begin{aligned} & n \to 360^{o} \\ & n_i \to x_i = \frac{360 \bullet n_i}{n} \end{aligned} \qquad (1.1)$$

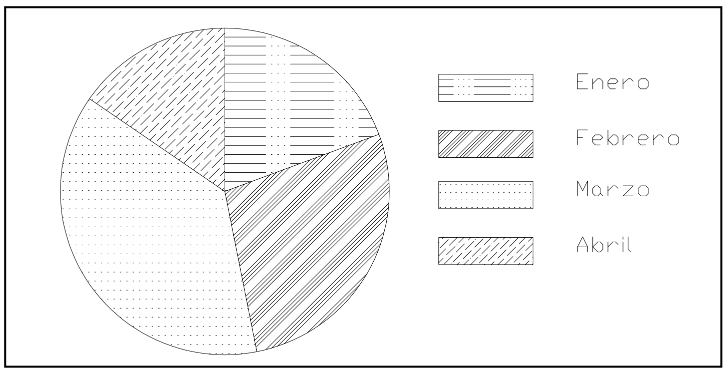


Figura 1.5 Diagrama de sectores que representa precipitaciones en un periodo del año.

Como en la situación anterior, puede interesar comparar dos poblaciones. En este caso también es aconsejable el uso de las frecuencias relativas (porcentajes) de ambas sobre gráficos como los anteriores. Otra posibilidad es comparar las 2 poblaciones usando para cada una de ellas un diagrama semicircular, al igual que en la Figura 1.6. Sean $\mathbf{n_1 \leq n_2}$ los tamaños respectivos de las 2 poblaciones. La población más pequeña se representa con un semicírculo de radio r_1y la mayor con otro de radio r_2. La relación existente entre los radios, es la que se obtiene de suponer que la relación entre las áreas de las circunferencias es igual a la de los tamaños de las poblaciones respectivas, es decir:

$$\frac{r_2^2}{r_1^2} = \frac{n_2}{n_1} \Leftrightarrow r_2 = r_1 \bullet \sqrt{\frac{n_2}{n_1}} \qquad (1.2)$$

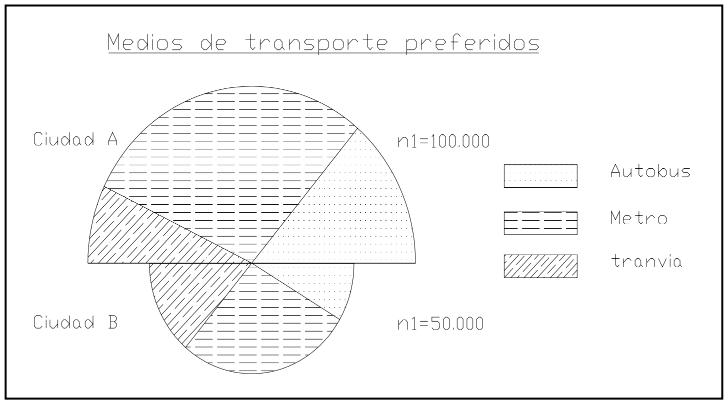


Figura 1.6 Diagrama de sectores para comparar dos poblaciones

c) Pictogramas

Expresan con dibujos alusivos al tema de estudio, las frecuencias de las categorias de la variable. Estos gráficos se hacen, representando a diferentes escalas un mismo dibujo, como en la Figura 1.7.

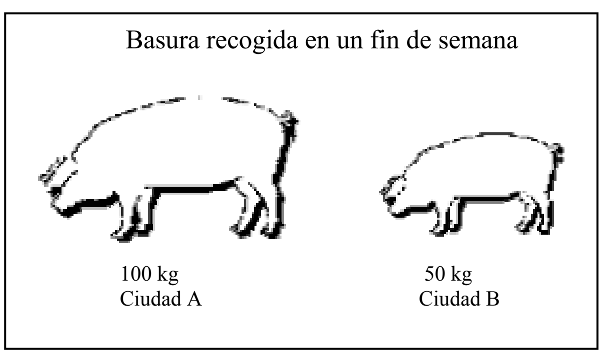


Figura 1.7 Pictograma. Las áreas son proporcionales a las frecuencias.

El escalamiento de los dibujos debe ser tal que el *área* de cada uno de ellos sea proporcional a la frecuencia de la modalidad que representa. Este tipo de gráficos suele usarse en los medios de comunicación, para que sean comprendidos por el público no especializado, sin que sea necesaria una explicación compleja.

1.4.7.3.2 Gráficos para variables cuantitativas

Para las variables cuantitativas, se considerará dos tipos de gráficos, en función de que para realizarlos se usen las frecuencias (absolutas o relativas) o las frecuencias acumuladas:

a) Diagramas diferenciales

Son aquellos en los que se representan frecuencias absolutas o relativas. En ellos se representa el número o porcentaje de elementos que presenta una modalidad dada (ver Figura 1.8).

b) Diagramas integrales

Son aquellos en los que se representan el número de elementos que presentan una modalidad inferior o igual a una dada. Se realizan a partir de las frecuencias acumuladas, lo que da lugar a gráficos crecientes, y es obvio que este tipo de gráficos no tiene sentido para variables cualitativas (ver Figura 1.8).

Según se ha visto existen dos tipos de variables cuantitativas: discretas y continuas.

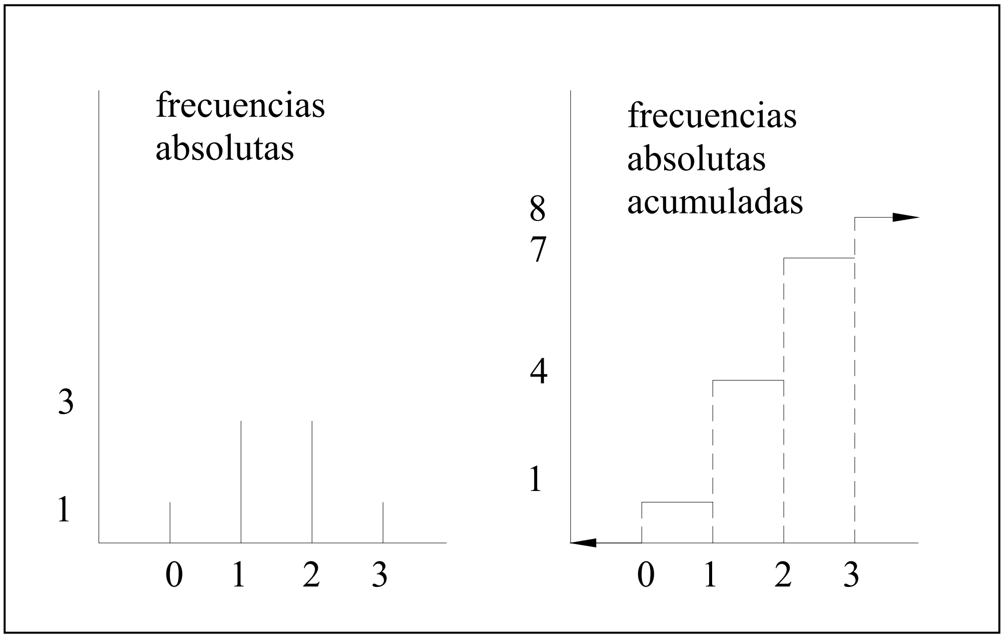


Figura 1.8 Diagrama diferencial (barras) e integral para una variable discreta. Obsérvese que el diagrama integral (creciente) contabiliza el número de observaciones de la variable inferiores o iguales a cada punto del eje de abscisas.

1.4.7.3.2.1 Gráficos para variables discretas

Cuando es representada una variable discreta, se usa el **diagrama de barras** cuando se pretende hacer una gráfica diferencial. Las barras deben ser estrechas para representar el que los valores que toma la variable son discretos. El diagrama integral o acumulado tiene, por la naturaleza de la variable, forma de escalera. Un ejemplo de diagrama de barras así como su diagrama integral correspondiente están representados en la Figura 1.8.

1.4.7.3.2.1 Gráficos para variables continuas

Cuando las variables son continuas, se utiliza como diagramas diferenciales los *histogramas* y los *polígonos de frecuencias*.

Definición 1.11 *Un* **histograma** *es una representación grafica de las distribuciones de frecuencias absolutas o relativas de datos cuantitativos continuos agrupados en clases. Se construye a partir de la tabla estadística, representando sobre cada intervalo, un rectángulo que tiene a este segmento como base. El criterio para calcular la altura de cada rectángulo es el de mantener la proporcionalidad entre las frecuencias absolutas (o relativas) de cada intervalo y el área de los mismos.*

Definición 1.12 *El* **polígono de frecuencias** *se construye fácilmente si se tiene representado previamente el histograma, ya que consiste en unir mediante líneas rectas los puntos del histograma que corresponden a las marcas de clase. Para representar el polígono de frecuencias en el primer y último intervalo, se supone que adyacentes a ellos existen otros intervalos de la misma amplitud y frecuencia nula, y se unen por una línea recta los puntos del histograma que corresponden a sus marcas de clase.*

Definición 1.13 *El diagrama integral para una variable continua se denomina también* ***polígono de frecuencias acumulado****, y se obtiene como la poligonal definida en abscisas a partir de los extremos de los intervalos en los que se ha organizado la tabla de la variable, y en ordenadas por alturas que son proporcionales a las frecuencias acumuladas*.

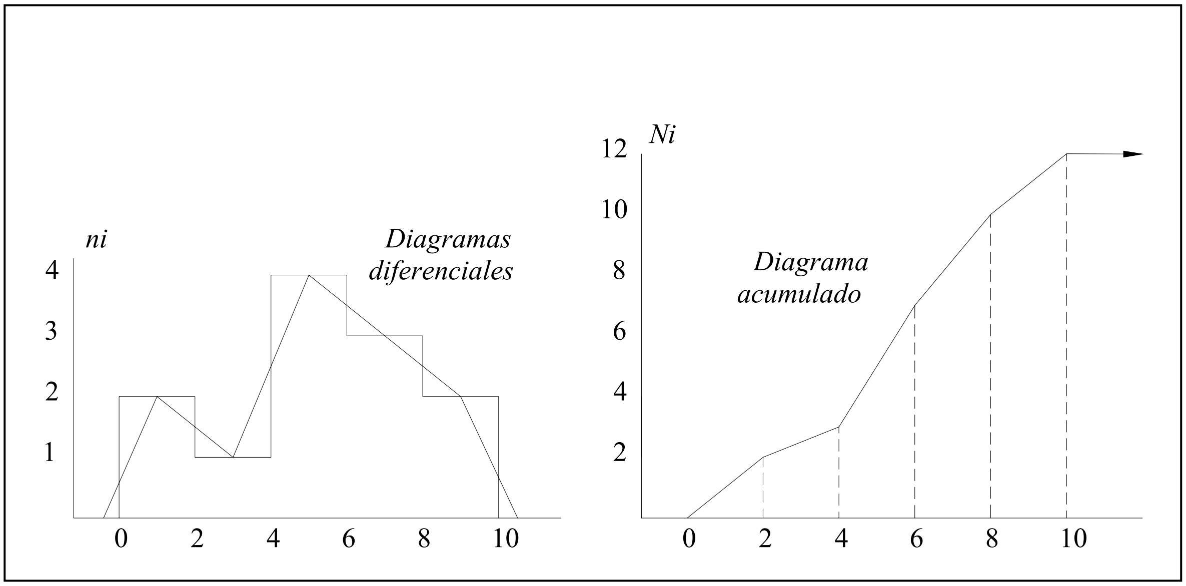


Figura 1.9 Diagramas diferenciales e integrales para una variable continua.

BIBLIOGRAFÍA

Azzimonti Renzo, JC; Apuntes de la Catedra de Estadística, UNAM, 1999/2000

Ríus Días , Francisca, Barón López, Francisco, Sánchez Font, Elisa, Parras Guijosa, Luis, Bioestadística: métodos y aplicaciones, Manual de la Universidad de Málaga.

Villarroel, Luis. Protocolo de observación, UMSS diciembre 2003

CAPÍTULO II

ESTADÍSTICA DESCRIPTIVA UNIDIMENSIONAL

CAPÍTULO II

ESTADÍSTICA DESCRIPTIVA UNIDIMENSIONAL

2.1 INTRODUCCIÓN

Al analizar un conjunto de datos a nivel uni-variado, se debe determinar primero si se tiene una muestra o una población completa y el tipo de variable, esa determinación afectará tanto los métodos que usemos como las conclusiones que saquemos. Utilizamos métodos de la estadística descriptiva para resumir o describir las características importantes de un conjunto conocido de datos, cuando un profesor calcula el promedio de las calificaciones del examen final para el grupo de estadística, ese resultado es un ejemplo de estadística descriptiva para una población. Tal como se ve a continuación:

Existen muchas formas de clasificar los datos, una manera útil, es dividirlo en categorías similares o clases, y luego contar con el número de observaciones que caen en cada categoría, lo que constituye una tabla de frecuencias o una distribución de frecuencias:

Nota de examen	Alumnos
0 – 50	20
51 – 90	40
90 - 100	5

Otra manera consistirá en un estudio de la tendencia central, estudio de la dispersión o fluctuación alrededor de este valor, estudio de la distribución de la serie.

20	23	……..	55	56	53	69	66	…….	90

Sin embargo, si se dice que el resultado es una estimación del promedio de calificaciones finales para todos los grupos de estadística, estamos haciendo inferencia estadística.

2.2 DISTRIBUCIÓN DE FRECUENCIAS

Si los datos que se dispone son numerosos, es indispensable clasificarlos en un cuadro o tabla resumen de las observaciones originales, a las que adelante se llaman *Tabla de distribución de frecuencias* o simplemente *Tabla de frecuencias.*

Definición 2.1 ***Distribución de frecuencias*** *es la tabla formada por las distintas categorías (valores o intervalos) del carácter X y por las frecuencias absolutas (relativas, absolutas acumuladas o relativas acumuladas).*

Considerando una población estadística de n elementos, descrita según un carácter o variable $\boldsymbol{X}$ cuyas categorías han sido agrupadas en un número $\boldsymbol{k}$ de clases, que se denota mediante $\boldsymbol{c_1, c_2,\ldots, c_k}$. Para cada una de las clases $\boldsymbol{c_i}$, $\boldsymbol{i = 1,\ldots,k}$, se introduce las siguientes definiciones:

Definición 2.2 *Se llama* ***clase c*** *a cada uno de los grupos en que se divide el conjunto de datos.*

Definición 2.3 *Se llama* ***frecuencia total*** *al número de datos* ***n****.*

Definición 2.4 *Se llamas* ***frecuencia absoluta*** $\boldsymbol{n_i}$ *de la categoría* C_i *(valor* x_i *o intervalo* I_i*) de la variable X al número de datos que presentan la categoría* C_i *(valor* x_i *o valor del intervalo* I_i*). Si existen* ***k*** *categorías posibles, se verificará:*

$$\sum_{i=1}^{k} n_j = n_1 + n_2 + \ldots + n_k = n \qquad (2.1)$$

Definición 2.5 *Se llama* ***Frecuencia relativa f_i*** *de la categoría C_i (valor x_i o intervalo I_i) de la variable X al cociente, entre las frecuencias absolutas de dicha categoría C_i (valor x_i o intervalo I_i) y el número total de datos u observaciones, es decir:*

$$f_i = \frac{n_i}{n} \tag{2.2}$$

Definición 2.6 *Se denomina* ***frecuencia absoluta acumulada N_i*** *hasta la categoría C_i (valor x_i o intervalo I_i) a la suma.*

$$N_i = n_1 + n_2 + \ldots + n_i = \sum_{j=1}^{i} n_j \tag{2.3}$$

Definición 2.7 *Se llama* ***Frecuencia relativa acumulada F_i*** *de la categoría C_i (valor x_i o intervalo I_i) de la variable X al cociente, entre las frecuencias absolutas acumulada y el número total de datos u observaciones.*

$$F_i = \frac{N_i}{n} = \frac{n_1 + n_2 + \ldots + n_i}{n} = f_1 + \ldots + f_i = \sum_{j=1}^{i} f_j \tag{2.4}$$

Como todas las categorías son exhaustivas e incompatibles ha de ocurrir que

$$\sum_{i=1}^{k} n_i = n_1 + n_2 + \ldots + n_k = n$$

o lo que es lo mismo,

$$\sum_{i=1}^{k} f_i = \sum_{i=1}^{k} \frac{n_i}{n} = \frac{\sum_{i=1}^{k} n_i}{n} = \frac{n}{n} = 1 \tag{2.5}$$

2.3 TABLA DE FRECUENCIAS PARA VARIABLES CUALITATIVAS

Las tablas de frecuencias para variables cualitativas son de dos tipos, para variables cualitativas nominales y para variables cualitativas ordinales.

2.3.1 Variables cualitativas nominales

Para el caso de variables cualitativas ordinales se tiene la siguiente tabla de distribución de frecuencias y su respectiva representación grafica.

2.3.1.1 Tabla de frecuencias

La tabla de distribución de frecuencias para variables cualitativas nominales constaría de tres columnas, la primera representada por las categorías seguida de sus frecuencias absolutas, la tercera columna ocupada por las frecuencias relativas, y por ultimo las frecuencias relativas acumuladas.

Categoría (C_i)	***Frecuencia absoluta (n_i)***	***Frecuencia relativa (f_i)***
C_1	n_1	f_1
C_2	n_2	f_2
. . .	. . .	. . .
C_i	n_i	f_i
. . .	. . .	. . .
C_k	n_k	f_k
Total:	$\sum n_i = n$	$\sum f_i = 1$

Tabla 2.1 Tabla para variables cualitativas nominales.

2.3.1.2 Representación Gráfica

Para la representación grafica de variables cualitativas nominales adoptaremos el gráfico de sectores o pastel.

Diagramas de sectores

También llamados *tartas (Pie),* se divide un círculo en tantas porciones como categorías existan, de modo que a cada categoría le corresponde un arco de círculo proporcional a su frecuencia absoluta o relativa (Figura 2.1).

El arco de cada porción se calcula usando la *regla de tres*:

$$\begin{aligned} &n \rightarrow 360^{o} \\ &n_i \rightarrow x_i = \frac{360 \bullet n_i}{n} \end{aligned} \qquad (2.6)$$

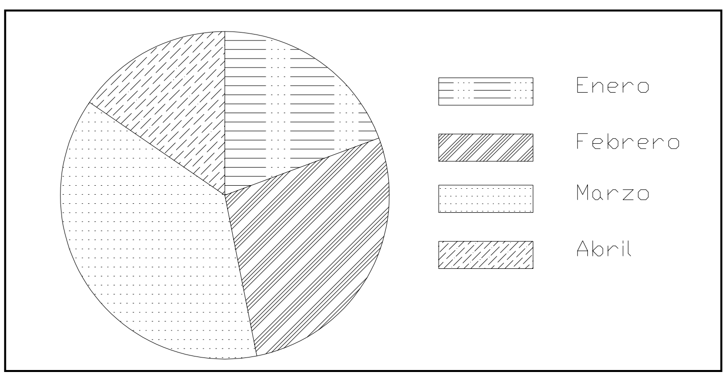


Figura 2.1 Diagrama de sectores que representa precipitaciones en un periodo del año.

2.3.2 Variables cualitativas ordinales

Se tiene a continuación la tabla de distribución de frecuencias y su representación gráfica para variables cualitativas ordinales.

2.3.2.1 Tabla de frecuencias

A la tabla para variables cualitativas nominales se agrega tres columnas, la primera con las Frecuencias absolutas acumuladas y la segunda con las Frecuencias relativas acumuladas, por ultimo la de frecuencias relativas acumuladas porcentuales y se obtiene la tabla de distribución de frecuencias para variables cualitativas ordinales.

C_i	Frecuencia absoluta (n_i)	Frec. Relativa (f_i)	Frec. Acum. absoluta (N_i)	Frec. Acum. Relativa (F_i)
Categoría$_1$	n_1	f_1	N_1	F_1
Categoría $_2$	n_2	f_2	N_2	F_2
. . .	. . .	. . .	. . .	. . .
Categoría $_i$	n_i	f_i	N_i	F_i
. . .	. . .	. . .	. . .	. . .
Categoría $_k$	n_k	f_k	$N_k = n$	$F_k = 1$
Total:	$\sum n_i = n$	$\sum f_i = 1$		

Tabla 2.2 Tabla para variables cualitativas.

2.3.2.2 Representación Grafica

Se adoptara para la representación de variables cualitativas ordinales los diagramas de barras.

Diagramas de barras

Siguiendo la Figura 2.2, Se representará en el eje de ordenadas las categorías y en abscisas las frecuencias absolutas o bien, las frecuencias relativas. Si, mediante el gráfico, se intenta comparar varias poblaciones entre sí, existen otras categorías, se usa los gráficos de barras dobles como las mostradas en la Figura 2.3. Cuando los tamaños de las dos poblaciones son diferentes, es conveniente utilizar las frecuencias relativas, ya que en otro caso podrían resultar engañosas.

Aunque no hay reglas estrictas para la construcción de gráficos de barras, se conveniente seguir las siguientes recomendaciones:

1. Todas las barras, rectángulos o paralelepípedos deben tener el mismo grosor.
2. El espacio entre las barras deben ser de la misma magnitud. No deben ser inferior que la mitad de una barra, ni mayor que el ancho de la misma.
3. Las barras, por estética deben ordenarse de mayor a menor cuando se pueda.
4. La escala de la frecuencia debe empezar por cero.
5. Deben dibujarse a buen criterio, líneas de fondo en la grafica; ellas facilitan la lectura de los valores.
6. No se deben recargar las barras tratando de expresar muchos productos en cada una de ellas.
7. Si el gráfico tiene muchas barras es preferible reemplazarlo por un diagrama lineal.

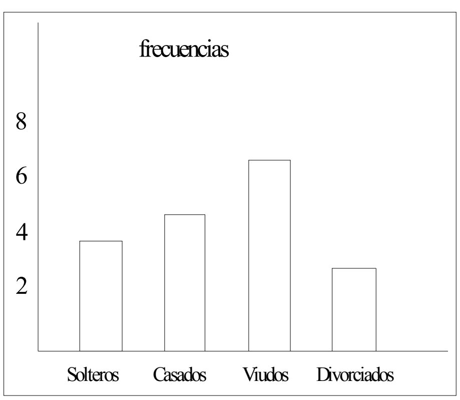


Figura 2.2 Diagrama de barras para una variable cualitativa.

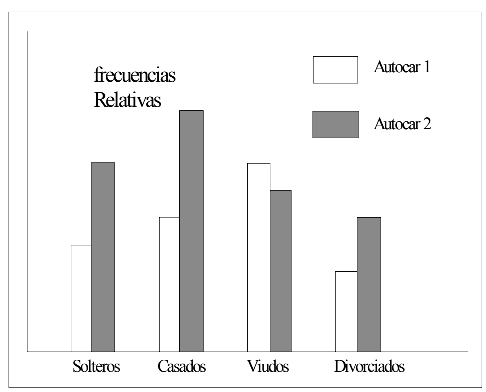


Figura 2.3 Diagramas de barras para comparar una variable cualitativa en diferentes poblaciones. Se ha de tener en cuenta que la altura de cada barra es *proporcional* al número de observaciones (frecuencias relativas).

2.4 TABLA DE FRECUENCIAS PARA VARIABLES CUANTITATIVAS

La distribución de frecuencias para variables cuantitativas por su naturaleza puede ser de dos tipos, para variables cuantitativas discretas o continuas.

2.4.1 Variables cuantitativas discretas

A continuación se presentan la tabla de frecuencia para variables cuantitativas discretas y su representación gráfica.

2.4.1.1 Tabla de frecuencias

Sean $x_1, x_2, ..., x_k$ un conjunto de ***n*** observaciones discretas , se tiene la siguiente tabla de distribución de frecuencias representada por cinco columnas ocupadas por los valores, las frecuencias absolutas, relativas, absolutas acumuladas y relativas acumuladas .

X_i	*Frecuencia absoluta* (n_i)	*Frec. Relativa* (f_i)	*Frec. Acum. absoluta* (N_i)	*Frec. Acum. Relativa* (F_i)
x_1	n_1	f_1	N_1	F_1
x_2	n_2	f_2	N_2	F_2
...	...	...	...	...
x_i	n_i	f_i	N_i	F_i
...	...	...	...	...
x_k	n_k	f_k	$N_k=n$	$F_k=1$
Total:	$\sum n_i = n$	$\sum f_i = 1$		

Tabla 2.3 Tabla para variables cuantitativas.

2.4.1.2 Representación gráfica

Para la representación de las frecuencias absolutas y relativas se usa los diagramas diferenciales y si se desea representar las frecuencias absolutas y relativas acumuladas se hace uso de los diagramas integrales.

Diagramas Diferencial

Si se desea representar gráficamente las distribuciones de frecuencias absolutas (o las relativas), se llevará sobre un eje horizontal los valores x_1, $x_2, \ldots, x_k$, y levantar sobre cada uno de ellos, un segmento vertical de longitud igual a la frecuencia absoluta (o relativa) correspondiente al valor, como se muestra en la Figura 2.4.

Diagramas integrales

Estas graficas consisten en llevar sobre el eje horizontal los diferentes valores x_1, $x_2, \ldots, x_k$, levantando sobre cada uno de estos valores, un segmento vertical de longitud igual a la frecuencia acumulada correspondiente y completando con tramos horizontales hasta el valor inmediato siguiente. Tal como se muestra en la Figura 2.4 que representa al grafico de frecuencias absolutas acumuladas correspondientes al ejemplo 2.1.

Ejemplo 2.1

Se lanzan tres monedas al aire en 8 ocasiones y se contabiliza el número de caras, *X*, obteniendose los siguientes resultados:

$$X \rightarrow 2,1,0,1,3,2,1,2$$

Representar gráficamente el resultado.

Solución:

En primer lugar se observa que la variable X es cuantitativa discreta, presentando las modalidades:

X Є 0, 1, 2, 3 Ordenando a continuación los datos en una tabla estadística, y se representa la misma en la Figura 2.6.

x_i	n_i	f_i	N_i	F_i
0	1	1/8	1	1/8
1	3	3/8	4	4/8
2	3	3/8	7	7/8
3	1	1/8	8	8/8
	$n=8$	1		

Tabla 2.4 Tabla estadística de los resultados del lanzamiento de moneda

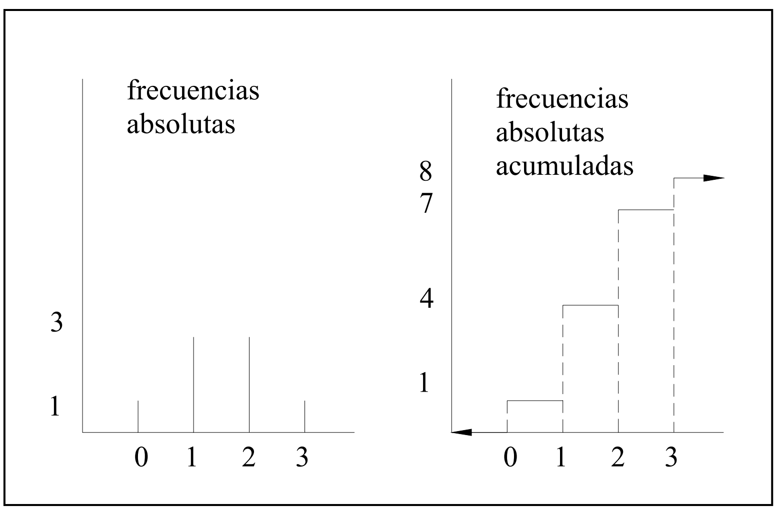


Figura 2.4 Diagrama diferencial (barras) e integral para una variable discreta. Obsérvese que el diagrama integral (creciente) contabiliza el número de observaciones de la variable inferiores o iguales a cada punto del eje de abcisas.

Ejemplo 2.2

Clasificadas 12 familias por su número de hijos se obtuvo:

Número de hijos (x_i)	1	2	3	4
Frecuencias (n_i)	1	3	5	3

Tabla 2.5 Tabla de variables y frecuencias

Comparar los diagramas de barras para frecuencias absolutas y relativas. Realizar el diagrama acumulativo creciente.

Solución:

En primer lugar, se escribe la tabla de frecuencias en el modo habitual:

Variable x_i	**Frec. Abs.** n_i	**Frec. Rel.** f_i	**Frec. Abs. Acumu.** N_i
1	1	0.083	1
2	3	0.250	4
3	5	0.416	9
4	3	0.250	12
	12	1	

Tabla 2.6 Tabla de frecuencias, relativas, absolutas y absolutas acumuladas

Con las columnas relativas a x_i y n_i se realiza el diagrama de barras para frecuencias absolutas, lo que se muestra en la Figura 2.5. Como puede verse es idéntico (salvo un cambio de escala en el eje de ordenadas) al diagrama de barras para frecuencias relativas y que ha sido calculado usando las columnas de x_i y f_i. El diagrama escalonado (acumulado) se ha construido con la información procedente de las columnas x_i y N_i.

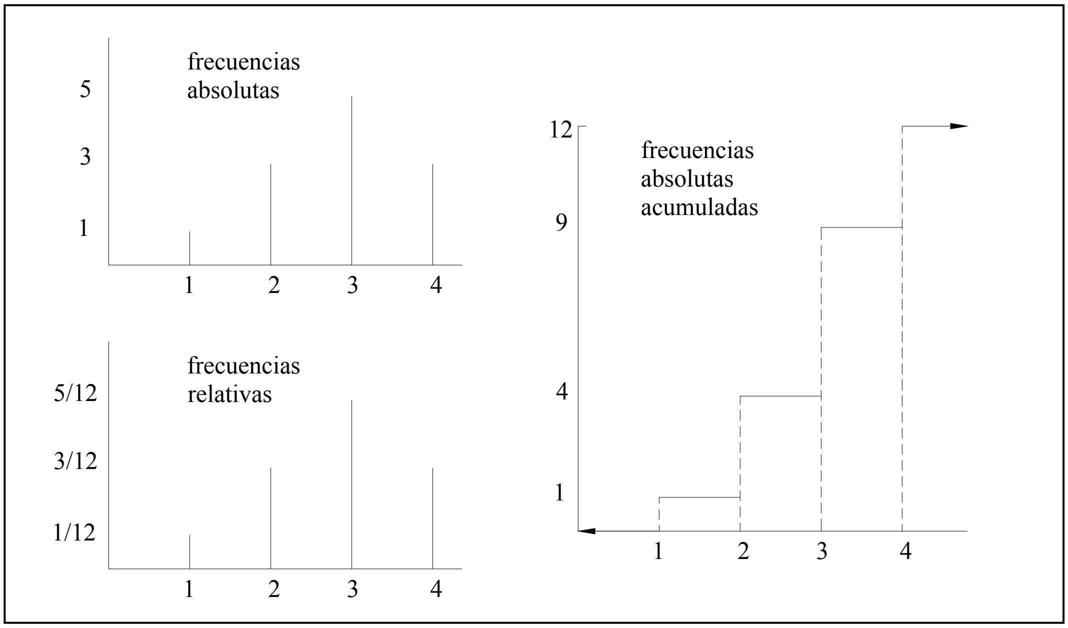


Figura 2.5 Diagramas de frecuencias para una variable discreta

2.4.2 Variables cuantitativas continuas

Para la distribución de frecuencias de variables cuantitativas continuas, se designará el conjunto de datos $x_1, x_2, ..., x_k$ un conjunto de $\boldsymbol{n}$ observaciones, y donde estos valores son del tipo continuo. En este caso debido a que las variables pueden tomar, al menos teóricamente, una infinidad de valores, el proceso de reducción, agrupación o condensación de los datos originales, que conducen a la construcción de tablas de frecuencia, no es tan simple como en el caso de variables cualitativas o cuantitativas discretas, para elaborar esta tabla se debe tener en cuenta las siguientes consideraciones.

2.4.2.1 Elección de las clases

Las clases vendrán definidas mediante lo que denominamos ***intervalos***. En este caso, las categorías que contiene una clase son todos los valores numéricos posibles contenidos en el intervalo, el cual viene normalmente definido de la forma:

$$(l_{i-1}, l_i) = \{x : l_{i-1} \leq x < l_i\}$$

o bien

$$(l_{i-1}, l_i) = \{x : l_{i-1} < x \leq l_i\}$$

Definición 2.8 *En estos casos se llama* ***amplitud del intervalo a_i*** *, a las diferencias entre el limite superior e inferior de un intervalo.*

$$a_i = l_i - l_{i-1} \qquad (2.7)$$

Definición 2.9 *Se denomina* ***marca de clase*** *C_i, a un punto representativo del intervalo. Si éste es acotado, se tiene como marca de clase al punto más representativo, es decir al punto medio del intervalo.*

$$C_i = \frac{l_i + l_{i-1}}{2} \qquad (2.8)$$

La marca de clase no es más que una forma abreviada de representar un intervalo mediante uno de sus puntos. Por ello se tomara como representativo, el punto medio del mismo. Esto está plenamente justificado si se recuerda que cuando se mide una variable continua como el peso, la cantidad con cierto número de decimales que expresa esta medición, no es el valor exacto de la variable, sino una medida que contiene cierto margen de error, y por tanto representa a todo un intervalo del cual ella es el centro.

2.4.2.2 Elección de intervalos

A la hora de seleccionar los intervalos, se plantean varios problemas como son el número de intervalos a elegir y sus tamaños respectivos. La notación más usual para un intervalo sea

$$l_{j-1} - l_j \equiv (l_{j-1}, l_j)$$

El primer intervalo, l_0 -- l_1, puede cerrarse en el extremo inferior para no excluir la observación más pequeña, l_0

$$l_0 - l_1 \equiv (l_0, l_1)$$

Éste es un convenio que se tomará en las páginas que siguen. El considerar los intervalos por el lado izquierdo y abrirlos por el derecho no cambia de modo significativo nada de lo que se expondrá.

Definición 2.10 *El **Número de intervalos (k)**, a utilizar no está determinado de forma fija y por tanto se tomará un **k** que permita trabajar cómodamente y ver bien la estructura de los datos; Como referencia se tomará una de los siguientes valores aproximados:*

$$\text{N}^{\text{o}}\text{ intervalos } \equiv k \approx \begin{cases} \sqrt{n} & \textit{si } \boldsymbol{n} \textit{ no es muy grande} \qquad (2.9) \\ 1+3{,}22 \log n & \textit{en otro caso (regla de Sturges)} \qquad (2.10) \end{cases}$$

Por ejemplo si el número de observaciones que se tiene es n = 100, un buen criterio será agrupar las observaciones en $k = \sqrt{100} = 10$ intervalos. Sin embargo si se tiene n = 1000000, será mas razonable elegir $k=1+3,22logn \approx 20$ intervalos, que $k=\sqrt{100000}=1000$.

En general se recomienda que el número de clase esté entre cinco y veinte.

$$5 \leq \boldsymbol{k} \leq 20$$

Definición 2.11 *La **Amplitud total (A)** esta definida como la diferencia entre la observación mas grande y la mas pequeña de la población (respectivamente* $\boldsymbol{l_0 = x_{min}}$ ***y*** $\boldsymbol{l_k = x_{max}}$*).*

$$A = l_k - l_0 \tag{2.11}$$

La amplitud de cada intervalo a_i suele tomarse constante de forma que la amplitud de cada intervalo sea:

$a_i = a \quad \forall i = 1,\dots,k$ entonces:

$$a = \frac{A}{k} \tag{2.12}$$

Así la división en intervalos podría hacerse tomando:

$$\begin{aligned}
&l_0 = x_{\min} \\
&l_1 = l_0 + a \\
&\dots \\
&l_k = x_{\max} = l_0 + ka
\end{aligned}$$

Observación

Podría ocurrir que la cantidad ***a*** fuese un número *muy desagradable* a la hora de escribir los intervalos (ej. a=10.325467). En este caso, se recomienda variar simétricamente los extremos, $l_0 < x_{min} < x_{max} < l_k$, de forma que se tenga que ***a*** es un número más simple (ej. a =10).

2.4.2.3 Tabla de frecuencias

La tabla de frecuencias para variables cuantitativas continuas está representada primeramente por una columna con los intervalos de clase, seguida de la marca de clase y posteriormente las frecuencias absolutas, relativas, absolutas acumuladas y relativas acumuladas, tal como se muestra en la Tabla 2.7.

Intervalo de clase (I_i)	***Marca de clase (C_i)***	***Frecuencia absoluta (n_i)***	***Frec. Relativa (f_i)***	***Frec. Acum. absoluta (N_i)***	***Frec. Acum. Relativa (F_i)***
$\boldsymbol{I_1}$	C_1	n_1	f_1	N_1	F_1
$\boldsymbol{I_2}$	C_2	n_2	f_2	N_2	F_2
. . .	. . .	. . .	. . .	. . .	. . .
$\boldsymbol{I_i}$	C_i	n_i	f_i	N_i	F_i
. . .	. . .	. . .	. . .	. . .	. . .
$\boldsymbol{I_k}$	C_k	n_k	f_k	$N_k=n$	$F_k=1$
		$\boldsymbol{n}$	$\boldsymbol{1}$		

Tabla 2.7 Tabla para variables cuantitativas continuas.

2.4.2.4 Representación Gráfica

Cuando las variables son continuas, se utiliza como diagramas diferenciales los *histogramas* y polígonos de frecuencias como diagramas integrales los *polígonos de frecuencias acumuladas*.

Histograma de frecuencias

Se usa para representar gráficamente las distribuciones de frecuencias absolutas o relativas de variables cuantitativas continuas agrupadas en clases. Consta de una serie de rectángulos semejantes a los del diagrama de barras; sin embargo las barras del histograma se colocan sólo verticalmente y deben ir uno al lado de las otras sin que haya un espacio que las separe. La base de cada rectángulo es la amplitud de la clase de la variable correspondiente. Se construye como sigue:

1. Se lleva sobre un eje horizontal los límites de los intervalos de clase.

   $l_0, l_1, l_2, \ldots, l_k$

2. Sobre cada intervalo de clase se levantan rectángulos que tengan como área exactamente la frecuencia absoluta (o relativa) correspondiente. Es decir, al intervalo de clase $l_{j-1} - l_j$ con amplitud $\boldsymbol{a_i}$ y frecuencia absoluta $\boldsymbol{n_i}$ (o frecuencia relativa $\boldsymbol{f_i}$) , le corresponderá un rectángulo de base $\boldsymbol{a_i}$ y altura $\boldsymbol{n_i/a_i}$ *(o respectivamente* $\boldsymbol{f_i/a_i}$) , con lo cual se garantiza que el área de dicho rectángulo , es decir base por altura, coincida con $\boldsymbol{n_i}$ (o con $\boldsymbol{f_i}$ según el caso)

$$a_i \times \frac{n_i}{a_i} = n_i \qquad \text{o} \qquad (a_i \times \frac{f_i}{a_i} = f_i)$$

A si la superficie total de los rectángulos o sea el área limitada por el histograma será igual a ***n*** (o a la unidad) , lo cual permite comparar visualmente dos distribuciones diferentes.

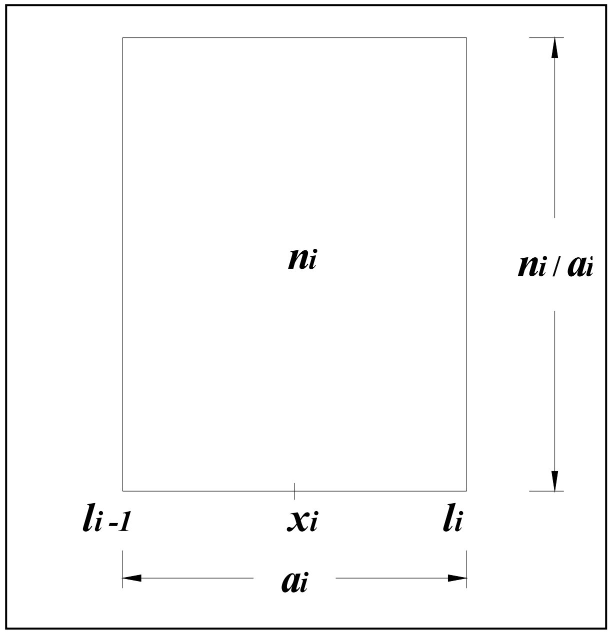


Figura 2.6 Construcción del histograma

Polígono de frecuencias

Los polígonos de frecuencias absolutas o relativas, se obtienen uniendo los puntos medios de las bases superiores de los rectángulos en el histograma de frecuencias absolutas o relativas, respectivamente.

Si se observa la Figura 2.6 se puede ver que en la grafica del polígono de frecuencias se ha agregado un intervalo de clase de cero observaciones en cada extremo de la distribución, lo cual permite al polígono alcanzar el eje horizontal en cada extremo de la distribución, lo cual permite al polígono alcanzar el eje horizontal en ambos extremos.

No existe diferencia al momento de usar un histograma o un polígono de frecuencias , sin embargo cuando se quiere presentar más de una serie estadística en el mismo grafico con fines comparativos , deben utilizarse polígonos de frecuentas. Ver Figura 2.7.

Polígono de frecuencias acumuladas (Ojivas)

Una ojiva es la representación gráfica de una distribución de frecuencias absolutas acumuladas o las frecuencias relativas acumuladas.

Véase la parte inferior de la Figura 2.7, en la que se representa a modo de ilustración los diagramas correspondientes a la variable cuantitativa continua expresada en la tabla siguiente:

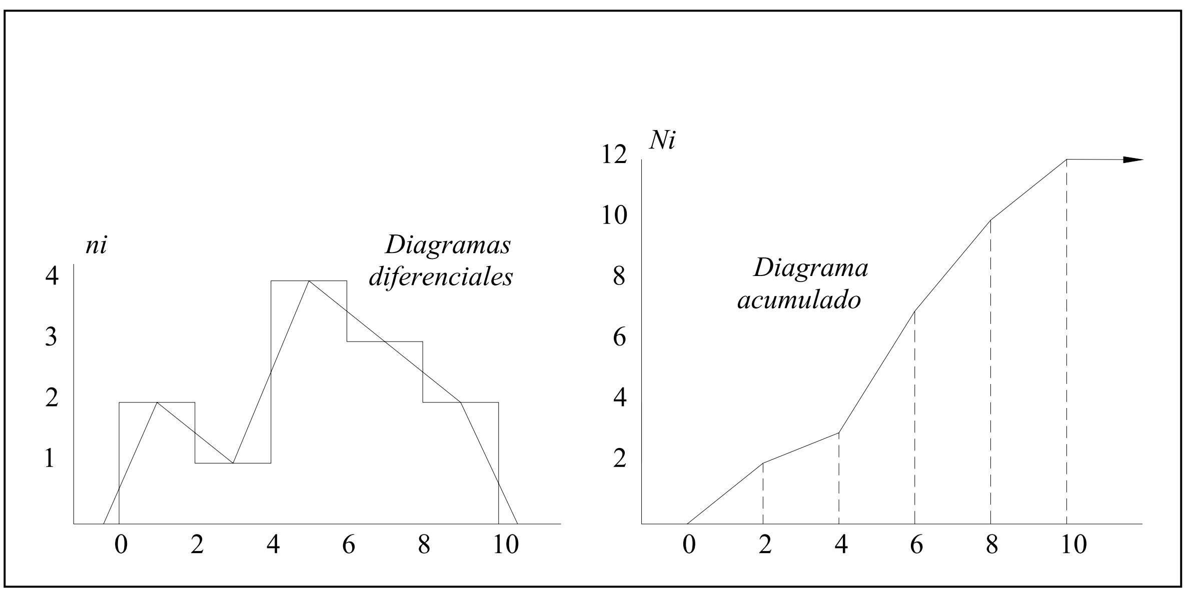


Figura 2.7 Diagramas diferenciales e integrales para una variable continua.

Intervalos	c_i	n_i	N_i
0 -- 2	1	2	2
2 -- 4	3	1	3
4 -- 6	5	4	7
6 -- 8	7	3	10
8 --10	9	2	12
		12	

Tabla 2.8 Tabla de frecuencias de la Figura 2.8

Ejemplo 2.3

Sobre un grupo de n=21 personas se realizan las siguientes observaciones de sus pesos, medidos en kilogramos:

$X \rightarrow x_1, x_2, \ldots, x_{21}$						
58	42	51	54	40	39	49
56	58	57	59	63	58	66
70	72	71	69	70	68	64

Tabla 2.9 Tabla estadística con valores de pesos de 21 personas

Agrupar los datos en una tabla estadística.

Solución:

En primer lugar se debe observar que si se denomina X a la variable "peso de cada persona" esta es una variable de tipo cuantitativa y continua. Por tanto a la hora de ser ordenados los resultados en una tabla estadística, esto se ha de hacer agrupándolos en intervalos de longitud conveniente. Esto lleva a perder cierto grado de precisión. Para que la pérdida de información no sea muy relevante se debe seguir el criterio $k = \sqrt{n} = \sqrt{21}$ intervalos (no son demasiadas las observaciones). En este punto se puede tomar bien k=4 o bien k=5. Arbitrariamente se elegirá una de estas dos posibilidades. Por ejemplo, tomando k=5.

Lo siguiente es determinar la longitud de cada intervalo, $\boldsymbol{a_i}$ $\forall i = 1,...,5$. Lo más cómodo es tomar la misma longitud en todos los intervalos, $\boldsymbol{a_i = a}$ (aunque esto no tiene por qué ser necesariamente así), donde :

$$a = \frac{A}{5} = \frac{33}{5} = 6.6$$
$$A = l_5 - l_0 = 72 - 39 = 33$$
$$l_0 = x_{\min} = 39$$
$$l_5 = x_{\max} = 72$$

Entonces se tomara k=5 intervalos de longitud a=6.6 comenzando por $l_0 = x_{min}$=39 y terminando en l_5=33:

	Interv. $l_{i-1} -- l_i$	**M. clase C**	**F.a.** n_i	**F.r.** f_i	**F.a.a.** N_i	**F.r.a.** F_i
$i=1$	39 -- 45.6	42.3	3	0.1428	3	0.1428
$i=2$	45.6 -- 52.2	48.9	2	0.0952	5	0.2381
$i=3$	52.2 -- 58.8	55.5	6	0.2857	11	0.5238
$i=4$	58.8 -- 65.4	62.1	3	0.1428	14	0.6667
$i=5$	65.4 -- 72	68.7	7	0.3333	21	**≈ 1**
			21	**≈ 1**		

Tabla 2.10 Calculo de frecuencia, relativa, absoluta y frecuencias acumuladas.

Otra posibilidad a la hora de construir la tabla, y que permite trabajar con cantidades más simples a la hora de construir los intervalos, es la siguiente. Como la regla para elegir l_0 y l_5 no es muy estricta se puede hacer la siguiente elección:

$$a' = 7$$
$$A' = a' \cdot 5 = 35$$
$$d = A' - A = 35 - 33 = 2$$
$$l_0 = x_{\min} - \frac{d}{2} = 39 - 1 = 38$$
$$l_5 = x_{\max} + \frac{d}{2} = 72 + 1 = 73$$

ya que así la tabla estadística no contiene decimales en la expresión de los intervalos, y el exceso ***d***, cometido al ampliar el rango de las observaciones desde ***A*** hasta ***A'***, se reparte del mismo modo a los lados de las observaciones menores y mayores:

	Interv. $l_{i-1} -- l_i$	**M. clase C**	**F.a.** n_i	**F.r.** f_i	**F.a.a.** N_i	**F.r.a.** F_i
$i=1$	38 -- 45	41,5	3	0.1428	3	0.1428
$i=2$	45 -- 52	48,5	2	0.0952	5	0.2381
$i=3$	52 -- 59	55,5	7	0.3333	12	0.5714
$i=4$	59 -- 66	62,5	3	0.1428	15	0.7143
$i=5$	66 -- 73	69,5	6	0.2857	21	**≈1**
			21	**≈1**		

Tabla 2.11 Calculo de frecuencia, relativa, absoluta y frecuencias acumuladas.

Ejemplo 2.5

La siguiente distribución se refiere a la duración en horas (completas) de un lote de 500 tubos:

Duración en horas	**Número de tubos**
300 -- 500	50
500 -- 700	150
700 -- 1.100	275
más de 1.100	25
	Total 500

Tabla 2.12 Datos ejemplo 2.5

a) Representar el histograma de frecuencias relativas y el polígono de frecuencias.

b) Trazar la curva de frecuencias relativas acumuladas.

c) Determinar el número mínimo de tubos que tienen una duración inferior a 900 horas.

Solución:

a) En primer lugar se observa que la variable en estudio es discreta *(horas completas)*, pero al tener un rango tan amplio de valores resulta más conveniente agruparla en intervalos, como si de una variable continua se tratase. La consecuencia es una ligera perdida de precisión.

El último intervalo está abierto por el límite superior. Dado que en él hay 25 observaciones puede ser conveniente cerrarlo con una amplitud ``razonable''. Todos los intervalos excepto el tercero tienen una amplitud de 200 horas, luego podría cerrarse el último intervalo en 1.300 horas.

Antes de realizar el histograma conviene hacer una observación importante. El histograma representa las frecuencias de los intervalos mediante *áreas* y no mediante *alturas*. Sin embargo nos es mucho más fácil hacer representaciones gráficas teniendo en cuenta estas últimas. Si todos los intervalos tienen la misma amplitud no es necesario diferenciar entre los conceptos de área y altura, pero en este caso el tercer intervalo tiene una amplitud doble a los demás, y por tanto debe repartirse su área en un rectángulo de base doble (lo que reduce su altura a la mitad).

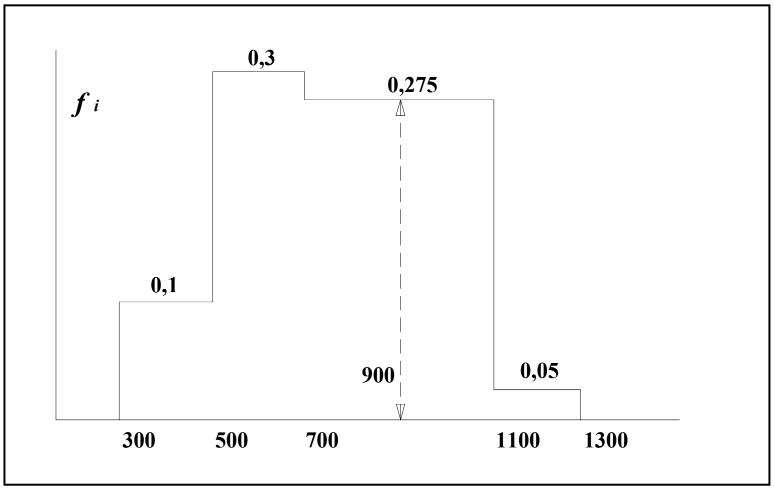


Figura 2.8 Histograma. Obsérvese que la altura del histograma en cada intervalo es f_i' que coincide en todos con f_s salvo en el intervalo 700 -- 1.100 en el que $f_i' = 1/2 f_i$ 'ya que la amplitud de ese intervalo es doble a la de los demás.

b) Así será conveniente añadir a la habitual tabla de frecuencias una columna que represente a las amplitudes a_i de cada intervalo, y otra de frecuencias relativas rectificadas, f_i', para representar la altura del histograma. Los gráficos requeridos se representan en las Figuras 2.8 y 2.9.

Intervalos	a_i	n_i	f_i	f_i'	F_i
300 -- 500	200	50	0.100	0.100	0.100
500 -- 700	200	150	0.300	0.300	0.400
700 -- 1.100	400	275	0.550	0.275	0.950
1.100 -- 1.300	200	25	0.050	0.050	1.000
		n=500			

Tabla 2.13 Tabla estadística que incluye frecuencias relativas rectificadas

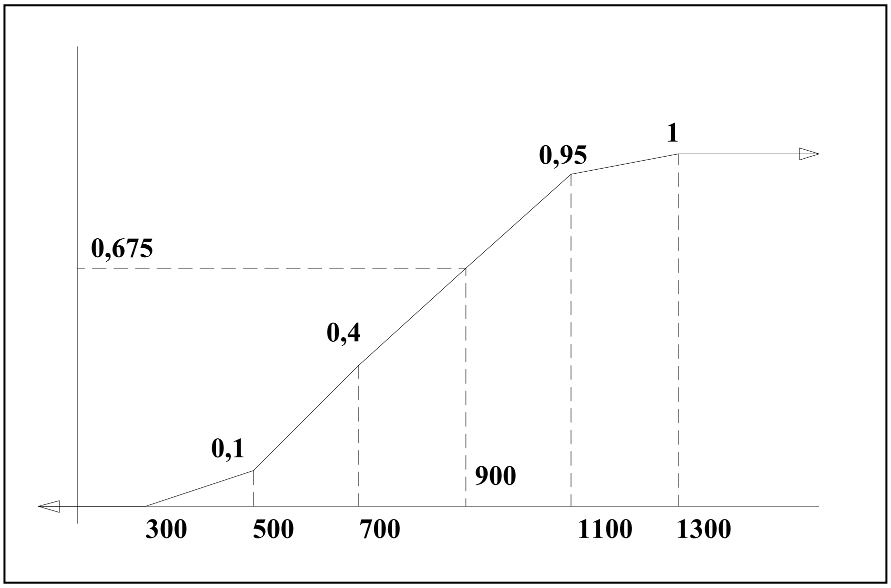


Figura 2.9 Diagrama acumulativo de frecuencias relativas

Intervalos	a_i	n_i	f_i	f_i'	F_i
300 -- 500	200	50	0.100	0.100	0.100
500 -- 700	200	150	0.300	0.300	0.400
700 -- 1.100	400	275	0.550	0.275	0.950
1.100 -- 1.300	200	25	0.050	0.050	1.000
		n=500			

Tabla 2.14 Tabla estadística que incluye frecuencias relativas rectificadas

c) Por otro lado, mirando la Figura 2.9, se ve que sumando frecuencias relativas, hasta las 900 horas de duración hay

0,10 + 0,30 + 0,275 = 0,675 = 67,5 % de los tubos.

Como en total son 500 tubos, el número de tubos con una duración igual o menor que 900 horas es **0.675 x 500 =337.5**, redondeando, 338 tubos.

Ejemplo 2.6

Dada la serie histórica de caudales medios anuales en m^3/s (tabla 2.15), de una estación, para el periodo 1911-1980, calcular las frecuencias absolutas, relativa, acumulada, función densidad, función acumulada, graficar Histograma de frecuencias y polígono de frecuencias.

Año	Caudal m3/s	Año	Caudal m3/s	Año	Caudal m3/s
1911	7,91	1935	24,58	1959	22,88
1912	8,01	1936	28,49	1960	17,57
1913	13,27	1937	10,05	1961	14,60
1914	16,39	1938	28,01	1962	31,14
1915	80,83	1939	34,92	1963	18,20
1916	60,08	1940	31,36	1964	24,69
1917	21,55	1941	42,74	1965	22,99
1918	27,71	1942	12,94	1966	11,78
1919	28,63	1943	41,16	1967	32,26
1920	30,27	1944	35,90	1968	4,76
1921	33,43	1945	33,76	1969	12,70
1922	35,16	1946	29,28	1970	16,19
1923	27,21	1947	19,17	1971	30,14
1924	15,58	1948	29,37	1972	30,57
1925	64,81	1949	30,06	1973	45,38
1926	51,26	1950	9,67	1974	18,91
1927	33,48	1951	10,42	1975	34,99
1928	25,79	1952	33,99	1976	21,49
1929	25,80	1953	42,17	1977	29,26
1930	18,93	1954	16,00	1978	4,58
1931	16,15	1955	22,78	1979	12,46
1932	38,30	1956	32,69	1980	3,14
1933	54,54	1957	34,28		
1934	59,40	1958	20,24		

Tabla 2.15 Serie histórica de caudales medios anuales en m^3/s, periodo (1911-1980).

Solución:

1. Ordenando los datos de la Tabla 2.15 Se obtiene la Tabla 2.16

3.14	4.53	4.76	7.91	8.01	9.67	10.05
10.42	11.78	12.46	12.70	12.92	13.27	14.60
15.58	16.00	16.15	16.19	16.39	17.57	18.20
18.91	18.93	19.77	20.24	21.49	21.55	22.78
22.88	22.99	23.99	24.58	24.69	25.79	25.80
27.21	27.71	28.01	28.49	28.63	29.26	29.28
29.37	30.06	30.14	30.27	30.57	31.14	31.36
32.26	32.69	33.43	33.48	33.76	34.28	34.92
34.99	35.16	35.90	38.30	41.16	42.17	42.74
45.38	51.26	54.54	59.40	60.08	64.81	80.83

Tabla 2.16 Serie histórica de caudales medios anuales en m^3/s, periodo (1911-1980) ordenado ascendentemente

2. Calculo de la *amplitud de intervalo total* ***A***

   De la ecuación (2.11) , se tiene :

   $\boldsymbol{A}$ = 80.83-3.14

   $\boldsymbol{A}$ = 77.69

3. Calculo del número de intervalos $\boldsymbol{k}$:

   De la ecuación (2.10) resulta:

   $\boldsymbol{k}$ =1+3,22 log n = 1+3.22log(70) = 6.94

   Redondeando

k = 7

4. Calculo de ***a*** y de *límites de clase*:

   De la ecuación (2.12) Se obtiene:

   $a = 77.69 / 7 = 11.09$

Si quisiéramos redondear a fin de que los límites y las marcas de clase resulten números mas simples, podría ser:

$$a' = 12$$
$$A' = a' \bullet k = 12 \times 7 = 84$$
$$d = A' - A = 84 - 77.69 = 6.31$$

5. Calculo de los límites de clase

$$l_0 = x_{\min} - \frac{d}{2} = 3.14 - \frac{6.31}{2} = -0.015 \approx 0$$
$$l_7 = x_{\max} + \frac{d}{2} = 80.83 + \frac{6.31}{2} = 83.98 \approx 84$$

6. Calculo de las marcas de clase

   De la ecuación (2.8) la marca de clase del primer intervalo es:

$$c_i = \frac{0+12}{2} = 6$$

   las marcas de clase de los otros intervalos se obtienen sumando ***a*** a la precedente.

7. Calculo de la frecuencia absoluta

A partir de los datos ordenados de la tabla 2.16 es fácil determinar el número de valores comprendidos en cada intervalo, así en el primer intervalo entre 0-12, hay 9 valores y así sucesivamente.

8. Calculo de la frecuencia relativa

De la ecuación (2.2) dividiendo la frecuencia absoluta de cada intervalo entre el número total de datos tenemos para el primer intervalo:

n_i=9

n=70

f_i=9/70=0.1286

para los demás intervalos se sigue el mismo procedimiento

9. Calculo de frecuencias acumuladas

Para encontrar las frecuencias acumuladas se tiene que efectuar una sumatoria de acuerdo a la ecuación (2.3)y (2.4).

Los resultados finales aparecen mostrados en la Tabla 2.13

Intervalo de clase ***I***	**M. clase C**	**Frec. Abs.** n_i	**Frec. Rel.** f_i	**Frec. Abs. Acum.** N_i	**Frec. Rel. Acum.** F_i
0 - 12	6	9	0.1286	9	0.1286
12 - 24	8	22	0.3143	31	0.4429
24 - 36	30	28	0.4000	50	0.8429
36 - 48	42	5	0.0714	33	0.9143
48 - 60	54	3	0.04296	8	0.9571
60 - 72	66	2	0.0286	5	0.9857
72 - 84	78	1	0.0143	3	1.0000

Tabla 2.17 Calculo de frecuencia, relativa, absoluta y frecuencias.

10. El histograma de frecuencias se obtiene graficando los valores de intervalos de clase con las frecuencias relativas.

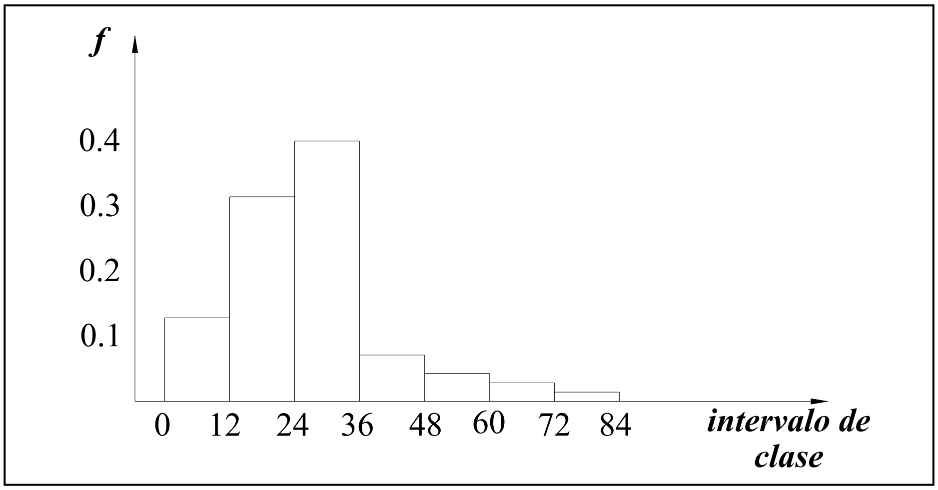


Figura 2.10 Histograma de frecuencias relativas

11. El polígono se obtiene graficando las columnas que contiene las marcas de clase y las de frecuencia relativa; se observa que se han agregado en las marcas de clase los valores de -6 y 90.

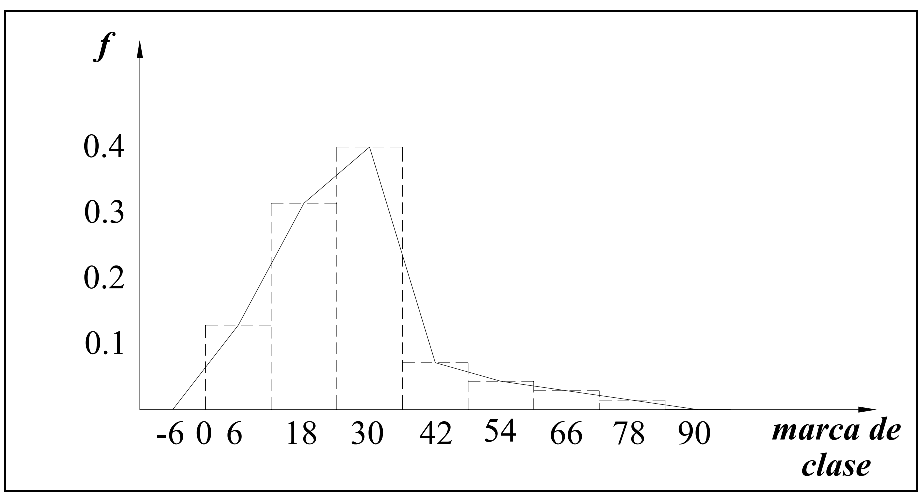


Figura 2.11 Polígono de frecuencias relativas

2.5 MEDIDAS DESCRIPTIVAS

Las tablas o cuadros estadísticos y los distintos tipos de gráficos que se han estudiado anteriormente, constituyen diversos modos de resumir o reducir un conjunto de datos a unas pocas cifras, que aisladamente, o dispuestas en forma tabular o gráfica, sirven para transmitir las características principales de la información representada en los datos y tienen elementos descriptivos que hacen innecesario el examen de todos los datos.

Las cifras descriptivas que se obtienen como función de una muestra $(X_1,X_2,X_3 \ldots X_n)$; es decir, como función de un conjunto de datos (que representa un subconjunto de la población), se llama Estadígrafo o números índices, estadísticas, estadísticos, etc.

Cuando se esta analizando la población entera un termino mas general para estas cifras descriptivas seria el de *medidas*.

Los fenómenos estudiados no suelen ser constantes, por lo que será necesario que junto a una medida que indique el valor alrededor del cual se agrupan los datos, se asocie una medida que haga referencia a la variabilidad que refleje dicha fluctuación.

En este sentido pueden examinarse varias características, las cuales pueden analizarse a través de cuatro tipos de Medidas:

- ***Medidas de posición.***
- ***Medidas de dispersión.***
- ***Medidas de concentración.***
- ***Medidas de forma.***

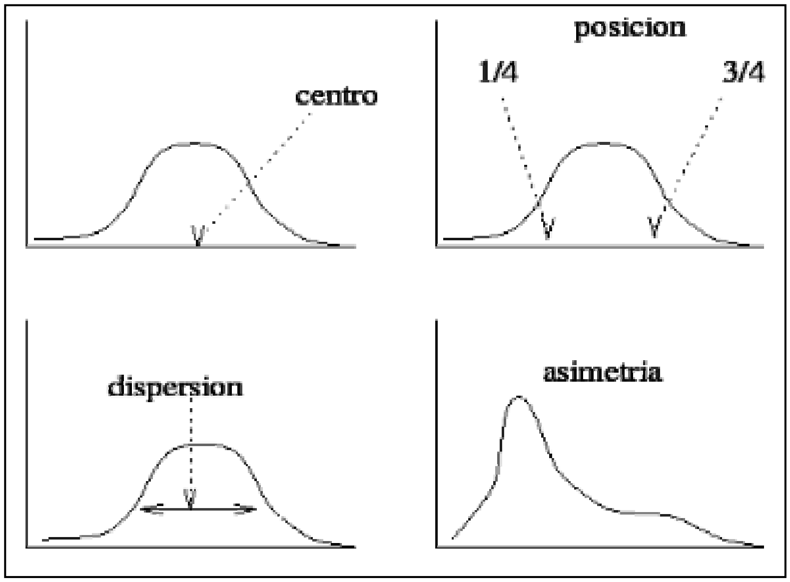


Figura 2.12 Medidas representativas de un conjunto de datos estadísticos

2.5.1 Medidas de posición

Se denominan así, a las medidas que indican un valor, alrededor del cual los datos parecen agruparse de cierta manera, como si fuese "centro de gravedad de los datos", tal es el caso de los estadígrafos de tendencia central. O bien, dejando una cierta cantidad de valores por debajo o por encima de él.

Las medidas de posición son de dos tipos:

- ***Medidas de posición central***: informan sobre los valores medios de la serie de datos.

- ***Medidas de posición no central***: informan de cómo se distribuye el resto de los valores de la serie.

2.5.1.1 Medidas de posición central

Las principales medidas de posición central son los siguientes:

2.5.1.1.1 Media

Es el valor medio ponderado de la serie de datos. Se pueden calcular diversos tipos de media, siendo las más utilizadas:

a) Media aritmética: También se le conoce como promedio ya que es el promedio de las lecturas o mediciones individuales que se tienen en la muestra, se determina con la fórmula siguiente:

- **Para Datos no agrupados:**

Definición 2.12 ***La media aritmética*** *o simplemente media (o promedio) de un conjunto de elementos X1, X2 ,..Xn de tamaño n de una variable o característica X, y se define como la suma de todos los valores observados, dividida por el número total de observaciones n. Cuando se calcula la media para una población, esta se denota por "μ", y cuando se trata de una muestra, por "$\overline{x}$ "*, *Es decir:*

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (2.13)$$

O bien :

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (2.14)$$

Donde:

μ = media poblacional

$\overline{x}$ = media muestral

x_i = valor i-ésimo de la muestra

n = número de datos de la población o muestra

- **Para datos agrupados**

Cuando se conocen los datos en forma individual, se puede calcular la media aritmética tradicional en forma directa aplicando las formulas anteriores. En cambio, si la información se suministra a través de ***n*** datos agrupados en ***m*** clases, se usa la siguiente formula:

$$\overline{x} = \frac{\sum_{i=1}^{m} n_i C_i}{n} \tag{2.15}$$

Donde :

$\boldsymbol{n_i}$ = Frecuencia absoluta (número de observaciones) en el intervalo ***i.***

$\boldsymbol{C_i}$ = marca de clase del intervalo ***i.***

i = número de intervalos de clase.

n = numero de observaciones de la muestra.

Definición 2.13 *La **media aritmética ponderada** permite calcular un promedio que toma en cuenta la ponderación de los datos con respecto a un factor, es un caso particular de la fórmula del cálculo de la media para los datos agrupados, su fórmula es:*

$$\overline{x}_p = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} \tag{2.16}$$

Donde:

$\overline{x}_p$ = media ponderada

x_i = valor i-ésimo de la muestra

f_i = valor del factor de ponderación del i-ésimo valor de la muestra.

n = número de observaciones

Proposición 2.1.- La sumatoria de los cuadrados de los desvíos es mínima. Más conocida como el principio de los mínimos cuadrados. Esto se demuestra así:

Sea un valor cualquiera a, para saber que valor a hace mínima la sumatoria, se saca la primera derivada y se iguala a cero:

$$\frac{d(di)}{da} = \frac{d((Xi-a)2)}{da} = 2(Xi-a)(-1) = -2(xi-a) = 0$$

El único valor que cumple con la propiedad anterior, es el promedio. O sea ,

$(Xi-a) = 0$ sólo si $a = \overline{x}$, de acuerdo a la propiedad anterior. Luego,

$$\frac{d^2(di^2)}{da^2} = \frac{d^2((Xi-a)^2)}{da^2} = (-2)(-a) = 2na$$

como $(2na)$ es una expresión siempre positiva, cuando $a = \overline{x}$ hay un mínimo

Proposición 2.2.- Si a cada valor medio se le suma (o resta) una constante ***a*** , la nueva media aritmética es igual a la anterior, más (o menos) dicha constante. Esto se demuestra así:

Si $Zi = Xi + a$, entonces $\overline{z} = \frac{\sum_{i=1}^{n}(z_i)}{n} = \frac{\sum_{i=1}^{n}(x_i + a)}{n} = \sum_{i=1}^{n}(\frac{xi}{n}) + \frac{n.a}{n}$

O sea $\bar{z} = \bar{x} + a$

Proposición 2.3.- Si a cada valor medido se le multiplica (o divide) por una constante ***a*** , la nueva media aritmética es igual a la anterior, multiplicada (o dividida) por dicha constante. Esto se demuestra:

Si $Zi = Xi\ *\ a$, entonces $\bar{z} = \dfrac{\sum_{i=1}^{n}(z_i)}{n} = \dfrac{\sum_{i=1}^{n}(x_i \times a)}{n} = a\sum_{i=1}^{n}(\dfrac{xi}{n}) = a\bar{x}$

Proposición 2.4.- La suma de las *diferencias de la variable con respecto a la media* es nula, es decir,

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

Demostración:

Basta desarrollar la sumatoria para obtener

$$\sum_{i=1}^{n}(x_i - \bar{x}) = (x_i - \bar{x}) + ... + (x_n - \bar{x}) = (x_1 + ... + x_n) - n\bar{x} = n\bar{x} - n\bar{x} = 0$$

Este resultado indica que el error cometido al aproximar un valor cualquiera de la variable, por ejemplo x_1, mediante el valor central $\bar{x}$, es compensado por los demás errores:

Error aprox. de x_1 $\equiv$ $x_1 - \bar{x} = \sum_{i=2}^{n}(x_i - \bar{x})$

Si los errores se consideran con signo positivo, en este caso no pueden compensarse. Esto ocurre si se toma como medida de error alguna de las siguientes:

$\sum_{i=1}^{n}(x_i - \overline{x})^2 \geq 0$ Error cuadrático

$\sum_{i=1}^{n}|x_i - \overline{x}| \geq 0$ Error absoluto

$\max_{i=1,\ldots,n}|x_i - \overline{x}| \geq 0$ Error máximo

que son cantidades *estrictamente* positivas si algún. $x_i = \overline{x}$

Ejemplo 2.7

Obtener las desviaciones con respecto a la media en la siguiente distribución y comprobar que su suma es cero.

l_{i-1} - l_i	n_i
0 - 10	1
10 - 20	2
20 - 30	4
30 - 40	3

Tabla 2.18 Datos agrupados en intervalos

Solución:

$l_{i-1} - l_i$	n_i	x_i	$x_i\ n_i$	$x_i - \overline{x}$	$(x_i - \overline{x})n_i$
0 - 10	1	5	5	-19	-19
10 - 20	2	15	30	-9	-18
20 - 30	4	25	100	1	4
30 - 40	3	35	105	11	33
	$n = 10$		$\sum x_i n_i = 240$		$\sum = 0$

Tabla 2.19 Tabla para el calculo de media aritmética y desviaciones

La media aritmética es:

$$\overline{x} = \frac{1}{n}\sum x_i n_i = \frac{240}{10} = 24$$

Como se puede comprobar sumando los elementos de la última columna,

$$\sum (x_i - \overline{x}) \bullet n_i = 0$$

Proposición 2.5.- Llamada ***Proposición König,*** para cualquier posible valor k que consideremos como candidato a medida central, $\overline{x}$ lo mejora en el sentido de los mínimos cuadrados, es decir;

$$\sum_{i=1}^{n} (x_i - \overline{x})^2 < \sum_{i=1}^{n} (x_i - k)^2 \qquad si\ k \neq \overline{x}$$

Demostración:

Sea $k \neq \overline{x}$. Se ve que el error cuadrático cometido por k es mayor que el de $\overline{x}$.

$$\sum_{i=1}^{n}(x_i-k)^2=\sum_{i=1}^{n}\left[x_i-(k-\overline{x}+\overline{x})\right]^2 \qquad \text{(sumando y restando } \overline{x}\text{)}$$

$$=\sum_{i=1}^{n}\left[(x_i-\overline{x})-(k-\overline{x})\right]^2 \qquad \text{(y usando el binomio de Newton…)}$$

$$=\sum_{i=1}^{n}(x_i-\overline{x})^2-2(\overline{x}-k)\underbrace{\sum_{i=1}^{n}(x_i-\overline{x})}_{0}+\underbrace{\sum_{i=1}^{n}(k-\overline{x})}_{n(k-\overline{x})^2>0}$$

$$>\sum_{i=1}^{n}(x_i-\overline{x})^2$$

Proposición 2.6.- Linealidad de la media

$$Y=a+bX\Rightarrow\overline{y}=a+b\overline{x}$$

Proposición 2.7.- Dados r grupos con n_1, n_2, ..., n_r observaciones y siendo $\overline{x}_1,\overline{x}_2,\ldots,\overline{x}_r$, las respectivas medias de cada uno de ellos. Entonces la media de las $\boldsymbol{n=n_1+\ldots+n_r}$ observaciones es **:**

$$\overline{x}=\frac{n_1\overline{x}_1+\ldots+n_r\overline{x}_r}{n_1+\ldots+n_r}$$

Demostración:

Se llamará x_{ij} a la j-ésima observación del grupo i; Entonces:

$$\left.\begin{array}{l} 1^{er}\, grupo \rightarrow x_{11} \quad \ldots \quad x_{1n_1} \\ 2^{o}\, grupo \rightarrow x_{21} \quad \ldots \quad x_{2n_2} \\ \quad\quad \ldots \\ r^{esimo}\, grupo \rightarrow x_{r1} \quad \ldots \quad x_{rn_r} \end{array}\right\} \quad \begin{array}{l} \bar{x}_1 = (\sum_{j=1}^{n_1} x_{ij})/n_1 \\ \bar{x}_2 = (\sum_{j=1}^{n_2} x_{ij})/n_2 \\ \quad \ldots \\ \bar{x}_r = (\sum_{j=1}^{n_r} x_{ij})/n_r \end{array}$$

Así, agrupando convenientemente las observaciones se llega a que

$$\bar{x} = \frac{(x_{11} + \ldots + x_{1n_1}) + (x_{22} + \ldots + x_{2n_2}) + \ldots + (x_{r1} + \ldots + x_{rn_r})}{n_1 + n_2 + \ldots + n_r}$$
$$= \frac{n_1\bar{x}_1 + \ldots + n_r.\bar{x}_r}{n}$$

Observación

A pesar de las buenas propiedades que ofrece la media, ésta posee algunos inconvenientes:

- Uno de ellos es que es muy sensible a los valores extremos de la variable: ya que todas las observaciones intervienen en el cálculo de la media, la aparición de una observación extrema, hará que la media se desplace en esa dirección. En consecuencia, no es recomendable usar la media como medida central en las distribuciones muy asimétricas;
- Depende de la división en intervalos en el caso de variables continuas.
- Si se considera una variable discreta, por ejemplo, *el número de hijos en las familias de Cochabamba* el valor de la media puede no pertenecer al conjunto de valores de la variable; Por ejemplo $\bar{x}$ = 2,5 hijos.

Cálculo abreviado

Se puede utilizar la *linealidad de la media* para simplificar las operaciones necesarias para su cálculo mediante un *cambio de origen* y de *unidad de medida*. El método consiste en lo siguiente:

1. Tomando ***a*** un número que exprese aproximadamente el tipo de unidad con la que se trabaja. Por ejemplo, si las unidades que se usan son millones, se toma a =1.000.000.

2. Seleccionando un punto cualquiera de la zona central de la tabla, x_0. Este punto jugará el papel de origen de referencia.

3. Cambiando a la variable

$$z = \frac{X - x_0}{a} \qquad \Rightarrow \quad \bar{z} = \frac{\bar{x} - x_o}{a}$$
$$\Rightarrow \quad \bar{x} = a\ \bar{z} + x_o$$

4. Se construye de este modo la tabla de la variable Z, para la que es más fácil calcular $\bar{z}$ directamente, y después se calcula $\bar{x}$ mediante la relación (Ecc. 2.14).

b) Media geométrica

Definición 2.14 *La **media geométrica** $\bar{x}_g$, es la media de los logaritmos de los valores de la variable:*

$$\log \bar{x}_g = \frac{\log x_1 + ... + \log x_n}{n} \tag{2.17}$$

Luego

$$\bar{x}_g = \sqrt[n]{x_1 x_2 \ldots x_n} \tag{2.18}$$

Si los datos están agrupados en una tabla, entonces se tiene:

$$\bar{x}_g = \sqrt[n]{x_1^{n_1} x_2^{n_2} \ldots x_k^{n_k}} \tag{2.19}$$

c) Media armónica

Definición 2.15 ***La media armónica*** $\bar{x}_a$*, se define como el recíproco de la media aritmética de los recíprocos, es decir,*

$$\frac{1}{\bar{x}_a} = \frac{\frac{1}{x_1} + \cdots + \frac{1}{x_n}}{n}$$

Por tanto,

$$\bar{x}_a = \frac{n}{\frac{1}{x_1} + \cdots + \frac{1}{x_n}} \tag{2.20}$$

Observación

- La media geométrica se suele utilizar en series de datos como tipos de interés anuales, inflación, etc., donde el valor de cada año tiene un efecto multiplicativo sobre el de los años anteriores. En todo caso, la media aritmética es la medida de posición central más utilizada.
- Lo más positivo de la media es que en su cálculo se utilizan todos los valores de la serie, por lo que no se pierde ninguna información.
- Sin embargo, presenta el problema de que su valor (tanto en el caso de la media aritmética como geométrica) se puede ver muy influido por valores

extremos, que se aparten en exceso del resto de la serie. Estos valores anómalos podrían condicionar en gran medida el valor de la media, perdiendo ésta representatividad.

d) Media cuadrática

Definición 2.16 *La **media cuadrática** $\overline{x}_c$, es la raíz cuadrada de la media aritmética de los cuadrados:*

$$\overline{x}_c = \sqrt{\frac{x_1^2 + \cdots + x_n^2}{n}} \tag{2.21}$$

Observación

- La media armónica se basa en todas las observaciones por lo que esta afectado por los valores extremos. Pero da a los valores grandes un peso menor que la que da la media geométrica; mientras que a los valores pequeños, le da un peso mayor que el que da la media aritmética y la media geométrica.
- La media armónica no esta definida, si alguno de los valores es cero.
- Se aplica cuando se presenta una relación inversa entre las variables implícitas, como por ejemplo, entre la productividad y el tiempo.

Relación entre medias aritméticas, geométricas y armónica

La media geométrica de una serie de números positivos X1,X2,…..Xn es menor o igual que su media aritmética pero es mayor o igual que su media armónica.

$$\text{H} \leq \text{G} \leq \overline{x} \tag{2.22}$$

2.5.1.1.2 La mediana

Definición 2.17 *La **Mediana** M_{ed}, es un valor único de un conjunto de datos que mide al elemento central en los datos. Este único elemento de los datos ordenados, es el más cercano a la mitad, o el más central en el conjunto de números. La mitad de los elementos quedan por encima de ese punto, y la otra mitad por debajo de él.*

- **Para datos no agrupados**

Sean X_1, X_2, X_3, X_n datos ordenados por magnitud creciente o decreciente y n el número impar de datos, la mediana (M_{ed}) es el dato situado en el centro, es decir:

$$M_{ed} = x_{(n+1)/2}\ , \qquad \text{para } n \text{ impar} \qquad (2.23)$$

si n es par, la mediana es el promedio de los números centrales, es decir:

$$M_{ed} = \frac{x_{n/2} + x_{n/2+1}}{2}\ , \qquad \text{para } n \text{ par} \qquad (2.24)$$

Ejemplo 2.8

Para el conjunto de datos ordenados: 2, 5, 8, 14, 21, la mediana es:
$M_{ed} = 8$ usando la ecuación (2.23)

Ejemplo 2.9

Para el conjunto de datos ordenados: 2, 5, 8, 14, 21, 32 la mediana es:
$M_{ed} = (8+14)/2 = 11$ usando la ecuación (2.24)

- **Para datos agrupados**

En el caso de datos agrupados, las clases vienen dadas por intervalos, y aquí la fórmula de la mediana se complica un poco más (pero no demasiado): Sea $(l_{i-1}, l_i]$ el intervalo donde se ha encontrado que por debajo están el **50%** de las observaciones. Entonces se obtiene la mediana a partir de las frecuencias absolutas acumuladas, mediante interpolación lineal (teorema de Thales) como sigue (Figura 2.13):

$$\frac{CC'}{AC} = \frac{BB'}{AB} \quad \Rightarrow \quad \frac{n_i}{a_i} = \frac{\frac{n}{2} - N_{i-1}}{M_{ed} - l_{i-1}}$$

$$\Rightarrow \quad M_{ed} = l_{i-1} + \frac{\frac{n}{2} - N_{i-1}}{n_i} \bullet a_i \qquad (2.25)$$

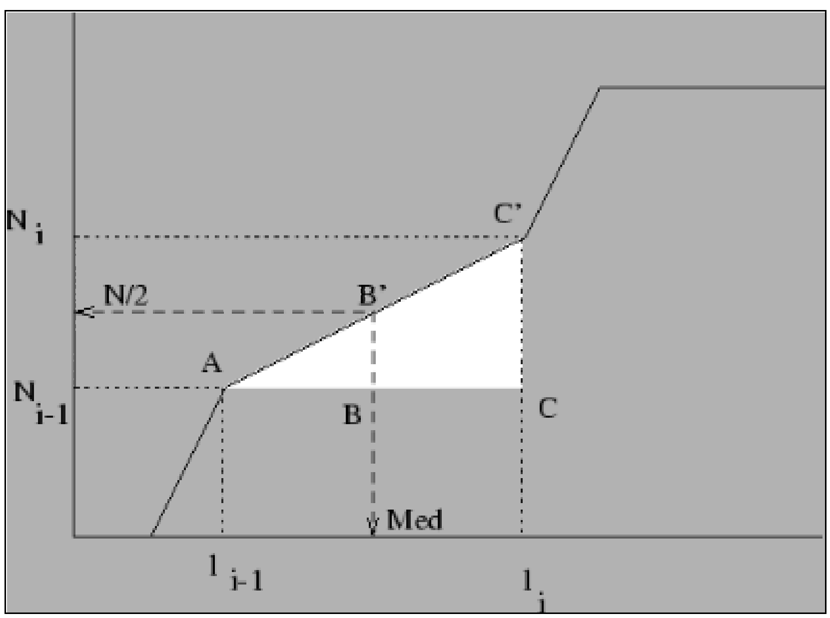


Figura 2.13 Cálculo geométrico de la mediana

Observación

Entre las propiedades de la mediana, Se destacan las siguientes:

- Como medida descriptiva, tiene la ventaja de no estar afectada por las observaciones extremas, ya que no depende de los valores que toma la variable, sino del orden de las mismas. Por ello es adecuado su uso en distribuciones asimétricas.
- Es de cálculo rápido y de interpretación sencilla.
- A diferencia de la media, la mediana de una variable discreta es siempre un valor de la variable que se estudia (ej. La mediana de una variable "*número de hijos"* toma siempre valores enteros).
- Si una población está formada por 2 subpoblaciones de medianas M_{ed1} y M_{ed2}, sólo se puede afirmar que la mediana, M_{ed}, de la población está comprendida entre M_{ed1} y M_{ed2}

$$\mathbf{M}_{ed1} \leq \mathbf{M}_{ed} \leq \mathbf{M}_{ed2}$$

- El mayor defecto de la mediana es que tiene unas propiedades matemáticas complicadas, lo que hace que sea muy difícil de utilizar en *inferencia estadística*.
- Es función de los intervalos escogidos.
- Puede ser calculada aunque el intervalo inferior o el superior no tenga límites.
- La suma de las diferencias de los valores absolutos de n puntuaciones respecto a su mediana es menor o igual que cualquier otro valor. Este es el equivalente al teorema de König (Proposición 2.5) con respecto a la media, pero donde se considera como medida de dispersión a:

$$\sum_{i=1}^{n} \left| x_i - M_{ed} \right|$$

Ejemplo 2.10

Sea X una variable discreta que ha presentado sobre una muestra las modalidades

$$X \rightarrow\ 2,5,7,9,12 \Rightarrow\ \bar{x} = 7, \quad M_{ed} = 7$$

Si se cambia la última observación por otra anormalmente grande, esto no afecta a la mediana, pero si a la media:

$$X \rightarrow\ 2,5,7,9,125 \Rightarrow\ \bar{x} = 29,6 \quad M_{ed} = 7$$

En este caso la media no es un posible valor de la variable (discreta), y se ha visto muy afectada por la observación extrema. Este no ha sido el caso para la mediana.

Ejemplo 2.11

Obtener la media aritmética y la mediana en la distribución adjunta. Determinar gráficamente cuál de los dos promedios es más significativo.

l_{i-1} - l_i	n_i
0 - 10	60
10 - 20	80
20 - 30	30
30 - 100	20
100 - 500	10

Tabla 2.20 Tabla de distribución de frecuencias absolutas

Solución:

l_{i-1} - l_i	n_i	a_i	x_i	$x_i\, n_i$	N_i	$n_i^{'}$
0 - 10	60	10	5	300	60	60
10 - 20	80	10	15	1200	140	80
20 - 30	30	10	25	750	170	30
30 - 100	20	70	65	1300	190	2.9
100 - 500	10	400	300	3000	200	0.25
	n=200			$\sum x_i n_i$ = 6.550		

Tabla 2.21 Tabla de frecuencias absolutas y acumuladas

La media aritmética es:

$$\bar{x} = \frac{1}{n}\sum x_i = \frac{6.550}{200} = 32,75$$

La primera frecuencia absoluta acumulada que supera el valor $n/2$=100 es N_i=140. Por ello el intervalo mediano es [10;20). Así:

$$M_{ed} = l_{i-1} + \frac{n/2 - N_{i-1}}{n_i} \cdot a_i = 10 + \frac{100-60}{80} \times 10 = 15$$

Para ver la representatividad de ambos promedios, se realiza el histograma de la Figura 2.14, y se observa que dada la forma de la distribución, la mediana es más representativa que la media.

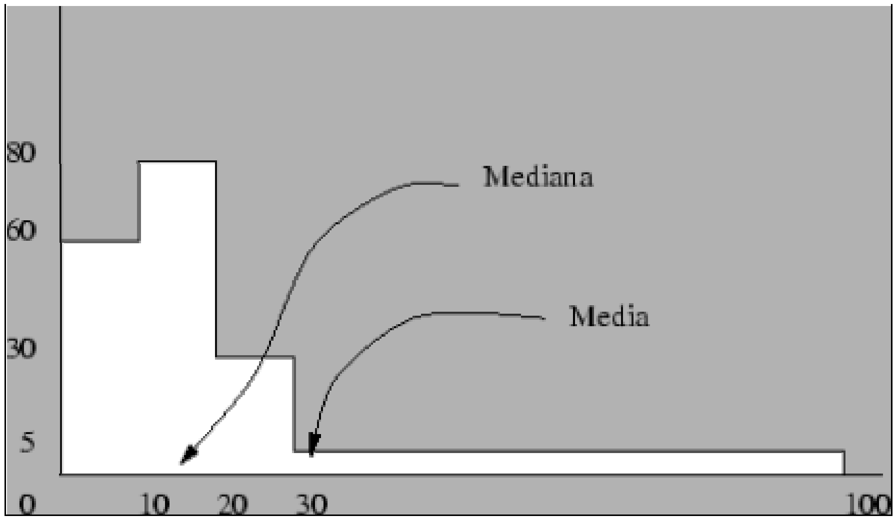


Figura 2.14 Para esta distribución de frecuencias es más representativo usar como estadístico de tendencia central la mediana que la media.

Observación

La mediana se emplea en lugar de la media aritmética cuando esta se ve influenciada por valores extremos. Cuando en el grupo de datos existen unos pocos valores muy altos o muy bajos, o ambos a la vez. La debilidad de la media aritmética, es esa sensibilidad a los valores extremos, que la corren mucho de la posición centralista que debería tener.

2.5.1.1.3 La moda

Definición 2.18 *Se llamará* ***moda*** *a cualquier máximo relativo de la distribución de frecuencias, es decir, cualquier valor de la variable que posea una frecuencia mayor que su anterior y su posterior.*

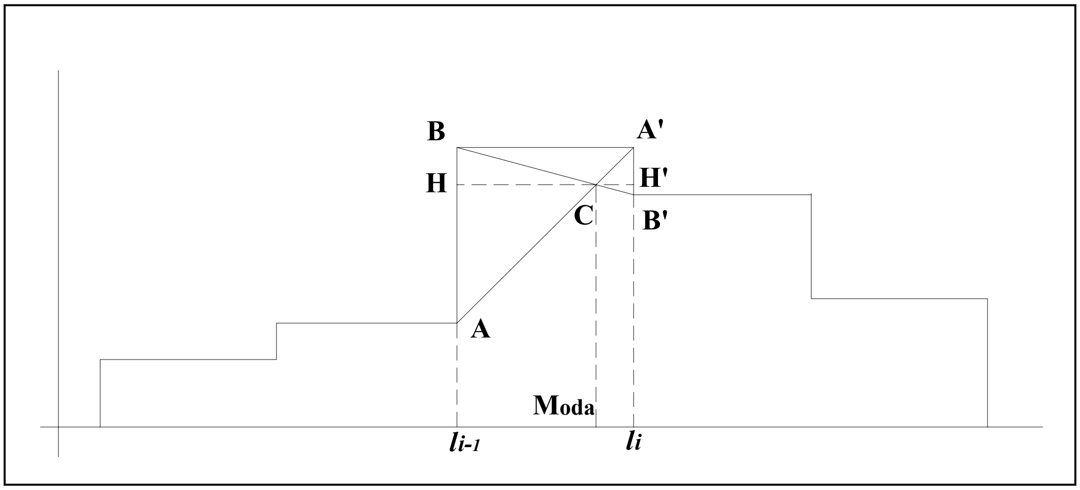


Figura 2.15 Cálculo geométrico de la moda

En el caso de variables continuas es más correcto hablar de *intervalos modales*. Una vez que este intervalo, $(l_{i-1}, l_i]$, se ha obtenido, se utiliza la siguiente fórmula para calcular la moda, que está motivada en la Figura 2.15:

$$\frac{HC}{AB} = \frac{H'C}{A'B'} \qquad = \frac{HC + H'C}{AB + A'B'}$$

$$\Rightarrow \frac{M_{oda} - l_{i-1}}{n_i - n_{i-1}} = \frac{a_i}{(n_i - n_{i-1}) + (n_i - n_{i+1})}$$

$$\Rightarrow M_{oda} = l_{i-1} + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} a_i \tag{2.26}$$

Observación

De la moda se destacan las siguientes propiedades:

- Puede no ser única.
- Es función de los intervalos elegidos a través de su amplitud, número y límites de los mismos.

- Aunque el primero o el último de los intervalos no posean extremos inferior o superior respectivamente, la moda puede ser calculada.
- Se lo usa en lugar de la media o mediana, cuando se desea señalar el valor máximo relativo.
- Es más ambiguo que la mediana, muy inestable y no permite tratamientos algebraicos sencillos.
- La moda no queda influenciada por valores extremos como la media y es el mas sencillo de comprender. A su antitesis, se lo llama antimoda, o sea el valor menos frecuente.

2.5.1.2 Medidas de posición no central

Las medidas de posición no centrales permiten conocer otros puntos característicos de la distribución que no son los valores centrales. Entre otros indicadores, se suelen utilizar una serie de valores que dividen la muestra en tramos iguales:

Definición 2.19 *Para una variable discreta, se define el* ***percentil de orden*** *k, como la observación,* ***P_k****, que deja por debajo de si el* ***k%*** *de la población. Esta definición hace recuerdo a la mediana, pues como consecuencia de la definición es evidente que*

$$M_{ed} = P_{50}$$

En el caso de una variable continua, el intervalo donde se encuentra ***$P_k \in (l_{i=1}, l_i)$***, se calcula buscando el que deja debajo de si al ***k%*** de las observaciones. Dentro de él, ***P_k*** se obtiene según la relación:

$$P_k = l_{i-1} + \frac{n\frac{k}{100} - N_{i-1}}{n_i} \cdot a_1 \qquad (2.27)$$

Por su propia naturaleza, el percentil puede estar situado en cualquier lugar de la distribución, por lo que no puede considerársele como una medida de tendencia central.

Definición 2.20 *Los **cuartiles**, Q_l, son un caso particular de los percentiles. Hay 3, y se definen como:*

$$\boldsymbol{Q_1 = P_{25}}$$

$$\boldsymbol{Q_2 = P_{50} \qquad = M_{ed}}$$

$$\boldsymbol{Q_3 = P_{75}}$$

Definición 2.21 *De forma análoga se definen los **deciles** como los valores de la variable que dividen a las observaciones en 10 grupos de igual tamaño. Más precisamente, se define $D_1, D_2, \ldots, D_9$ como:*

$$\boldsymbol{D_i = P_{10\,i} \qquad i = 1,\ldots,9}$$

Ejemplo 2.12

Dada la siguiente distribución en el número de hijos de cien familias, calcular sus cuartiles.

x_i	n_i	N_i
0	14	14
1	10	24
2	15	39
3	26	65
4	20	85
5	15	100
	n=100	

Tabla 2.22 Distribución de frecuencias absolutas y acumuladas

Dada la siguiente distribución en el número de hijos de cien familias, calcular sus cuartiles.

x_i	n_i	N_i
0	14	14
1	10	24
2	15	39
3	26	65
4	20	85
5	15	100
	n=100	

Tabla 2.23 Distribución de frecuencias absolutas y acumuladas

Solución:

1. Primer cuartil:

   $n/4 = 25$; Primera $N_i > n / 4 = 39$; luego $Q_1 = 2$.

2. Segundo cuartil:

   $2n/4 = 50$; Primera $N_i > 2n / 4 = 65$; luego $Q_2 = 3$

3. Tercer cuartil:

   $3n/4 = 75$; Primera $N_i > 3n / 4 =85$; luego $Q_3 = 4$

Ejemplo 2.13

Calcular los cuartiles en la siguiente distribución de una variable continua:

$l_{i\text{-}1}$ - l_i	n_i	N_i
0 - 1	10	10
1 - 2	12	22
2 - 3	12	34
3 - 4	10	44
4 - 5	7	51
	n=51	

Tabla 2.24 Distribución de frecuencias absolutas y acumuladas

Solución:

1. Primer cuartil

$\frac{N}{4} = 12,75$; Primera $N_i > n/4 = 22$; la línea $\boldsymbol{i}$ es la del intervalo $\left[1;2\right)$

$$Q_1 = l_{i-1} + \frac{\frac{n}{4} - N_{i-1}}{n_i} a_i = 1 + \frac{12,75 - 10}{12} \times 1 = 1,23$$

2. Segundo cuartil:

$\frac{2n}{4} = 25,5$; Primera $N_i > 2\,n/4 = 34$; La línea $\boldsymbol{i}$ es la del intervalo $\left[2;3\right)$

$$Q_2 = l_{i-1} + \frac{\frac{2n}{4} - N_{i-1}}{n_i} a_i = 2 + \frac{25,5 - 22}{12} \times 1 = 2,29$$

3. Tercer cuartil

$\frac{3n}{4} = 38,25$; Primera $N_i > 3n/4 = 44$; la línea ***i*** es la del intervalo $[3;4)$

$$Q_3 = l_{i-1} + \frac{\frac{3n}{4} - N_{i-1}}{n_i} a_i = 3 + \frac{38,25 - 34}{10} \times 1 = 3,445$$

Ejemplo 2.14

Han sido ordenados los pesos de 21 personas en la siguiente tabla:

Intervalos l_{i-1} -- l_i	***f.a.*** n_i
38 -- 45	3
45 -- 52	2
52 -- 59	7
59 -- 66	3
66 -- 73	6
	21

Tabla 2.25 Tabla de frecuencias acumuladas

Encontrar aquellos valores que dividen a los datos en 4 partes con el mismo número de observaciones.

Solución:

Las cantidades que se busca son los tres cuartiles: Q_1, Q_2 y Q_3. Para calcularlos, se añade a la tabla las columnas con las frecuencias acumuladas, para localizar qué intervalos son los que contienen a los cuartiles buscados:

l_{i-1} -- l_i	n_i	N_i	
38 -- 45	3	3	
45 -- 52	2	5	
52 -- 59	7	12	$\ni Q_1, Q_2$
59 -- 66	3	15	
66 -- 73	6	21	$\ni Q_3$
	21		

Tabla 2.26 Tabla de frecuencias absolutas y acumuladas

Q_1 y Q_2 se encuentran en el intervalo 52--59, ya que N_3=12 es la primera f.a.a. que supera a $21 \cdot 1/4$ y $21 \cdot 2/4$.

Q_3 está en 66--73, pues N_5=21 es el primer N_i mayor que $21 \cdot 3/4$.

Así se tiene que:

$$\frac{1}{4} \cdot 21 = 5,25 \Rightarrow \quad i = 3 \Rightarrow Q_1 \quad = l_{i-1} + \frac{\frac{1}{4}n - N_{i-1}}{n_i} \cdot a_i$$

$$= 52 + \frac{10,5 - 5}{7} \cdot 7 = 57,5$$

$$\frac{3}{4}\cdot 21 = 15,75 \Rightarrow \quad i = 5 \Rightarrow Q_3 \quad = l_{i-1} + \frac{\frac{3}{4}n - N_{i-1}}{n_i}\cdot a_i$$

$$= 66 + \frac{15,75-15}{6}\cdot 7 = 66,875$$

Obsérvese que $\boldsymbol{Q_2 = M_{ed}}$. Esto es lógico, ya que la mediana divide a la distribución en dos partes con el mismo número de observaciones, y $\boldsymbol{Q_2}$, hace lo mismo, pues es deja a dos cuartos de los datos por arriba y otros dos cuartos por abajo.

Ejemplo 2.15

La distribución de una variable tiene por polígono acumulativo de frecuencias el de la figura 2.16. Si el número total de observaciones es 50:

1. Elaborar una tabla estadística con los siguientes elementos: intervalos, marcas de clase, frecuencia absoluta, frecuencia absoluta acumulada, frecuencia relativa y frecuencia relativa acumulada.

2. Cuántas observaciones tuvieron un valor inferior a 10, cuántas inferior a 8 y cuántas fueron superior a 11.

3. Calcule las modas.

4. Determine los cuartiles.

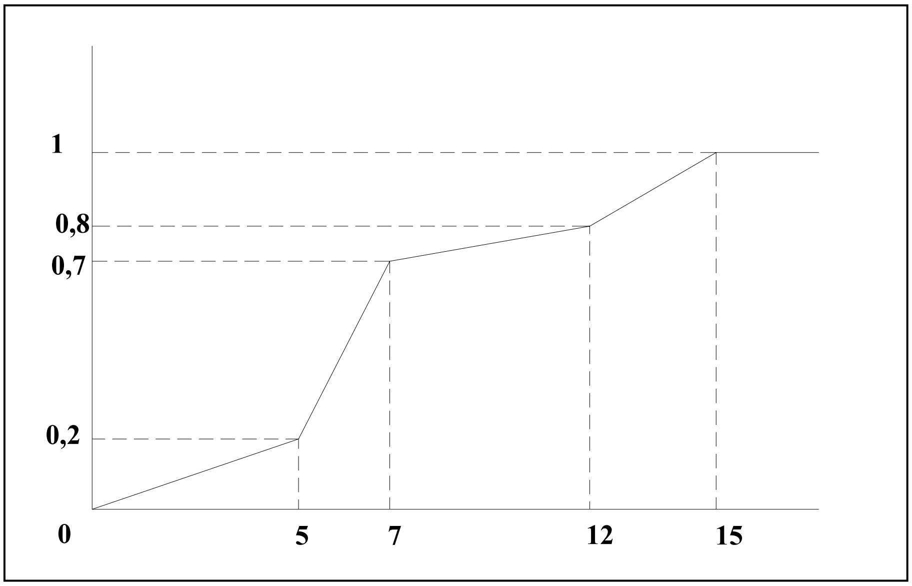


Figura 2.16 Diagrama acumulado de frecuencias relativas.

Solución:

1. En la siguiente tabla se proporciona la información pedida y algunos cálculos auxiliares que nos permitirán responder a otras cuestiones.

Intervalos	n_i	N_i	f_i	F_i	x_i	a_i	n_i'
0 - 5	10	10	0.2	0.3	2.5	5	2
5 - 7	25	35	0.5	0.7	6	2	12,5
7 - 12	5	40	0.1	0.8	9.5	5	1
12 - 15	10	50	0.2	1	13.5	7	3.33

Tabla 2.27 Datos encontrados a partir del diagrama acumulado de frecuencias relativas

2. Se calcula el número de observaciones pedido:

7 a 12 ------- 5 $\Leftrightarrow$ 5 ------- 5 $\Rightarrow$ $x = \dfrac{3 \times 5}{5} = 3$

7 a 10 ------- x 3 ------- x

10 + 25+3 = 38 observaciones tomaron un valor inferior a 10

7 a 12 ------- 5 $\Leftrightarrow$ 5 ------- 5 $\Rightarrow$ $x = \dfrac{1 \times 5}{5} = 1$

7 a 8 ------- x 1 ------- x

10 + 25+1 = 36 observaciones tomaron un valor inferior a 8

7 a 12 ------- 5 $\Leftrightarrow$ 5 ------- 5 $\Rightarrow$ $x = \dfrac{4 \times 5}{5} = 4$

7 a 11 ------- x 4 ------- x

50 -(10 + 25+4) = 50-39=11 observaciones tomaron un valor superior a 11

3. Hay dos modas. Calculando la más representativa:

$$M_{oda} = l_{i-1} + \frac{n_{i-1}'}{n_{i-1}' + n_{i+1}'} \cdot a_i = 5 + \frac{1}{2+1} \cdot 2 = 5,66$$

4. Cuartiles:

$$Q_1 = l_{i-1} + \frac{n/4 - N_{i-1}}{n_i} \cdot a_i = 5 + \frac{12,5-10}{25} \cdot 2 = 5,2$$

$$Q_2 = l_{i-1} + \frac{2n/4 - N_{i-1}}{n_i} \cdot a_i = 5 + \frac{25-10}{25} \cdot 2 = 6,2$$

$$Q_3 = l_{i-1} + \frac{3n/4 - N_{i-1}}{n_i} \cdot a_i = 7 + \frac{37,5-35}{5} \cdot 5 = 9,5$$

2.5.2 Medidas de dispersión

Se denominan así a las medidas que cuantifican la separación de los datos entre sí, respecto de un punto de referencia central, como la media aritmética. Cuanto menor es la dispersión, tanto mayor será la precisión del sistema de medición.

Cuando se tiene una muestra de datos obtenida de una población cualquiera, es importante determinar sus medidas de tendencia central así como también es básico el determinar que tan dispersos están los datos en la muestra, por lo que se hace necesario determinar su rango, la varianza, la desviación estándar, etc., ya que una excesiva variabilidad o dispersión en los datos indica la inestabilidad del proceso en análisis en la mayoría de los casos.

2.5.2.1 Rango

Definición 2.22 *El **rango** es la diferencia entre el valor mayor y el valor menor encontrado en el conjunto de elementos, también se le denomina recorrido ya que indica entre que valores hace su recorrido la variable de interés; y se determina de la siguiente manera:*

$$\boldsymbol{R = V_M - Vm} \tag{2.28}$$

Donde:

$\boldsymbol{R}$: rango o recorrido

$\mathbf{V_M}$: valor mayor en la muestra

$\mathbf{V_m}$:valor menor en la muestra

Rango Intercuartilico. Se define como la diferencia entre el tercer y primer cuartel (o también entre la diferencia entre los percentiles 75^{avo} y 25^{avo}). Es decir :

Dada una serie de datos

Rango intercuartilico = $Q = Q_3 + Q_1$ ò $Q = P_{75} - P_{25}$

Rango interdecil.- es la diferencia entre los percentiles 90^{avo} y décimo. Es decir

$$\boldsymbol{RID = P_{90} - P_{10}} \tag{2.29}$$

Desviación del cuartil.- la mitad del rango intercuartilico es una medida llamada desviación del cuartil. Es decir.

$$Q = \frac{Q_3 - Q_1}{2} \tag{2.30}$$

La desviación del cuartil mide el recorrido de un cuarto de los datos. Es representativo de la dispersión de los datos, ya que calcula, tomando el promedio de la mitad de los elementos del medio en lugar de escoger uno de los cuartos.

2.5.2.2 Varianza

Definición 2.23 *La **Varianza** mide la distancia existente entre los valores de la serie y la media. Se calcula como sumatoria de las diferencias al cuadrado entre cada valor y la media, multiplicadas por el número de veces que se ha repetido cada valor. La sumatoria obtenida se divide por el tamaño del conjunto.*

- **Para datos no agrupados**

  - Varianza Poblacional $$\sigma^2 = \frac{\sum_{i=1}^{n}\left(xi - \bar{x}\right)^2}{\mathrm{N}} \quad (2.31)$$

  - Varianza muestral $$S^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n} \quad (2.32)$$

Donde:

Xi : es la i-esima medición o valor

n : es el tamaño de la muestra

N : es el tamaño de la población

$\overline{x}$: es la media

Para el calculo es útil expresar la sumatoria de la siguiente manera :

$$\sum(x_i - \overline{x})^2 = \sum(x_i^2 - 2\overline{x}x_i + \overline{x}^2)$$

$$= \sum x_i^2 - 2\overline{x}\sum x_i + n\overline{x}^2 \quad (2.33)$$

pero:

$$\sum x_i = n\frac{\sum x_i}{n} = n\overline{x} \quad (2.34)$$

Luego sustituyendo la ecuación 2.34 en la ecuación 2.33

$$\sum(x_i - \overline{x})^2 = \sum x_i^2 - 2n\overline{x}^2 + n\overline{x}^2$$

$$\sum(x_i - \overline{x})^2 = \sum x_i^2 - n\overline{x}^2$$

Sustituyendo, se tiene:

$$S^2 = \frac{1}{n}\sum x_i^{\,2} - \overline{x}^2 \qquad (2.35)$$

Al ser mínima la sumatoria del cuadrado de los desvíos, como se vio para la media, entonces la varianza respecto de la media aritmética, será un mínimo. Es decir, la medida de la dispersión de los valores alrededor de la media usando la varianza, será menor que si se toma respecto de otra medida de posición cualquiera.

- **Para datos agrupados**

$$S^2 = \frac{\sum_{i=1}^{m}(c_i - \overline{c}_i)^2 . n_i}{n} \qquad (2.36)$$

Donde :

n_j : es la frecuencia de clase ***i***

c_j : su respectiva marca de clase

$\overline{c}$: media aritmética de las marcas de clase

n : número total de datos

m : número de intervalos de clase

Para su calculo, la ecuación se puede expresar como:

$$S^2 = \frac{1}{n}\sum_{i=1}^{m} c_i^{\,2} n_i - \overline{c}^2) \qquad (2.37)$$

2.5.2.3 Desviación media, Dm

Definición 2.24 *Se define la* ***desviación media*** *como la media de las diferencias en valor absoluto de los valores de la variable a la media, es decir, si tenemos un conjunto de n observaciones,* $x_1, \ldots, x_n$*, entonces*

$$D_m = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \overline{x}\right| \tag{2.38}$$

Si los datos están agrupados en una tabla estadística es más sencillo usar la relación

$$D_m = \frac{1}{n}\sum_{i=1}^{k}\left|x_i - \overline{x}\right| n_i \tag{2.39}$$

Como se observa, la desviación media guarda las mismas dimensiones que las observaciones. La suma de valores absolutos es relativamente sencilla de calcular, pero esta simplicidad tiene un inconveniente: Desde el punto de vista geométrico, la distancia que induce la desviación media en el espacio de observaciones no es la *natural* (no permite definir ángulos entre dos conjuntos de observaciones). Esto hace que sea muy engorroso trabajar con ella a la hora de hacer inferencia a la población.

2.5.2.4 Desviación estándar típica

Si bien es conocido como expresar cuantitativamente la dispersión de un conjunto de observaciones, ocurre un inconveniente en cuanto a la interpretación de esta cantidad, ya que ella está dada en el cuadrado de la dimensión en que se expresa la característica, y en ocasiones trae confusión. Es conveniente, entonces contar con otro estadígrafo que basado en el valor de la varianza, sirva para dar una medida de la dispersión en la misma dimensión en que están los datos. Esta medida es la desviación típica o desviación estándar.

Definición 2.25 *Se puede definir a la* ***Desviación Estándar*** *como la desviación o diferencia promedio que existe entre cada dato del conjunto y la media aritmética del conjunto. Y se obtiene a partir de la varianza, sacándole raíz cuadrada.*

$$s = \sqrt{s^2} \qquad (2.40)$$

Observación

- El desvió estándar poblacional (σ) o maestral (s) se obtienen con la raíz cuadrada de las respectivas varianzas. Para poder conocer y calcular tanto la media como la varianza poblacionales, se necesita conocer la población completa, o en su defecto se pueden estimarse a partir de sus respectivas medias muestrales.
- Debe considerarse al desvió estándar como un estadígrafo y a la varianza como a otro. Aquí se desarrollan en conjunto para simplificar el texto, aprovechando la gran similitud de sus fórmulas. Pero, mientras el primero se puede comparar directamente con cualquiera de los valores medidos o de sus medidas de posición, la varianza no. A sí la media y el desvió se emplearan para construir intervalos donde se espera encontrar adentro al verdadero valor de la población, llamado a veces *parámetro.*

2.5.3 Tipificación

Se conoce por **tipificación** al proceso de restar la media y dividir por su desviación típica a una variable *X*. De este modo se obtiene una nueva variable

$$Z = \frac{X - \overline{x}}{S} \tag{2.41}$$

de media $\overline{z} = 0$ y desviación típica $\boldsymbol{S_z = 1}$, que se denomina **variable tipificada**. Esta nueva variable carece de unidades y permite hacer comparables dos medidas que en un principio no lo son, por aludir a conceptos diferentes. Así por ejemplo puede preguntarse si un elefante es más grueso que una hormiga determinada, cada uno en relación a su población. También es aplicable al caso en que se quieran comparar individuos semejantes de poblaciones diferentes. Por ejemplo si deseamos comparar el nivel académico de dos estudiantes de diferentes Universidades para la concesión de una beca de estudios, en principio sería injusto concederla directamente al que posea una nota media más elevada, ya que la dificultad para conseguir una buena calificación puede ser mucho mayor en un centro que en el otro, lo que limita las posibilidades de uno de los estudiante y favorece al otro. En este caso, lo más correcto es comparar las calificaciones de ambos estudiantes, pero tipificadas cada una de ellas por las medias y desviaciones típicas respectivas de las notas de los alumnos de cada Universidad.

2.5.4 Coeficiente de variación (Coeficiente de Pearson)

Se ha visto que las medidas de centralización y dispersión nos dan información sobre una muestra. Puede preguntarse si tiene sentido usar estas magnitudes para comparar dos poblaciones. Por ejemplo, si se pide comparar la dispersión de los pesos de las poblaciones de elefantes de dos circos diferentes, $\boldsymbol{S}$ dará información útil.

¿Pero qué ocurre si lo que se compara es la altura de unos elefantes con respecto a su peso? Tanto la media como la desviación típica, $\overline{x}$ y $\boldsymbol{S}$, se expresan en las mismas unidades que la variable. Por ejemplo, en la variable altura puede usarse como unidad de longitud el metro y en la variable peso, el kilogramo. Comparar una desviación (con respecto a la media) medida en metros con otra en kilogramos no tiene ningún sentido.

El problema no deriva sólo de que una de las medidas sea de longitud y la otra sea de masa. El mismo problema se plantea si se mide cierta cantidad, por ejemplo la masa, de dos poblaciones, pero con distintas unidades. Este es el caso en que se compara el peso en *toneladas* de una población de 100 elefantes con el correspondiente en *miligramos* de una población de 50 hormigas.

El problema no se resuelve tomando las mismas escalas para ambas poblaciones. Por ejemplo, se puede tratar de medir a las hormigas con las mismas unidades que los elefantes (toneladas). Si la ingeniería genética sorprendiera con alguna barbaridad, lo lógico es que la dispersión de la variable *peso de las hormigas* sea prácticamente nula (¡Aunque haya algunas que sean 1.000 veces mayores que otras!)

En los dos primeros casos mencionados anteriormente, el problema viene de la *dimensionalidad* de las variables, y en el tercero de la diferencia enorme entre las medias de ambas poblaciones.

Definición 2.26 *El **coeficiente de variación** es lo que permite evitar estos problemas, pues elimina la dimensionalidad de las variables y tiene en cuenta la proporción existente entre medias y desviación típica. Se define del siguiente modo:*

$$CV = \frac{S_x}{\overline{x}} \tag{2.42}$$

Basta dar una rápida mirada a la definición del coeficiente de variación, para ver que las siguientes consideraciones deben ser tenidas en cuenta:

- Sólo se debe calcular para variables con todos los valores positivos. Todo índice de variabilidad es esencialmente no negativo. Las observaciones pueden ser positivas o nulas, pero su variabilidad debe ser siempre positiva. De ahí que sólo se debe trabajar con variables positivas, para la que se tiene con seguridad que $\overline{x}$ **>0**.

- No es invariante ante cambios de origen. Es decir, si a los resultados de una medida se le suma una cantidad positiva, $b>0$, para tener $Y=X+b$, entonces $CV_Y < CV_X$, ya que la desviación típica no es sensible ante cambios de origen, pero si la media. Lo contrario ocurre si se resta ($b<0$).

$$CV_Y = \frac{S_Y}{\overline{y}} = \frac{S_X}{\overline{x}+b} < \frac{S_X}{\overline{x}} = CV_X$$

- Es invariante a cambios de escala. Si se multiplica X por una constante a, para obtener $\boldsymbol{Y = a\,X}$, entonces

$$CV_Y = \frac{S_Y}{\overline{y}} = \frac{S_{aX}}{a\overline{x}} = \frac{aS_X}{a\overline{x}} = CV_X$$

Observación

Es importante destacar que los *coeficientes de variación* sirven para comparar las variabilidades de dos conjuntos de valores (muestras o poblaciones), mientras que si se desea comparar a dos *individuos* de cada uno de esos conjuntos, es necesario usar los *valores tipificados*.

Ejemplo 2.16

Dada la distribución de horas trabajadas en una empresa con 100 trabajadores, obtener:

1. La variable tipificada Z.
2. Valores de la media y varianza de Z.
3. Coeficiente de variación de Z.

Dada la distribución de horas trabajadas en una empresa con 100 trabajadores, obtener:

1. La variable tipificada *Z*.
2. Valores de la media y varianza de *Z*.
3. Coeficiente de variación de *Z*.

Horas trabajadas	**Num. empleados**
0 -- 4	47
4 -- 10	32
10 -- 20	17
20 -- 40	4
	100

Tabla 2.28 Tabla de datos de horas trabajadores y números de empleados

Solución:

Para calcular la variable tipificada

$$Z = \frac{X - \overline{x}}{S}$$

Se parte de los datos del enunciado. Será necesario calcular en primer lugar la media y desviación típica de la variable original (*X*= años).

$l_{i\text{-}1}$ -- l_i	x_i	n_i	$x_i\, n_i$	$x_i^2\, n_i$
0 -- 4	2	47	94	188
4 -- 10	7	32	224	1.568
10 -- 20	15	17	255	3.825
20 -- 40	30	4	120	3.6
		n=100	693	9.181

Tabla 2.29 Tabla para el calculo de la media aritmética y la desviación típica

$$\overline{x} = \frac{693}{100} = 6,93$$

$$S_x^2 = \frac{9.181}{100} - 6,93^2 = 43,78 \quad \textit{años al cuadrado}$$

$$S_x = \sqrt{43,78} = 6,6 \quad \textit{años}$$

A partir de estos valores se puede calcular los valores tipificados para las marcas de clase de cada intervalo y construir su distribución de frecuencias:

$$z_1 = \frac{15 - 6,93}{6,6} = 1,22$$

$$z_2 = \frac{7 - 6,93}{6,6} = 0,011$$

$$z_3 = \frac{15 - 6,93}{6,6} = 1,22$$

$$z_4 = \frac{30 - 6,93}{6,6} = 3,486$$

z_i	n_i	$z_i\ n_i$	$z_i^2\ n_i$
-0.745	47	-35.015	26.086
0.011	32	0.352	0.004
1.22	17	20.72	25.303
3.486	4	13.944	48.609
	n=100	0.021	100.002

Tabla 2.30 Tabla de distribución de frecuencias para la variable tipificada

$$\overline{z} = \frac{0,021}{100} \approx 0$$

$$S_Z^2 = \frac{100,02}{100} - 0^2 \approx 1$$

$$S_X = \sqrt{1} = 1$$

A pesar de que no se debe calcular el coeficiente de variación sobre variables que presenten valores negativos (y Z los presenta), se puede calcular con objeto de ilustrar el porqué:

$$CV = \frac{S_Z}{\overline{z}} = \frac{1}{0} = \infty$$

Es decir, el coeficiente de variación no debe usarse nunca con variables tipificadas.

2.5.5 Medidas de forma

Con frecuencia al analizar una serie de datos es necesario el conocer si éstos se distribuyen de forma simétrica con respecto a un valor central, o si bien la gráfica que representa la distribución de frecuencias es *de una forma diferente del lado derecho que del lado izquierdo*.

Si la simetría ha sido determinada, puede ser necesario conocer si la curva es más o menos *apuntada* (larga y estrecha). Este apuntamiento se tendrá que medir comparando a cierta distribución de frecuencias que consideramos *normal* (no por casualidad es éste el nombre que recibe la distribución de referencia).
Estas características se llaman Asimetría y Apuntamiento.

2.5.5.1 Medidas de asimetría

Una distribución es simétrica cuando su mediana, su moda y su media aritmética coincidan.

Una distribución es *asimétrica a la derecha* si las frecuencias (absolutas o relativas) descienden más lentamente por la derecha que por la izquierda.
Si las frecuencias descienden más lentamente por la izquierda que por la derecha se dirá que la distribución es *asimétrica a la izquierda.*

Existen varias medidas de la asimetría de una distribución de frecuencias. Aquí estudiaremos dos de ellas.

a. **Coeficiente de Asimetría de Pearson**

Se define como:

$$A_p = \frac{\overline{x} - M_{oda}}{S} \tag{2.43}$$

Siendo cero cuando la distribución es simétrica, positivo cuando existe asimetría a la derecha y negativo cuando existe asimetría a la izquierda.

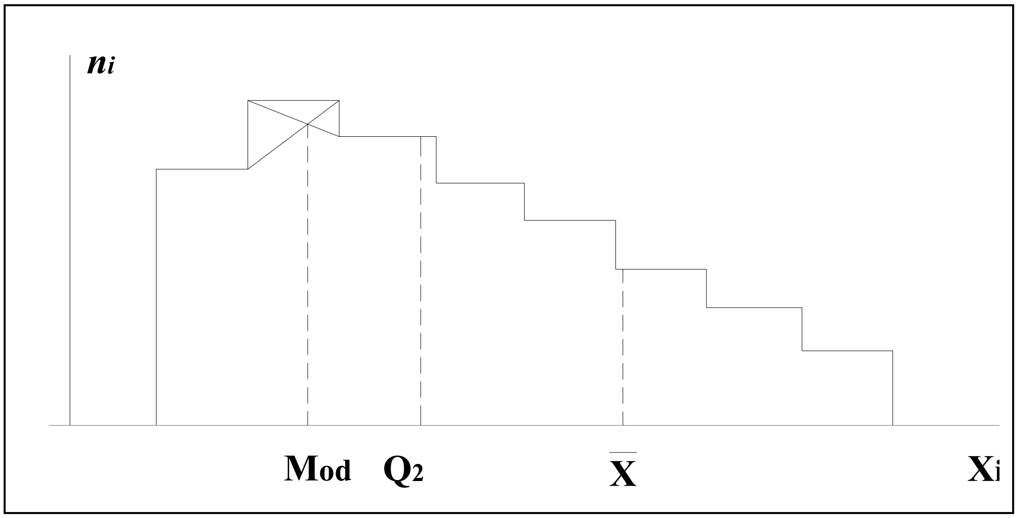


Figura 2.17 Diferencias importantes entre la media y la moda o la media y la mediana indican asimetría

b. **Coeficiente de Asimetría de Fisher**

Se define como:

$$A_f = \frac{\sum_{i=1}^{k} (x_i - \overline{x})^3 \cdot n_i}{n \cdot S^3} \tag{2.44}$$

siendo x_i los valores de la variable o las marcas de clase y S= $\sqrt{S^2}$.

La interpretación del *coeficiente de Fisher* es la misma que la del *coeficiente de Pearson*: si la distribución es simétrica vale cero, siendo positivo o negativo cuando exista *asimetría a la derecha o izquierda respectivamente.*

2.5.5.2 Medidas de apuntamiento o Curtosis

Definición 2.27 ***Curtosis*** *es la medida de deformación vertical de una distribución de frecuencias, es decir, la medida de apuntamiento o achatamiento de una distribución***.**

La idea de apuntamiento de una distribución surgió de la comparación de la frecuencia de los valores centrales de una distribución con la frecuencia de dichos valores en la distribución normal o gaussiana (será introducida posteriormente) que le corresponde. Entonces, el apuntamiento de una distribución de frecuencias indica la mayor o menor altura del máximo central, con respecto a la altura de la curva normal con media y desviación típica igual que la distribución que se estudia.

Definición 2.28 *Se define el coeficiente de aplastamiento de* ***Fisher*** *como:*

$$\gamma_2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^4 / n}{S^4} - 3 \tag{2.45}$$

$$\gamma_2 = \frac{m^4}{S^4} - 3 \tag{2.46}$$

donde m^4 es el momento empírico de cuarto orden. Es éste un coeficiente adimensional, invariante ante cambios de escala y de origen. Sirve para medir si una distribución de frecuencias es muy apuntada o no. Para decir si la distribución es larga y estrecha, hay que tener un patrón de referencia. El patrón de referencia es la *distribución normal o gaussiana* para la que se tiene

$$\frac{m^4}{S^4} = 3 \Rightarrow \gamma_2 = 0$$

De este modo, atendiendo a γ_2, se clasifican las distribuciones de frecuencias en:

Leptocúrtica:

Cuando $\gamma_2 > 0$, o sea, si la distribución de frecuencias es más apuntada que la normal;

Mesocúrtica:

Cuando $\gamma_2 = 0$, es decir, cuando la distribución de frecuencias es tan apuntada como la normal;

Platicúrtica:

Cuando $\gamma_2 < 0$, o sea, si la distribución de frecuencias es menos apuntada que la normal;

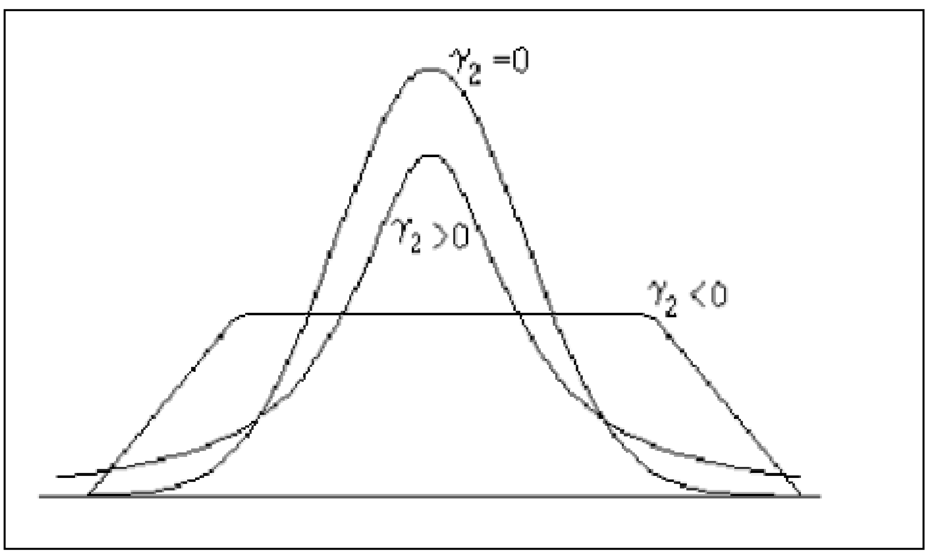


Figura 2.18 Apuntamiento de distribuciones de frecuencias

2.5.6 Medidas de Concentración

El concepto de ***concentración*** fue introducido por el estadístico italiano CORRADO GINI, a propósito de la distribución de salarios y rentas. En general se aplica a la descripción de unidades económicas según el tamaño (empresas; según ventas, el número de asalariados, de producción, etc.), pero también se utiliza para medir el grado de concentración o desigualdad de cualquier distribución.

2.5.6.1 Curva de concentración

Gráficamente puede ser representada esta característica mediante la denominada **curva de concentración o curva de LORENZ**, que representa en el eje de las absisas el porcentaje acumulado de elementos y en las ordenadas el porcentaje acumulado de los valores: variable x frecuencia absoluta, como se observa en la Figura 2.19

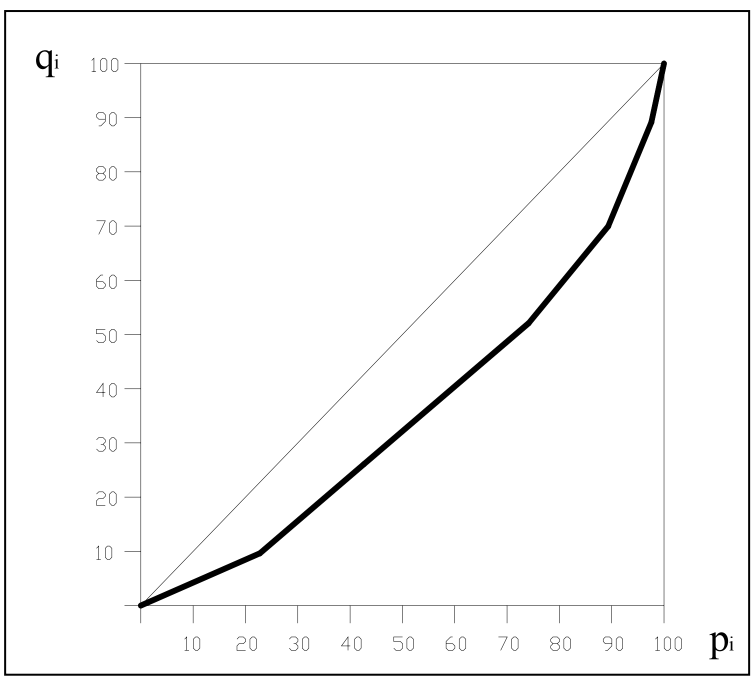


Figura 2.19 Curva de Lorenz

2.5.6.2 Índice de concentración

Para medir el nivel de concentración de una distribución de frecuencia se pueden utilizar distintos indicadores, entre ellos el **Índice de Gini** "G", éste es el doble del área comprendida entre la curva de concentración y la diagonal. El valor de **G** está comprendido entre 0 y 1 : pues el área vale cero si la curva coincide con la diagonal y un medio si la curva coinciden con los lados del cuadrado; es decir el área es la del triangulo formado por los dos lados del cuadrado y su diagonal.

Su cálculo puede hacerse por comparación de áreas o bien gráficamente (por métodos de valoración grafica de áreas) o bien aplicando la formula:

$$G = \frac{\sum_{i=1}^{n-1} (p_i - q_i)}{\sum_{i=1}^{n-1} p_i} \qquad (2.47)$$

la sumatoria es sólo hasta *n*-1, porque siempre $p_n - q_n = 100\% - 100\% = 0$.
Conociéndose que p_i mide el porcentaje de individuos que presentan un valor igual o inferior al de x_i.

$$p_i = \frac{n_1 + n_2 + n_3 + \ldots + n_i}{n} \times 100 \qquad (2.48)$$

mientras que q_i se calcula aplicando la siguiente formula:

$$q_i = \frac{x_i n_1 + x_2 n_2 + x_3 n_3 + \ldots + x_i n_i}{x_i n_1 + x_2 n_2 + x_3 n_3 + \ldots + x_n n_n} \times 100 \qquad (2.49)$$

El **Indice Gini** (G) puede tomar valores entre 0 y 1:

G = 0 : concentración mínima. La muestra está uniformemente repartida a lo largo de todo su rango.

G = 1 : concentración máxima. Un sólo valor de la muestra acumula el 100% de los resultados.

BIBLIOGRAFÍA

Villón Béjar, Máximo, Hidrológica Estadística; Lima Peru

Ríus Días , Francisca, Barón López, Francisco, Sánches Font, Elisa, Parras Guijosa, Luis, Bioestadística: métodos y aplicaciones, Manual de la Universidad de Málaga.

Moya Rufino – Saravia A. Gregorio. Estadística Descriptiva; Editorial Limusa

Azzimonti Renzo, JC; Apuntes de la Catedra de Estadística, UNAM, 1999/2000

CAPÍTULO III

ESTADÍSTICA DESCRIPTIVA BIDIMENSIONAL

CAPÍTULO III

ESTADÍSTICA DESCRIPTIVA BIDIMENSIONAL

3.1 INTRODUCCIÓN

En lo estudiado anteriormente se ha podido aprender cómo a partir de la gran cantidad de datos que describen una muestra mediante una variable, *X*, se representan gráficamente los mismos de modo que resulta más intuitivo hacerse una idea de como se distribuyen las observaciones.

Otros conceptos que según se ha visto, también ayudan en el análisis, son las medidas de tendencia central, que indican hacia donde tienden a agruparse los datos (en el caso en que lo hagan), y las medidas de dispersión, que indican si las diferentes categorías que presenta la variable están muy agrupadas alrededor de cierto valor central, o si por el contrario las variaciones que presentan las categorías con respecto al valor central son grandes.

También se sabe determinar si los datos se distribuyen de forma simétrica a un lado y a otro de un valor central.

En este capítulo se estudiará una situación muy usual y por tanto de gran interés en la práctica:

Si *Y* es otra variable definida sobre la misma población que *X*, ¿será posible determinar si existe alguna relación entre las categóricas de *X* y de *Y*?

Casi siempre, en el desarrollo de los diferentes capítulos que anteceden al presente , se ha venido refiriendo o analizando únicamente una variable, es decir, se realiza una muestra de la cual se obtienen datos para una o varias características , a fin de hacer inferencias acerca de cada parámetro en forma independiente , constituyendo así las llamadas *distribuciones unidimensionales* o datos univariantes.

En este capitulo se estudiará dos o mas variables relacionadas y analizadas simultáneamente ; el primer caso se refiere a distribuciones bidimensionales o a datos bivariantes y el caso especial de distribuciones pluridimensionales, multidimensionales o datos multivariantes ya que se está relacionando mas de dos variables.

Un ejemplo trivial consiste en considerar una población formada por alumnos de primer semestre de Ingeniería Civil y definir sobre ella las variables

X ≡ altura medida en centímetros,

Y ≡ Altura medida en metros,

ya que la relación es determinista y clara: $Y = X/100$. Obsérvese que aunque la variable Y, como tal puede tener cierta dispersión, vista *como función* de X, su dispersión es nula.

Un ejemplo más parecido a lo que interesa realmente se tiene cuando sobre la misma población se define las variables

X ≡ altura medida en centímetros,

Y ≡ Peso medida en kilogramos,

Intuitivamente se espera que exista cierta relación entre ambas variables, por ejemplo,

$$Y = X - 110 \pm \textbf{dispersión}$$

que expresa que (en media) a mayor altura se espera mayor peso. La relación no es exacta y por ello será necesario introducir algún término que exprese la dispersión de $\boldsymbol{Y}$ con respecto a la variable $\boldsymbol{X}$.

Es fundamental de cara a realizar un trabajo de investigación experimental, conocer muy bien las técnicas de estudio de variables bidimensionales (y *n*-dimensionales en general). Baste para ello pensar que normalmente las relaciones entre las variables no son tan evidentes como se mencionó arriba.

La relación que puede existir entre dos variables se puede clasificar de la siguiente forma:

a) *Dependencia casual unilateral*. Esta relación se da cuando una de las variables influye en la otra pero no al contrario.
b) *Interdependencia*. Se presenta cuando la influencia entre las dos variables es reciproca. Sería en caso de dependencia bilateral.
c) *Dependencia indirecta*. Dos variables pueden mostrar una correlación a través de una tercera variable que influye en ellas.
d) *Concordancia*. Se presenta en dos variables independientes a las cuales se les determina la correlación que pueda existir.
e) *Covariación casual*. Cuando la correlación que se presenta entre dos variables es totalmente casual o accidental

3.3 TIPOS DE RELACIÓN GRAFICA ENTRE DOS VARIABLES

La investigación de una relación entre dos variables, comienza con un intento de descubrir la *forma aproximada* de la relación, para lo cual se representan los datos observados en un sistema de coordenadas. Esta grafica recibe el nombre de ***diagrama de dispersión***, el cual muestra la ubicación de los valores o puntos (x_i , y_i) de la variable bidimensional (x,y), en un sistema de coordenadas rectangulares. En la grafica se puede observar si existe o no una relación acentuada entre las variables ***x*** e ***y***, y si se puede ver que forma tiene: lineal u otra. Además puede observarse los valores extremos que puede presentar el diagrama.

3.3.1 Diagrama de dispersión

Es una grafica en la que cada punto trazado representa un par de valores observados de las variables independiente y dependiente. El valor de la variable independiente "***x***" se identifica respecto del eje horizontal, mientras que el valor de la variable independiente "***y***" se identifica respecto del eje vertical.

Ejemplo 3.1

Estudiando la talla, medida en cm. y el peso, medido en kg. de un grupo de 10 personas, se puede obtener los siguientes valores:

Talla (cm)	160	165	168	170	171	175	175	180	180	182
Peso (kg)	55	58	58	61	67	62	66	74	79	83

Tabla 3.1 Tallas y pesos medidos en diez personas

Podemos llamar **X** a la talla e **Y** al peso con lo que se obtendría la variable bidimensional (**X, Y**) que toma 10 valores, que son las 10 parejas de valores de la tabla anterior: (160,55), (165,58), etc.

En la primera fila se colocan los valores de una de las características o variable que componen la variable bidimensional y en la primera columna los de la otra.

Graficando los valores se obtendría el siguiente diagrama de dispersión: (En el eje **X** se representa la talla en cm. y en el eje **Y** el peso en kg.)

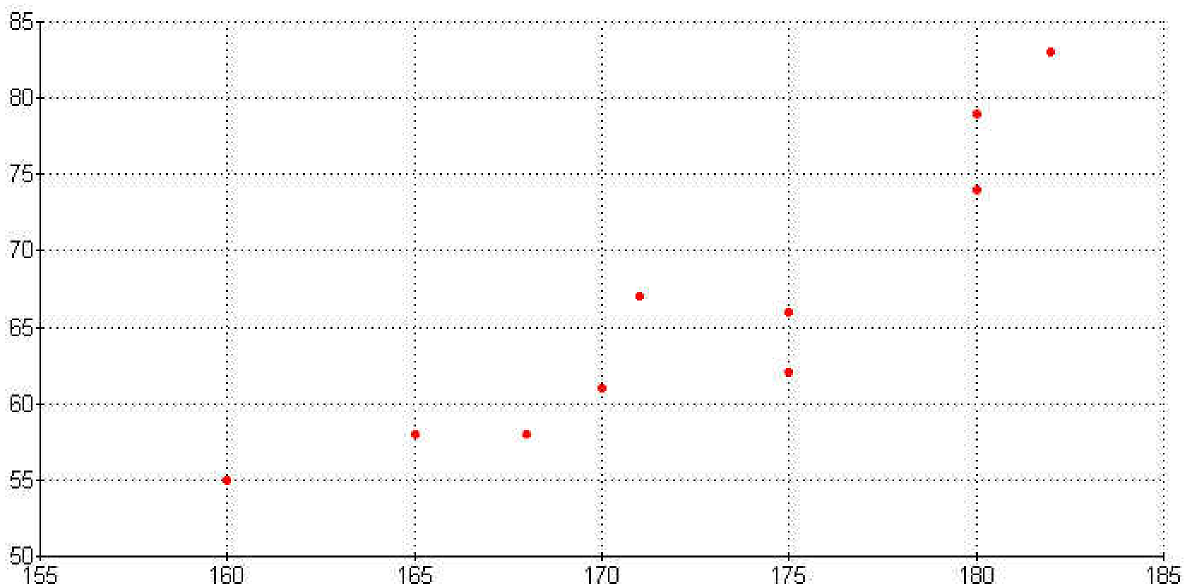


Figura 3.1 Diagrama de dispersión de las tallas y pesos de diez personas

Se puede ver en la figura 3.1, que correspondía al diagrama de **talla – peso,** que existe una interdependencia entre estos valores y además que la serie de puntos presenta una tendencia "**ascendente**". Se dice en este caso que existen entre las dos variables una "**dependencia directa**" .

En caso en que la tendencia sea "**descendente**" se diría que estaríamos ante una "**dependencia inversa**"

Naturalmente en caso en que no se pueda observar una tendencia clara se estaría ante una dependencia muy débil que no se puede observar mediante la nube de puntos y se verá cómo estudiarla posteriormente.

3.2.2 Relación directa e inversa

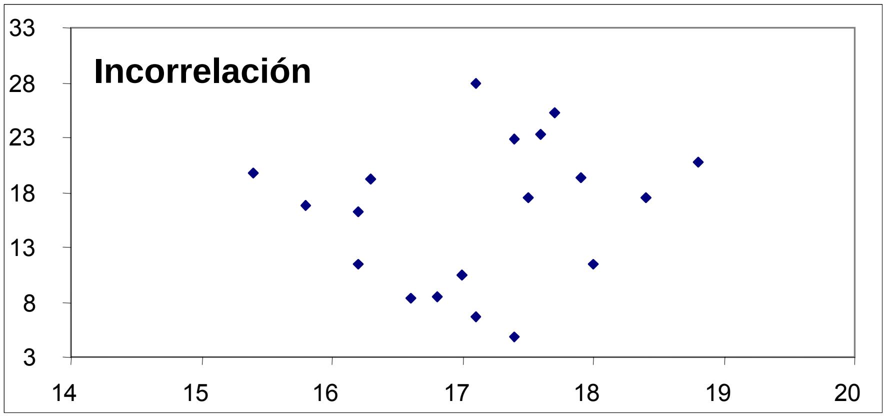


Figura 3.2 Grafica de dispersión valores incorrelacionados

Para valores de X por encima de la media se tiene valores de Y por encima y por debajo en proporciones similares. Incorrelación.

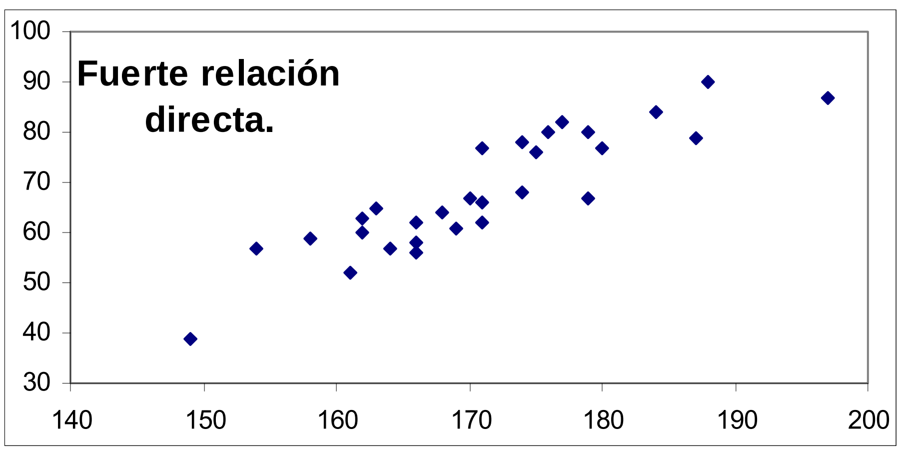


Figura 3.3 Grafica de dispersión valores con fuerte relación directa

- Para los valores de X mayores que la media le corresponden valores de Y mayores también.

- Para los valores de X menores que la media le corresponden valores de Y menores también.

- Esto se llama relación directa.

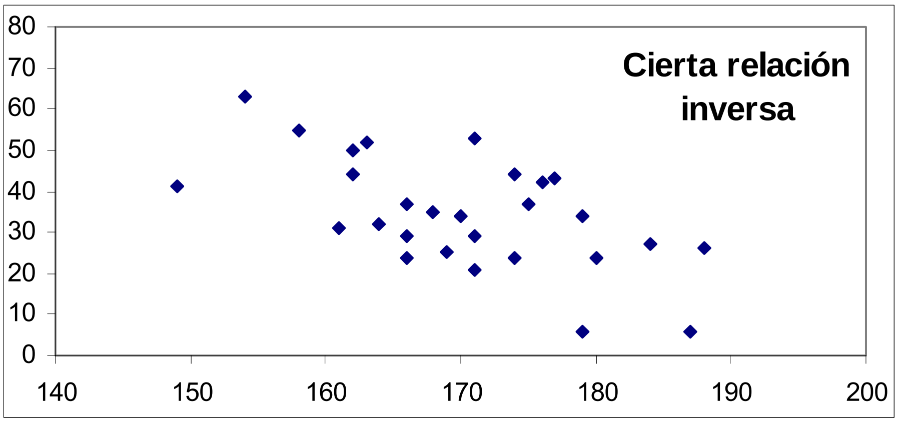


Figura 3.4 Grafica de dispersión valores con cierta relación inversa

Para los valores de X mayores que la media le corresponden valores de Y menores. Esto es relación inversa o decreciente.

3.3 REGRESIÓN

El análisis de Regresión da lugar a una ecuación matemática que permite describir la relación existente entre dos variables. Es decir, obtener una línea "ideal" conocida como *línea de regresión* que describa la relación o dependencia entre dos variables.

Así como el análisis de correlación permite medir la fuerza de asociación entre dos variables el análisis de regresión permite la predicción o sea la estimación de un valor o promedio de una variable denominada dependiente, con base en un valor o un promedio supuestamente conocido para la otra variable denominada independiente.

Esta línea o función matemática, en el caso de una sola variable independiente o explicativa puede ser expresada, a través de una ecuación de curva de aproximación

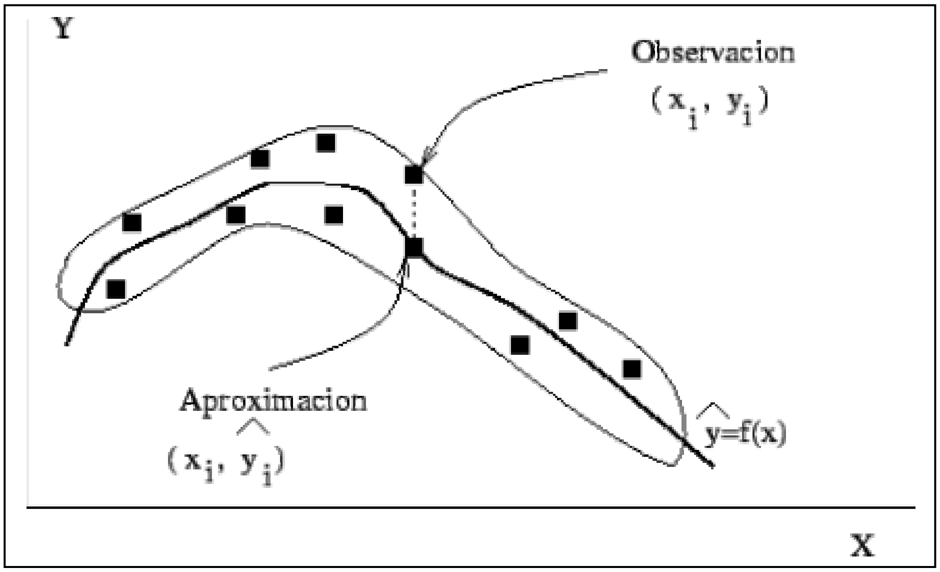


Figura 3.5 Mediante las técnicas de regresión de una variable Y sobre una variable X, se busca una función que sea una buena aproximación de una nube de puntos (x_i,y_i), mediante una curva del tipo $\hat{Y} = f(X)$. Para ello se debe asegurar que la diferencia entre los valores y_i e $\hat{y}_i$ sea tan pequeña como sea posible.

Mediante las técnicas de regresión se inventa una variable $\hat{Y}$ como función de otra variable X (o viceversa),

$$\hat{Y} = f(x)$$

Esto es lo que se denomina **relación funcional**. El criterio para construir $\hat{Y}$, tal como se citó anteriormente, es que la diferencia entre Y e $\hat{Y}$ sea pequeña.

$$\hat{Y} = f(X), \qquad Y - \hat{Y} = error,$$

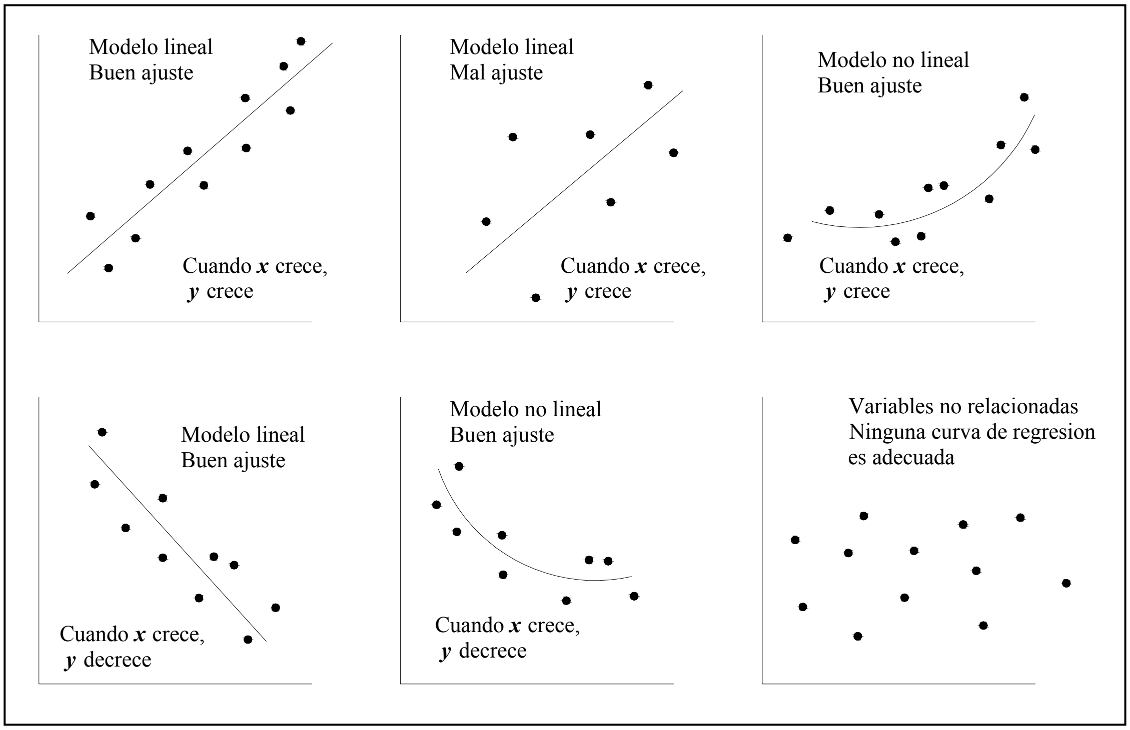


Figura 3.6 Diferentes nubes de puntos y modelos de regresión para ellas.

El término que se ha denominado **error** debe ser tan pequeño como sea posible (Figura 3.5). El objetivo será buscar la función (también denominada **modelo de regresión**) $\hat{Y} = f(x)$ que lo minimice. Véase la Figura 3.6.

3.3.1 Ecuaciones de Curvas de Aproximación

Matemáticamente las ecuaciones serían:

Polinomiales de 1er, 2do, 3er y *n* grados

$y = a_o + a_1 x$	Lineal
$y = a_o + a_1 x + a_2 x^2$	Parábola o curva cuadrática
$y = a_o + a_1 x + a_2 x^2 + a_3 x^3$	Curva cúbica

$y = a_o + a_1 x + a_2 x^2 + ...a_n x^n$ Curva de grado *n*

Otras posibles ecuaciones:

$y = \frac{1}{a_o + a_1 x}$ ó $\frac{1}{y} = a_o + a_1 x$ Hiperbólica

$y = ab^x$ ó $\log y = \log a + [\log b] x$ Exponencial

$y = ax^b$ ó $\log y = \log a + b \log x$ Geométrica

$y = ax^b + g$ Geométrica modificada

$y = ab^x + g$ Exponencial modificada

$y = pq^{b^x}$ Curva de Gompertz

$y = \frac{1}{ab^x + g}$ Curva logístico

3.3.2 Regresión lineal

La forma de la función $\boldsymbol{f}$ en principio podría ser arbitraria, y tal vez se tenga que la relación más exacta entre las variables, peso y altura, definidas anteriormente sea algo de la forma

$$\hat{Y} = f(X) = \frac{1}{\log 8,325 X^{\pi}} \sqrt{e + senX}, \qquad Y - \hat{Y} = error$$

Por el momento no se pretende encontrar relaciones tan complicadas entre variables, pues el estudio se limitara al caso de la **regresión lineal**. Con este tipo de regresiones bastará con encontrar relaciones funcionales de tipo lineal, es decir, se buscará cantidades a y b tales que se pueda escribir:

$$\hat{Y} = a + b \cdot X \tag{3.1}$$

con el menor error posible entre $\hat{Y}$ e Y, o bien

$$\hat{X} = a + b \cdot Y \tag{3.2}$$

de forma que X - $\hat{X}$ sea una variable que toma valores próximos a cero.

Observación

Obsérvese que la relación 3.1 explica cosas como que si X varía en 1 unidad, $\hat{Y}$ varía la cantidad b. Por tanto:

- Si b>0, las dos variables aumentan o disminuyen a la vez;
- Si b<0, cuando una variable aumenta, la otra disminuye.

Por tanto, en el caso de las variables peso y altura lo lógico será encontrar que b>0.

El problema que se plantea es entonces el de cómo calcular las cantidades a y b a partir de un conjunto de n observaciones

$$\begin{cases} (x_1, y_1) \\ (x_2, y_2) \\ \cdots \\ (x_n, y_n) \end{cases}$$

de forma que se minimice el error. Las etapas en que se divide el proceso que se va a desarrollar son de forma esquemática, las que siguen:

1. Dadas dos variables X, Y, sobre las que se define

$$\hat{Y} = a + b \cdot X$$

se mide el error que se comete al aproximar Y mediante $\hat{Y}$ calculando la suma de las diferencias entre los valores reales y los aproximados al cuadrado (para que sean positivas y no se compensen los errores):

$$\sum_{i=1}^{n}(y_i - \hat{y}(x_i))^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2 \tag{3.3}$$

2. Una aproximación $\hat{Y} = a + b \cdot X$ de Y, se define a partir de dos cantidades a y b. Se calculara aquellas que minimizan la función

$$E_{rror}(a,b) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - a - bx_i)^2 \tag{3.4}$$

3. Posteriormente se encontrara fórmulas para el cálculo directo de a y b que sirvan para cualquier problema.

3.3.3 Regresión de *Y* sobre *X*

Para calcular la recta de regresión de Y sobre X tomando como referencia la Figura 3.7.

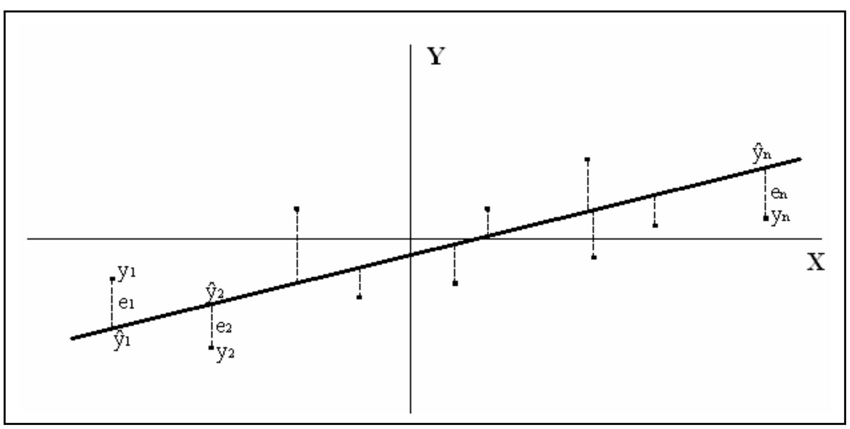


Figura 3.7 Los errores a minimizar son las cantidades $e_i^2 = (y_i - \hat{y}_i)^2$

Una vez que se tiene definido el error de aproximación mediante la ecuación 3.4, las cantidades que lo minimizan se calculan derivando con respecto a ambas e igualando a cero *(procedimiento de los mínimos cuadrados)*:

$$(a,b) \text{ minimizan } E_{rror}(a,b) \quad \Leftrightarrow \quad \begin{cases} \dfrac{E_{rror}}{\partial a}(a,b) = 0 \\ \dfrac{E_{rror}}{\partial b}(a,b) = 0 \end{cases}$$

$$\Leftrightarrow \quad \begin{cases} \dfrac{E_{rror}}{\partial a}(a,b) = -2\sum_{i=1}^{n}(y_i - a - bx_i) = 0 \\ \dfrac{E_{rror}}{\partial b}(a,b) = -2\sum_{i=1}^{n}(y_i - a - bx_i)x_i = 0 \end{cases}$$

$$\Leftrightarrow \quad \begin{cases} \dfrac{E_{rror}}{\partial a}(a,b) = -2\sum_{i=1}^{n} e_i = 0 \\ \dfrac{E_{rror}}{\partial b}(a,b) = -2b\sum_{i=1}^{n} e_i x_i = 0 \end{cases}$$

De las relaciones anteriores se puede se puede obtener otra relación, que se denomina *ecuaciones normales*. La primera de se escribe como

$$\sum_{i=1}^{n} y_i - an - b\sum_{i=1}^{n} x_i = 0 \quad \Leftrightarrow \quad a = \overline{y} - b\overline{x}$$

Sustituyendo se tiene que

$$\sum_{i=1}^{n} y_i x_i - a\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} x_i^2 = 0 \quad \Leftrightarrow \quad \frac{1}{n}\sum_{i=1}^{n} y_i x_i - \overline{x}(\overline{y} - b\overline{x}) - b\frac{1}{n}\sum_{i=1}^{n} x_i^2 = 0$$

$$\Leftrightarrow \quad S_{XY} - bS_X^2 = 0$$

Lo que da las relaciones buscadas:

$$a = \overline{y} - b\overline{x} \tag{3.5}$$

$$b = \frac{S_{XY}}{S_X^2} \tag{3.6}$$

La cantidad ***b*** se denomina *coeficiente de regresión de* Y *sobre* X.

3.3.4 Regresión de *X* sobre *Y*

Las mismas conclusiones se sacan cuando se intenta hacer la regresión de *X* sobre *Y*, pero ¡atención!: Para calcular la recta de regresión de *X* sobre *Y* es *totalmente incorrecto* despejar de

$$\hat{Y} = a + bX \quad \Rightarrow \quad X = \frac{1}{b}(\hat{Y} - a)$$

Pues esto da la regresión de *X* sobre $\hat{Y}$, que no es lo que se busca. La regresión de *X* sobre *Y* se hace aproximando *X* por $\hat{X}$, del modo

$$\hat{X} = a + bY$$

donde:

$$a = \overline{x} - b\overline{y} \tag{3.7}$$

$$b = \frac{S_{XY}}{S_Y^2} \tag{3.8}$$

pues de este modo se minimiza, en el sentido de los mínimos cuadrados, los errores entre las cantidades x_i y las $\hat{x}_i = a + by_i$ (Figura 3.8.)

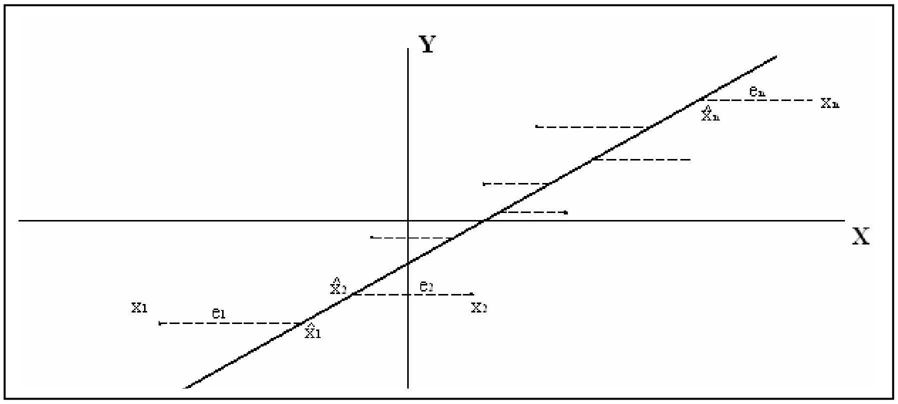


Figura 3.8 Los errores a minimizar son las cantidades $e_i^2 = (x_i - \hat{x}_i)^2$

Ejemplo 3.2

En una muestra de 1.500 individuos se recogen datos sobre dos medidas antropométricas *X* e *Y*. Los resultados se muestran resumidos en los siguientes estadísticos:

$$\overline{x} = 14 \quad S_X = 2$$

$$S_{XY} = 45$$

$$\overline{y} = 100 \quad S_Y = 25$$

Obtener el modelo de regresión lineal que mejor aproxima *Y* en función de *X*. Utilizando este modelo, calcular de modo aproximado la cantidad *Y* esperada cuando *X*=15.

Solución:

Lo que se busca es la recta, $\hat{Y} = a + b \cdot X$, que mejor aproxima los valores de *Y* (según el criterio de los mínimos cuadrados) en la nube de puntos que resulta de representar en un plano (*X*,*Y*) las 1.500 observaciones. Los coeficientes de esta recta son:

$$b = \frac{S_{XY}}{S_X^2} = \frac{45}{4} = 11,25$$

$$a = \overline{y} - b \cdot \overline{x} = 100 - 11,25 \times 14 = -57,5$$

Así, el modelo lineal consiste en:

$$\hat{Y} = -57,5 + 11,25 \cdot X$$

Por tanto, si x=15, el modelo lineal predice un valor de Y de:

$$\hat{y} = -57,5 + 11,25 \cdot x = -57,5 + 11,25 \times 15 = 111,25$$

En este punto debe analizarse si realmente esta predicción puede considerarse fiable. Para dar una respuesta, es necesario estudiar propiedades de la regresión lineal que están a continuación.

3.3.5 Varianza Residual, y Error Estándar de Estimación =VR

La Varianza Residual, varianza no explicada o la suma de los cuadrados del error son los nombres que se empelan para identificar a esta medida de dispersión, que permite calcular el grado de variación o de dispersión que presenta el conjunto de puntos u observaciones que no quedaron explicados por la recta de regresión.

Es decir, la dispersión de aquellos valores que no quedan dentro de la recta establecida para el conjunto de observaciones.

Definición 3.1.- *La **Varianza Residual** Se define como la media aritmética del cuadrado de las diferencias, entre los valores observados y los valores estimados. Las fórmulas que se emplean para su cálculo son varias, entre las cuales tenemos:*

$$S_{yx}^2 = \frac{\sum (y_1 - \hat{y})^2}{n} \tag{3.9}$$

$$S_{yx}^2 = \frac{\sum (z_1 - \hat{z})^2}{n} \tag{3.10}$$

$$S_{yx}^2 = S_y^2 - \frac{m_{xy}^2}{s_x^2} \tag{3.11}$$

$$S_{yx}^2 = S_y^2 - 2bm_{xy} + b^2 S_x^2 \tag{3.12}$$

$$S_{yx}^2 = \frac{\sum y_i^2 - c\sum y_i - b\sum y_i x_i}{n} \tag{3.13}$$

$$S_{yx}^2 = S_y^2 (1 - r^2) \tag{3.14}$$

3.3.6 Error estándar de estimación

Es una expresión dada a la desviación estándar de los valores observados respecto a una línea de regresión. Se podría decir que es una estimación de la variación que probablemente acompañara a las predicciones realizadas por medio de la ecuación de regresión.

Se define como la raíz cuadrada de la varianza residual

$$S_{yx} = \sqrt{S_{yx}^2} \tag{3.15}$$

3.3.7 Varianza Explicada = VE

Hasta el momento se han calculado las varianzas para la variable **x** y para la variable **y** $(S_x^2 = S_y^2)$, denominados como; varianza total de $\boldsymbol{x}$ y Varianza total de $\boldsymbol{y}$. También se han calculado los valores para las varianzas no explicadas o varianza residual (S_{yx}^2). Ahora debe tratarse el calculo de otra varianza, de gran utilidad en el análisis de regresión, conocida como *varianza explicada*, simbolizada por S_{ay}^2.

Definición 3.2 *La* ***varianza explicada*** *es una medida que permite calcular el grado de dispersión de aquellos valores que quedaron explicados por la recta de represión, es decir, la dispersión de aquellos puntos en la gráfica que se confunden con los puntos de la recta o valores estimados. Se define como la media aritmética, de los cuadrados de las diferencias entre los valores estimados y la media aritmética de esa variable. Su formula es:*

$$S_{ay}^2 = \frac{\sum (\hat{y}_i - \overline{y})}{n} \tag{3.16}$$

Es necesario saber que la varianza total debe ser igual a la suma de la varianza residual mas la varianza explicada: $VT = VR + VE$. De acuerdo a dicha igualdad se tendrá que $VE = VT - VR$.

3.3.8 Regresión lineal ponderada

Imaginando que se maneja una cantidad de datos muy grande, el manejo de esa información seria complicada si no es agrupada. Este procedimiento permite agilizar operaciones y reducir el espacio de trabajo. Las formulas estarán ponderadas y el procedimiento será exactamente igual al desarrollo para el método simple.

Calculemos los coeficientes angulares (**b**) y de posición (**c**) cuando se va a estimar $\hat{y}$.

Ecuaciones Normales:

$$\sum y_i n_i = b \sum x_i n_i \tag{3.17}$$

$$\sum x_i y_i n_i = b \sum x_i^2 n_i + c \sum x_i n_i \tag{3.18}$$

teniendo en cuenta las siguientes ecuaciones:

$$\overline{x} = \frac{\sum x_i n_i}{n}$$

$$\overline{y} = \frac{\sum y_i n_i}{n}$$

$$S_y^2 = \frac{\sum y_i^2 n_i - n\overline{y}^2}{n}$$

$$S_x^2 = \frac{\sum x_i^2 n_i - n\overline{x}^2}{n}$$

$$S_{xy} = \frac{\sum x_i y_i n_i}{n} - \overline{x} \cdot \overline{y}$$

Entonces se tiene:

$$b = \frac{S_{xy}}{S_x^2} = \frac{n\sum x_i y_i n_i - (\sum x_i n_i)(\sum y_i n_i)}{n\sum x_i^2 n_i - (\sum x_i n_i)^2} \tag{3.19}$$

$$c = \overline{y} - b\overline{x} = \frac{\sum y_i n_i - b\sum x_i n_i}{n} \tag{3.20}$$

En cuanto a la varianza residual se tiene;

$$S_{yx}^2 = \frac{\sum (y_i - \hat{y})^2 n_i}{n} = \frac{\sum y_i^2 n_i - c\sum y_i n_i - b\sum x_i y_i n_i}{n} \tag{3.21}$$

El error estándar de estimación es igual a

$$S_{yx} = \sqrt{S_{yx}^2} \tag{3.22}$$

En cuanto al *coeficiente de correlación*, se tiene:

$$R^2 = \frac{S_{xy}^2}{S_x^2 S_y^2} = \frac{\left[n\sum x_i y_i n_i - (\sum x_i n_i)(\sum y_i n_i)\right]^2}{\left[n(\sum x_i^2 n_i) - (\sum x_i n_i)^2\right]\left[n\sum y_i^2 n_i - (\sum y_i n_i)^2\right]} \tag{3.23}$$

3.3.9 Regresión parabólica simple

Aplicada a todos aquellos fenómenos que en un diagrama de dispersión se pueden observar, como la concentración de puntos que presenta una parte ascendente y en seguida otra descendente (puede darse lo contrario).

La ecuación general de la parábola es $\hat{y} = ax^2 + bx + c$

Para encontrar los parámetros, se tienen las siguientes ecuaciones normales:

$$\sum y = a\sum x^2 + b\sum x + nc \tag{3.24}$$

$$\sum xy = a\sum x^3 + b\sum x^2 + c\sum x \tag{3.25}$$

$$\sum xy = a\sum x^4 + b\sum x^3 + c\sum x^2 \tag{3.26}$$

A partir de acá el procedimiento es igual

3.4 MEDIDA DE LA FUERZA DE LA RELACIÓN

3.4.1 Covarianza

Definición 3.3.- *La* ***covarianza*** $\boldsymbol{S_{XY}}$*, es una medida que permite determinar que tan independiente es una variable aleatoria de otra, es decir, el grado de independencia de dos variables aleatorias. Es una manera de generalizar la varianza y se define como:*

$$S_{XY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_1 - \overline{y}) \tag{3.27}$$

o lo que es lo mismo

$$S_{XY} = \frac{1}{n}\sum_{i=1}^{n} x_i y_1 - \overline{xy} \tag{3.28}$$

3.4.2 Una interpretación geométrica de la covarianza

Considerando la *nube de puntos* formadas por las *n* parejas de datos (x_i,y_i). El centro de gravedad de esta nube de puntos es, $(\overline{x}, \overline{y})$. Trasladando los ejes *XY* al nuevo centro de coordenadas $(\overline{x}, \overline{y})$. Queda así dividida la nube de puntos en cuatro cuadrantes como se observa en la Figura 3.9. Los puntos que se encuentran en el primer y tercer cuadrante contribuyen positivamente al valor de $\boldsymbol{S_{XY}}$, y los que se encuentran en el segundo y el cuarto lo hacen negativamente.

De este modo:

- Si hay mayoría de puntos en el tercer y primer cuadrante, ocurrirá que $S_{XY} \geq 0$, lo que se puede interpretar como que la variable ***Y*** tiende a aumentar cuando lo hace ***X***;

- Si la mayoría de puntos están repartidos entre el segundo y cuarto cuadrante entonces $S_{XY} \leq 0$, es decir, las observaciones ***Y*** tienen tendencia a disminuir cuando las de ***X*** aumentan;

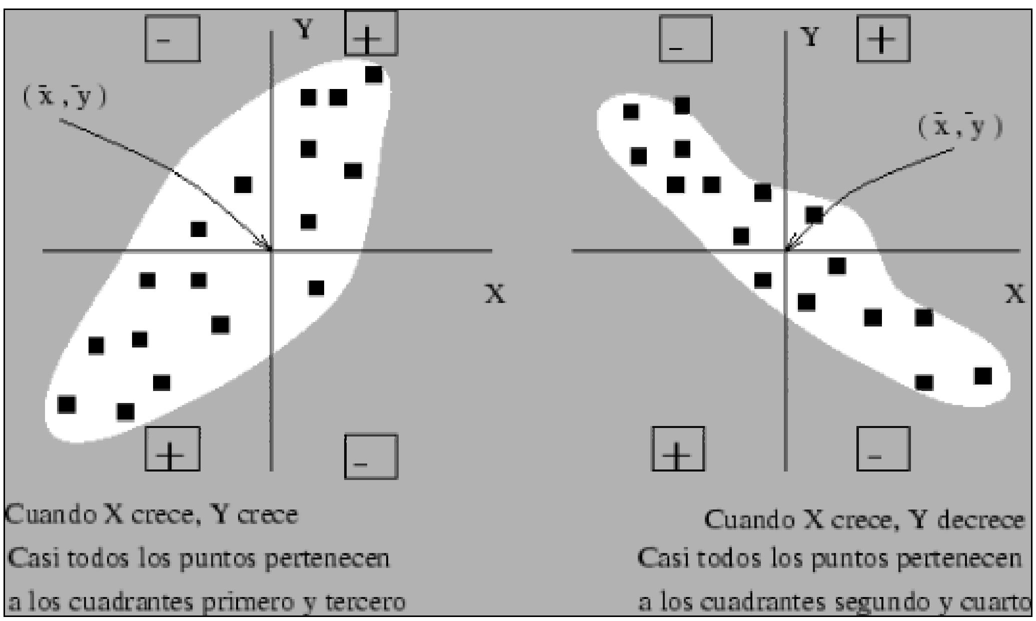


Figura 3.9 Interpretación geométrica de S_{XY}

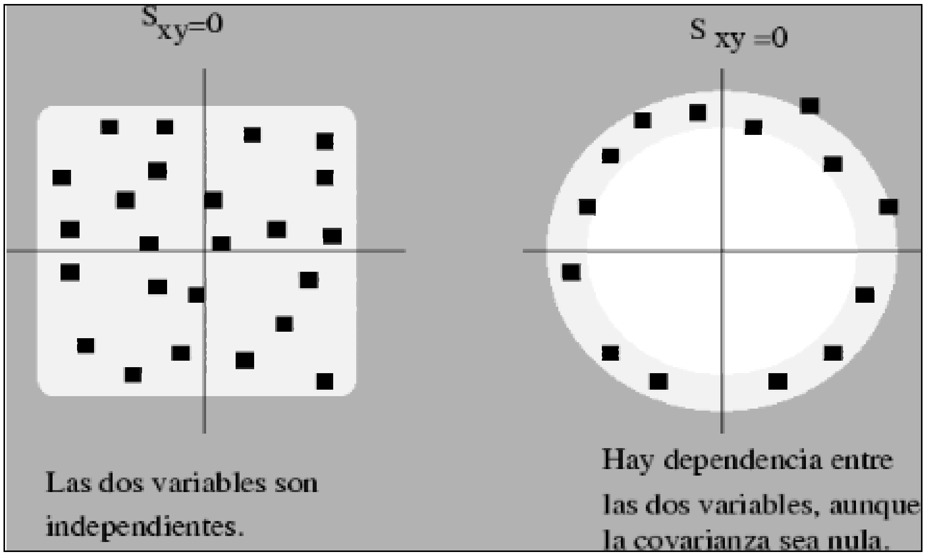


Figura 3.10 Cuando los puntos se reparten de modo más o menos homogéneo entre los cuadrantes primero y tercero, y segundo y cuarto, se tiene que $S_{XY} \approx 0$. Eso no quiere decir de ningún modo que no pueda existir ninguna relación entre las dos variables, ya que ésta puede existir como se aprecia en la figura de la derecha.

- Si $S_{XY} > 0$ las dos variables crecen o decrecen a la vez (nube de puntos creciente).
- Si $S_{XY} < 0$ cuando una variable crece, la otra tiene tendencia a decrecer (nube de puntos decreciente).
- Si los puntos se reparten con igual intensidad alrededor de $(\bar{x}, \bar{y})$, $S_{XY} = 0$ (no hay relación lineal).

De este modo se puede utilizar la covarianza para medir la variación conjunta (*covariación*) de las variables X e Y. Esta medida no debe ser utilizada de modo exclusivo para medir la relación entre las dos variables, ya que es sensible al cambio de unidad de medida, como se observa en el siguiente resultado:

3.4.3 Correlación

Se conoce que la covarianza mide la relación entre dos variables, el inconveniente de la covarianza como medida de asociación es su dependencia de las unidades de medida de las variables. En consecuencia, para construir una medida adimensional, tendremos que dividir la covarianza por un término con sus mismas dimensiones. Si se divide por el producto de sus desviaciones típicas se define el *coeficiente de correlación* entre dos variables.

Definición 3.4 *El **análisis de correlación** describe el grado o fuerza con que se produce esta relación, para ello se utilizara una medida conocida **como coeficiente de correlación.***

3.4.4 Coeficiente de correlación =R=r

Definición 3.5 *El **coeficiente de correlación** es una medida de interdependencia de dos variables aleatorias.*

En primer lugar deberá considerarse el cálculo del coeficiente de correlación al cuadrado denominado también como *coeficiente de determinación*, simbolizado por R^2. Las formulas para el calculo de este coeficiente es variado, siendo el resultado igual que cualquiera de los métodos que se utilice.

$$R^2 = \frac{VE}{VT} \qquad (3.29)$$

$$R^2 = 1 - \frac{VR}{VT} \tag{3.30}$$

para ajustes de tipo lineal se tiene que el coeficiente de determinaciones igual a r^2 , y por tanto representan además la proporción de de varianza explicada por la regresión lineal , de esta manera se tiene que el coeficiente de correlación es igual a:

$$r = \frac{S_{XY}}{S_X S_Y} \tag{3.31}$$

El coeficiente de correlación lineal posee las siguientes propiedades:

EL COEFICIENTE DE CORRELACION

- Carece de unidades de medida (adimensional)
- Es invariable para transformaciones lineales (cambio de origen y escala) de las variables.
- Sólo toma valores comprendidos entre -1 y 1, **$-1 \leq r \leq 1$.**
- Cuando $|r|$ esté próximo a uno, se tiene que existe una relación lineal muy fuerte entre variables.
- Cuando $r \approx 0$, puede afirmarse que no existe relación lineal entre ambas variables.

3.4.5 Interpretación geométrica de *r*

Si los datos son observaciones que no están ordenadas en una tabla bidimensional, se tendrá parejas de valores para cada sujeto o elemento

$$(x_i, y_i), \qquad i = 1, \ldots, n$$

la fórmula de la covarianza, en este caso, es

$$S_{XY} = \sum_{i=1}^{n} (x_i - \overline{x}) \cdot (y_i - \overline{y})$$

se puede escribir las observaciones en forma de vectores de la siguiente manera:

$$\vec{X} = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$$

$$\vec{Y} = (y_1, y_2, ..., y_n) \in \mathbb{R}^n$$

$$\vec{\overline{x}} = (\overline{x}, \overline{x}, ..., \overline{x}) \in \mathbb{R}^n$$

$$\vec{\overline{y}} = (\overline{y}, \overline{y}, ..., \overline{y}) \in \mathbb{R}^n$$

Si se denota $\vec{\upsilon} \cdot \vec{\omega}$ al producto escalar de los vectores $\vec{\upsilon}$ y $\vec{\omega}$, es inmediato comprobar que en realidad las definiciones de varianza y covarianza tienen una idea geométrica muy simple: son productos escalares en los que intervienen los vectores $\vec{X} - \vec{\overline{x}}$ e $\vec{Y} - \vec{\overline{y}}$

$$S_{XY} = \frac{1}{n}(\vec{X} - \vec{\overline{x}}) \cdot (\vec{Y} - \vec{\overline{y}})$$

$$S_X^2 = \frac{1}{n}(\vec{X} - \vec{\overline{x}})(\vec{X} - \vec{\overline{x}}) = \frac{1}{n}\left|\vec{X} - \vec{\overline{x}}\right|^2 = S_{XX}$$

$$S_Y^2 = \frac{1}{n}(\vec{Y} - \vec{\overline{y}})(\vec{Y} - \vec{\overline{y}}) = \frac{1}{n}\left|\vec{Y} - \vec{\overline{y}}\right|^2 = S_{YY}$$

Con esta descripción geométrica de las varianzas y covarianzas, se puede poner de manifiesto la existencia de *paralelismo* entre las desviaciones de las variables *X* e *Y*, con respecto a sus centros de gravedad ya que

$$(\vec{X} - \vec{\overline{x}}) \cdot (\vec{Y} - \vec{\overline{y}}) = \left|\vec{X} - \vec{\overline{x}}\right| \cdot \left|\vec{Y} - \vec{\overline{y}}\right| \cdot \cos\theta,$$

donde θ es el ángulo entre los vectores $\vec{X} - \vec{\overline{x}}$ e $\vec{Y} - \vec{\overline{y}}$ (véase la figura 3.11). Despejando:

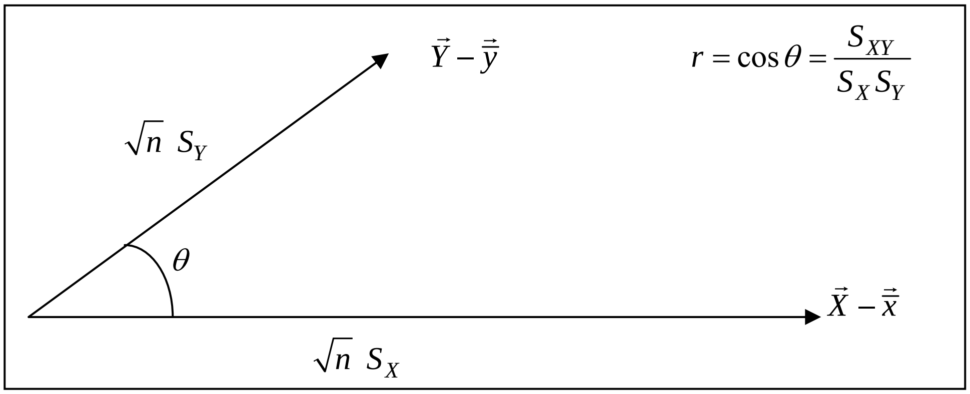


Figura 3.11 Interpretación geométrica de *r* como el coseno del ángulo que forman los vectores de las desviaciones con respecto a sus respectivas medias de *X* y de *Y*.

$$\cos\theta=\frac{(\vec{X}-\vec{\bar{x}})\cdot(\vec{Y}-\vec{\bar{y}})}{\left|\vec{X}-\vec{\bar{x}}\right|\cdot\left|\vec{Y}-\vec{\bar{y}}\right|}=\frac{\frac{1}{n}(\vec{X}-\vec{\bar{x}})\cdot(\vec{Y}-\vec{\bar{y}})}{\sqrt{\frac{1}{n}}\left|\vec{X}-\vec{\bar{x}}\right|\cdot\sqrt{\frac{1}{n}}\left|\vec{Y}-\vec{\bar{y}}\right|}=\frac{S_{XY}}{S_X S_Y} \tag{3.32}$$

Si los vectores $\vec{X}-\vec{\bar{x}}$ e $\vec{Y}-\vec{\bar{y}}$ son totalmente paralelos entonces **cosθ = ±1**. En este caso existirá una constante de proporcionalidad *m* tal que:

$$(\vec{Y}-\vec{\bar{y}})=m(\vec{X}-\vec{\bar{x}})$$

Esta es la ecuación de una recta (véase la Figura 3.4). Es decir:

$$r=\pm 1$$
$$\Updownarrow$$
Las desviaciones con respecto a la media de ambas variables son proporcionales
$$\Updownarrow$$
Las observaciones están perfectamente alineadas

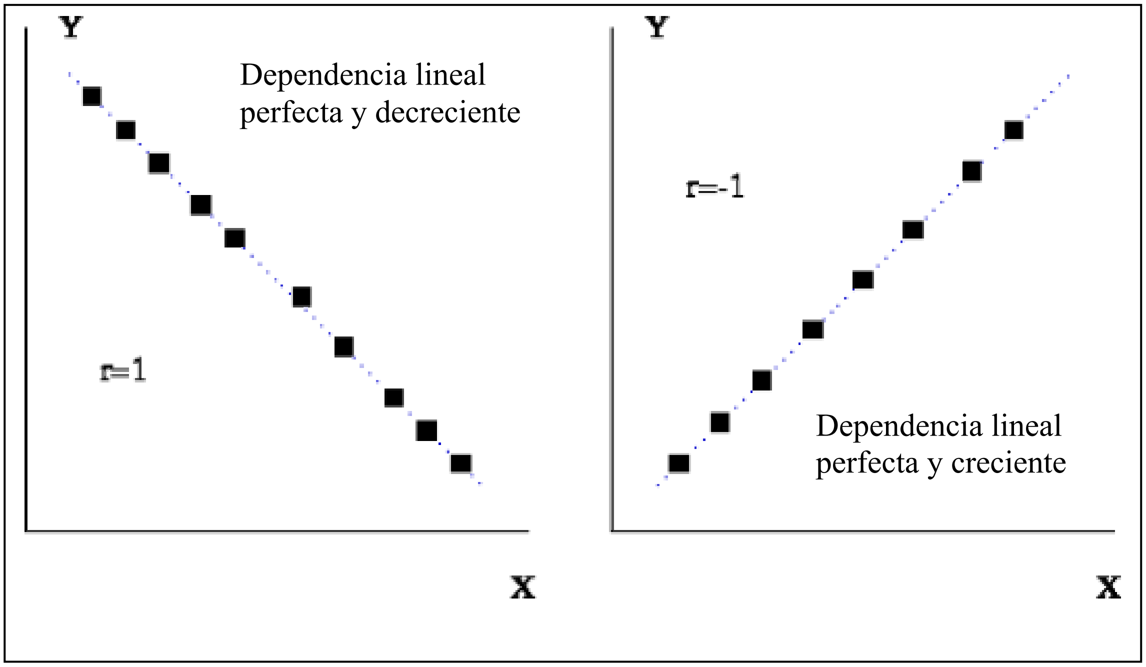


Figura 3.12 $r = \pm 1$ es lo mismo que decir que las observaciones de ambas variables están perfectamente alineadas. El signo de *r*, es el mismo que el de S_{XY}, por tanto indica el crecimiento o decrecimiento de la recta.

La magnitud que expresa el coseno del ángulo que forman los vectores $\vec{X} - \vec{\bar{x}}$ e $\vec{Y} - \vec{\bar{y}}$ tiene un papel muy destacado como se ver más adelante en regresión lineal. Se ha denominado anteriormente como *coeficiente de correlación lineal de Pearson* y se representa mediante la letra *r*:

$$r = \cos\theta = \frac{S_{XY}}{S_X S_Y} \tag{3.33}$$

Son evidentes entonces las siguientes propiedades de *r*

- Cualesquiera que sean los valores (x_i, y_i) , $\boldsymbol{i = 1,\ldots,n}$, se tiene que $\boldsymbol{-1 < r < 1}$, ya que ***r*** es el coseno del ángulo que forman las variaciones con respecto a sus valores medios de las observaciones $\boldsymbol{x_i}$ e $\boldsymbol{y_i}$. Si cuando r es calculado en un caso práctico se obtiene un valor no comprendido en ese rango, es signo evidente de que se ha cometido un error de cálculo, que por tanto ha de ser revisado.

- Si las desviaciones con respecto al valor central de las observaciones x_i, son proporcionales a las desviaciones de y_i con respecto a su valor centra $\overline{y}$,

  $$\vec{Y}-\vec{\overline{y}}=m(\vec{X}-\vec{\overline{x}}) \Leftrightarrow y_i=(\overline{y}-m\overline{x})+m\cdot x_i \quad \forall i=1,\ldots,n$$

  entonces los vectores $\vec{X}-\vec{\overline{x}}$ e $\vec{Y}-\vec{\overline{y}}$ son paralelos y por tanto $\mathbf{r = \pm 1}$. En este caso se puede decir de modo exacto que conocido X lo es también Y, (y recíprocamente).

- Por el contrario si no existe dicha relación, el ángulo que formen $\vec{X}-\vec{\overline{x}}$ e $\vec{Y}-\vec{\overline{y}}$ será mayor, siendo el caso extremo en que ambos sean perpendiculares (r=0). Cuando $\boldsymbol{r}$**=0** decimos que las variables X e Y son **incorreladas**.

Otra propiedad interesante de r es la siguiente:

Proposición

El coeficiente de correlación entre dos variables no se ve afectada por los cambios de unidades.

Demostración

Considerando la variable bidimensional (X,Y) y sometiendo a Y a un cambio de unidad $Z = a + bY$. Entonces

$$\frac{S_{XZ}}{S_X S_Z}=\frac{bS_{XY}}{bS_X S_Y}=\frac{S_{XY}}{S_X S_Y}$$

Por tanto ambas variables *XZ* y *XY* tienen el mismo coeficiente de correlación.

3.4.6 Regresión y correlación exponencial

Si las dos o más variables crecen o decrecen en forma de progresión aritmética, se debe aplicar la regresión lineal o rectilínea, si por el contrario se comprueba que esas variables, crecen o decrecen en forma de progresión geométrica, se adoptara una función exponencial.

La ecuación general es $\hat{y} = cb^x$, la cual se reduce al de la función lineal cuando trabajamos con logaritmos, ya sea neperianos o con base 10. Veamos como queda

$$\log \hat{y} = \log c + (x)\log b \tag{3.34}$$

para su representación grafica se debe utilizar papel semilogarítmico cuando la viable X, localizada en el eje horizontal o abscisa se presenta en forma de progresión aritmética, mientras que en la ordenada, donde se ubica la variable Y, se expresa en forma logarítmica. Si ambas variables tienen crecimiento geométrico, la representación grafica se hace en papel logarítmico.

Ecuaciones Normales:

$$\sum \log y = n\log c + (\sum x)\log b \tag{3.35}$$

$$\sum x\log y = (\sum x)\log c + (\sum x^2)\log b \tag{3.36}$$

Se debe observar que $\sum y_i \neq \sum \hat{y}_i$, por el contrario $\sum \log y_i \cong \sum \log \hat{y}_i$

Calculo de b y c directamente

$$\log b = \frac{n\sum x \log y - (\sum x)(\sum \log y)}{n} \tag{3.37}$$

$$\log c = \frac{\sum \log y - \sum \log b(\sum x)}{n} \tag{3.38}$$

Varianza residual y error estándar de estimación

Aplicando la formula que se presenta con mayor frecuencia en los textos de estadística, siendo:

$$S_{x\log y} = S^2_{x\log y} = \frac{\sum(\log y_i - \log \hat{y})^2}{n} \quad \text{ó}$$

También se puede calcular la varianza residual, mediante las siguientes formulas:

$$S^2_{x\log y} = \frac{\sum(\log y_i)^2 - \log c(\sum \log y_i) - \log b(\sum x_i \log y_i)}{n} \tag{3.39}$$

Coeficiente de correlación al cuadrado

$$r = \frac{n\sum x_i \log y_i - (\sum x_i)(\sum \log y_i)}{\sqrt{\left[n\sum x_i^2 - (\sum x_i)^2\right]\left[n\sum(\log y_i)^2 - (\sum \log y_i)^2\right]}} \tag{3.40}$$

Algunos autores denominan al coeficiente de correlación (de cualquier ajuste) como coeficiente de correlación de Pearson.

3.4.7 Coeficiente de correlación por rangos

Los coeficientes de correlación por rangos más utilizados son los de Pearson y Spearman, éste último de gran aplicación especialmente cuando se desean investigar preferencias, actitudes o atributos, por medio de algún sistema de calificación que va en orden ascendente de graduación. Puede utilizarse en los fenómenos cuantitativos o variables, si que éstas tengan significado numérico.

La formula aplicada es:

$$r_s = 1 - \frac{6\sum d_i^2}{n^3 - n} \tag{3.41}$$

Donde:

d_i = representa la diferencia entre los rangos pareados de cada par de variables x_i e y_i

n = es el numero de rangos o sea el número de variantes pareadas

El procedimiento que se sigue en el cálculo del coeficiente, consiste en colocar en orden: primeo, segundo, tercero, etc. Con respecto a cada una de las variables y la diferencia entre los dos rangos se usa como base para el cálculo.

3.5 CASO ESPECIAL (REGRESIÓN Y CORRELACIÓN MÚLTIPLE)

Cuando se trabaja con más de dos variables independientes, relacionadas entre si, el análisis de regresión recibe el nombre de *regresión múltiple*. El grado de relación que pueda haber entre estas variables es cuantificado mediante el cálculo del *coeficiente de correlación múltiple.*

En estadística con fines de explicación y aplicación , por lo general se trabaja con dos variables independientes, esto permite la realización de operaciones menos engorrosas que trabajando con mas de dos, lo que obligaría a la aplicación de álgebra matricial o al manejo de algunos paquetes estadísticos que contienen programas de regresión , entre otros SPSS, TSP, SAS, MICROSTAT. Y otros que son muy comunes.
La expresión de la función utilizada en la regresión múltiple se puede ver a continuación

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + ... + a_mx_m \qquad (3.42)$$

a los parámetros $\boldsymbol{a_1}$, $\boldsymbol{a_2}$,$\boldsymbol{a_3}$ se les denomina coeficientes de regresión , el coeficiente $\boldsymbol{a_1}$, mide la cantidad (+/-) que aumenta o disminuye la variable dependiente $\hat{y}$ o $\hat{x}$, cuando la primera variable independiente X aumenta en una unidad, manteniendo constante los otros coeficientes; lo mismo puede decirse para los demás coeficientes de regresión , siempre que las demás variables independientes en la ecuación permanezcan constantes. Los coeficientes de regresión miden las elasticidades de la variable dependiente con respecto a la variable independiente respectiva, el símbolo $\boldsymbol{a_o}$, corresponden al coeficiente de posición llamado también *intercepto*, representa un punto donde la línea de regresión cruza el eje ***y***.

Para la ecuación general (ecuación 3.42) expuesta anteriormente, se tiene el siguiente sistema de ecuaciones normales, cuando tenemos *m* variables independientes y una dependiente.

$$\begin{aligned}
&\sum y = a_o n + a_1 \sum x_1 + a_2 \sum x_2 + \ldots + a_m \sum x_m \\
&\sum x_1 y = a_o \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_2 x_1 + \ldots + a_m \sum x_m x_1 \\
&\sum x_2 y = a_o \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x_2^2 + \ldots + a_m \sum x_m x_2 \\
&\quad \vdots \\
&\sum x_m y = a_o \sum x_m + a_1 \sum x_1 x_m + a_2 \sum x_2 x_m + \ldots + a_m \sum x_m^2
\end{aligned} \tag{3.43}$$

donde:

n: numero de grupos de elementos de la muestra.

La solución del sistema (3.43) proporcionan valores de a_o , a_1, a_2,..., a_m.

El número de ecuaciones normales, deben ser tantas como incógnitas se tienen, a fin de que se tenga un sistema resoluble y así encontrar: a_o , a_1, a_2,..., a_m.

3.5.1 Error estándar del estimado para regresión múltiple (Se)

Definición 3.6 *El **error estándar de estimación** es la medida de la dispersión que se obtiene, mediante la raíz cuadrada de la varianza residual.*

$$S_e = \sqrt{\frac{\sum (y-\hat{y})^2}{n-p}} = \sqrt{\frac{\sum e^2}{n-p}} \tag{3.44}$$

Donde:

S_e= error estándar de estimación

$e = y - \hat{y}$ = error entre el valor observado y estimado de la variable dependiente.

n = número de grupos de la muestra.

$p = m+1$ = número de parámetros a estimar a partir de la muestra.

n-p = grados de libertad

La ecuación 3.44 es muy tediosa de calcular, por lo que los estadísticos han derivado una formula mas corta de calcular por el método computacional, la cual se muestra en la ecuación 3.45.

$$S_e = \sqrt{\frac{\sum y^2 - a_o \sum y - a_1 \sum x_1 y - a_2 \sum x_2 y - ... - a_m \sum x_m y}{n-p}} \tag{3.45}$$

3.5.2 Coeficiente de determinación múltiple

Representa la proporción de la variación total de ***y*** que es explicada por las variables involucradas en la ecuación de regresión múltiple, se puede calcular a partir de la ecuación 3.46 o 3.47.

$$R^2 = 1 - \frac{S_e^2}{S_y^2} \tag{3.46}$$

$$R^2 = \frac{a_o \sum y + a_1 \sum x_1 y + a_2 \sum x_2 y + ... + a_m \sum x_m y + n\bar{y}^2}{\sum y^2 - n\bar{y}^2} \tag{3.47}$$

donde:

R^2 = coeficiente de determinación.

Se = error estándar del estimado, calculando con las ecuaciones 3.45 o 3.46.

S^2_y = varianza de la variable dependiente ***y***.

$$S_y^2 = \frac{1}{n-1}(\sum (y - (\bar{y})^2)) = \frac{1}{n-1}\sum (y^2 - n(\bar{y})^2) \tag{3.48}$$

$\bar{y} = \frac{1}{n}\sum y$ media de la variable dependiente

n = número de grupos de la muestra

3.5.3 Coeficiente de correlación múltiple

El coeficiente de correlación múltiple, se puede calcular a partir de las ecuaciones 3.49 o 3.50.

$$R = \sqrt{1 - \frac{S_e^2}{S_y^2}} \tag{3.49}$$

$$R = \sqrt{\frac{a_o \sum y + a_1 \sum x_1 y + a_2 \sum x_2 y + ... + a_m \sum x_m y + n\overline{y}^2}{\sum y^2 - n\overline{y}^2}} \tag{3.50}$$

Ejemplo 3.3

Del estudio de una región de Bolivia, se ha obtenido para 14 subcuencas, el caudal anual (de los cuales máximos anuales) Q, en m^3/s, el área de la cuenca A, en km^2, y la intensidad máxima de precipitación I, en cm/24 hr , siendo los resultados los que se muestran en la tabla 3.2.

Se desea saber si éstas variables se correlacionan linealmente, es decir, si se puede establecer el modelo:

$$Q = a_o + a_1 A + a_2 I \tag{3.51}$$

Se pide:

1. Calcular el intercepto a_o *los coeficientes de regresión* a_1, a_2, y definir la ecuación de regresión lineal múltiple.

2. Calcular $\hat{Q}$, caudal estimado con la ecuación (3.51), para cada conjunto de valores ***A*** e ***I***.

3. Calcular los errores $e_i = Q - \hat{Q}$.

4. Calcular el error estándar de estimación (*Se*).

5. Calcular la varianza de la variable dependiente.

6. Calcular los coeficientes de determinación y correlación múltiple.

7. Estimar el valor de Q, si A = 4 km^2 e I = 1.5 cm/24 horas.

Estación	A (km2)	I (cm/24 horas)	Q (m2/s)
1	1.250	1.70	15.50
2	0.871	2.10	8.50
3	5.690	1.90	85.00
4	8.270	1.90	105.00
5	1.620	2.10	24.80
6	0.175	2.40	3.80
7	0.148	3.20	1.76
8	1.400	2.70	18.00
9	0.297	2.90	8.75
10	0.322	2.90	8.25
11	0.178	2.80	3.56
12	0.148	2.70	1.90
13	0.872	2.10	16.50
14	0.091	2.90	2.80

Tabla 3.2 Valore de A, I y Q para 14 subcuencas

Solución:

1. Cálculo de los para metros.

Para el ejemplo, utilizando las ecuaciones normales (3.44), se tiene:

$$\begin{aligned} \sum Q &= a_o n + a_1 \sum A + a_2 \sum I \\ \sum AQ &= a_o \sum A + a_1 \sum A^2 + a_2 \sum A \times I \\ \sum IQ &= a_o \sum I + a_1 \sum A \times I + a_2 \sum I^2 \end{aligned} \tag{3.52}$$

donde $n = 14$

En la tabla 3.3 se encuentra tabulado los valores de las $\sum$s , que reemplazado en el conjunto de ecuaciones 3.43, resulta:

$$304.12 = 14a_o + 21.332a_1 + 34.3a_2 \tag{3.53}$$

$$1465.8929 = 21.332a_o + 108.7412a_1 + 43.3419a_2 \tag{3.54}$$

$$627.80 = 34.3a_o + 43.3419a_1 + 86.99a_2 \tag{3.55}$$

En el sistema de ecuaciones 3.53, 3.54 y 3.55 se obtiene:

a_o = 1.656991

a_1 = 13.151048

a_2 = 0.011194

A (1)	I (2)	Q (3)	A x I (4)	I x Q (5)	2 (7)	2 (8)	Q^2 (9)
1.250	1.7	15.50	2.125	19.3750	26.3500	2.8900	240.2500
0.871	2.1	8.50	1.8291	7.4035	17.8500	4.4100	72.2500
5.690	1.9	85.00	10.8110	483.6500	161.500	3.6100	7225.0000
8.270	1.9	105.00	15.7130	868.3500	199.500	3.6100	11025.0000
1.620	2.1	24.80	3.4020	40.1760	52.0800	4.4100	615.0400
0.175	2.4	3.80	0.4200	0.6650	9.1200	5.7600	14.4400
0.148	3.2	1.76	0.4736	0.2605	5.6320	10.240	3.0976
1.400	2.7	18.00	3.7800	25.2000	48.6000	7.2900	324.0000
0.297	2.9	8.75	0.8613	2.5987	25.3750	8.4100	76.5625
0.322	2.9	8.25	0.9338	2.6565	23.9250	8.4100	68.0625
0.178	2.8	3.56	0.4984	0.6337	9.9680	7.8400	12.6736
0.148	2.7	1.90	0.3996	0.2812	5.1300	7.2900	3.6100
0.872	2.1	16.50	1.8312	14.3880	34.6500	4.4100	272.2500
0.091	2.9	2.80	0.2639	0.2548	8.1200	8.4100	7.8400
21.332	**34.3**	**304.12**	**43.3419**	**1465.8929**	**627.800**	**86.990**	**19960.0742**

Tabla 3.3 Valores para el cálculo de parámetros

Siendo la ecuación de regresión múltiple:

$$Q = 1.656991 + 13.151048A + 0.011194I \qquad (3.56)$$

2. Utilizando la ecuación 3.56 para los valores experimentales, se obtienen los valores estimados del caudal, los que se muestran en la columna (4) de la tabla 3.4.

3. La diferencia entre el valor experimental y el valor estimado del caudal con la ecuación 3.56, representa el error, estos valores se muestran en la columna (5) de la tabla 3.4.

A (1)	I (2)	Q experimental (3)	$\hat{Q}$ estimado (4)	$e = Q - \hat{Q}$ (5)	e^2 (6)
1.250	1.7	15.5	18.1148	-2.6148	6.8373
0.871	2.1	8.5	13.1351	-4.6351	21.4838
5.690	1.9	85	76.5077	8.4923	72.1188
8.270	1.9	105	110.4374	-5.4374	29.5657
1.620	2.1	24.8	22.9852	1.8148	3.2935
0.175	2.4	3.8	3.9853	-0.1853	0.0343
0.148	3.2	1.76	3.6392	-1.8792	3.5313
1.400	2.7	18	20.0987	-2.0987	4.4045
0.297	2.9	8.75	5.5953	3.1547	9.9520
0.322	2.9	8.25	5.9241	2.3259	5.4099
0.178	2.8	3.56	4.0292	-0.4692	0.2202
0.148	2.7	1.9	3.6336	-1.7336	3.0053
0.872	2.1	16.5	13.1482	3.3518	11.2345
0.091	2.9	2.8	2.8862	-0.0862	0.0074
				$\sum$	**171.0984**

Tabla 3.4 cálculo del caudal estimado y del error

4. De la ecuación 3.44 se tiene:

$$S_e = \sqrt{\frac{\sum e^2}{n-p}}$$

donde :

$\sum e^2$ =171.084, $\sum$ de la columna (6) de la tabla 3.4

n = 14, número de grupos de la muestra

p = 3, número de parámetros a estimas (a_o, a_1, a_2)

luego:

$$S_e = \sqrt{\frac{171.0984}{14-3}}$$

$S_e = 3.943907$

5. De la ecuación 3.48, se tiene:

$$S_Q^2 = \frac{1}{n-1}(\sum Q^2 - n\bar{Q}^2)$$

Donde:

$\sum Q^2 = 19960.0742$, Columna (9) de la tabla 3.4

$\bar{Q} = \frac{304.12}{14} = 21.722857$

$n = 14$

Reemplazando y resolviendo se obtiene:

$S_Q^2 = 1027.209152$

6. De la ecuación 3.46, se tiene:

$$R = \sqrt{1 - \frac{S_e^2}{S_Q^2}}$$

Donde:

$S_e^2 = 3.943907^2 = 15.5544$

S_Q^2 =1027.209152

Reemplazando y resolviendo se obtiene:

$R^2 = 0.984858$

$r = (0.984858)^{1/2} = 0.9924$

7. En la ecuación 3.56:

$Q = 1.656991 + 13.151048A + 0.011194I$

para :

$A = 4$

$I = 1.5$

se tiene:

$\hat{Q} = 1.656991 + 13.151048 \times 4 + 0.011194 \times 1.5$

$\hat{Q}$ =54.28 m^3/s

BIBLIOGRAFÍA

Villón Béjar, Máximo, Hidrológica Estadística; Lima Peru

Ríus Días , Francisca, Barón López, Francisco, Sánches Font, Elisa, Parras Guijosa, Luis, Bioestadística: métodos y aplicaciones, Manual de la Universidad de Málaga.

Moya Rufino – Saravia A. Gregorio. Estadística Descriptiva; Editorial Limusa

Romero Mauricio. Texto de Probabilidad y Estadística; U.M.S.S.

CAPÍTULO IV

TEORÍA DE LA PROBABILIDAD

CAPÍTULO IV

TEORÍA DE LA PROBABILIDAD

4.1 INTRODUCCIÓN

En ocasiones cuando se habla de probabilidad o posibilidad de que un evento ocurra, se pierde la credibilidad acerca del evento en cuestión, debido a que no es posible tener la certeza total de su ocurrencia, en si por que el llevar a efecto dicho evento o proyecto en particular por más simple que este sea está sujeto a una gran diversidad de factores que afectan su ocurrencia, es aquí donde la *probabilidad* es de gran ayuda, ya que basándose en datos, se cuantifica la posibilidad de ocurrencia de los eventos y por consiguiente se toma una buena decisión basada en esta información.

Es así que la probabilidad se encarga de evaluar todas aquellas actividades en las cuales se tiene incertidumbre acerca de los resultados que se pueden esperar, esto quiere decir que la probabilidad está presente en casi todas las actividades que se pretenda realizar.

Ejemplos:

- Cualquier proyecto de Ingeniería o de otras áreas
- Competencias deportivas
- Juegos de azar, etc.

En éste capítulo, como introductorio a la probabilidad, se definen términos que son usados mas adelante y facilitan la comprensión más exacta de todo el tema en general, se determinan además los axiomas, la regla de la multiplicación como también de la regla de la adición, y las técnicas de conteo que son herramientas usadas para determinar la probabilidad de ocurrencia de distintos casos que se puedan presentar.

4.2 CONCEPTOS Y DEFINICIONES

4.2.1 Definiciones previas

Antes de iniciar con lo que es la aplicación de la probabilidad en sí y otros subtítulos de interés mayor es imprescindible conocer algunas definiciones que se van a utilizar a lo largo de este capítulo.

Definición 4.1 ***Experimento*** *es un proceso para obtener los resultados posibles que sean de interés en la investigación.*
Ej.: Determinar el acelerado de un coche desde el reposo hasta 70 Km/h.

Los experimentos se dividen en:

- ***Deterministicos,*** *cuando los resultados del experimento son ya determinados y pueden describirse por una formula matemática (modelo matemático).*
  Ej.: "Soltar una piedra en el aire", pues la piedra caerá y su movimiento esta determinado por las ecuaciones de caída libre.

- ***No deterministico****, cuando al repetirlo en condiciones análogas no se puede predecir el resultado.*
  Ej.: "Lanzar una moneda y observar sello".

Definición 4.2 *Un* ***Experimento aleatorio (ε)*** *es no deterministico, pudiendo repetirse indefinidamente sin cambiar esencialmente las condiciones.*

Definición 4.3 ***Espacio muestral*** **(Ω)** *es el conjunto de todos los resultados posibles del experimento aleatorio.*

Los espacios muestrales según su número se clasifican en:

- ***Finito:*** *Si el número de elementos que tiene* Ω *está acotado.*

- ***Infinito numerable:*** *Cuando, a pesar de tener infinitos elementos, no siempre es posible intercalar uno entre dos dados.*

  Ej.: Nº de veces que hay que lanzar un dado hasta que salga un 6. Este número, en teoría es ilimitado, pero nunca puede estar entre 5 y 6 (siempre será entero).

- ***Infinito no numerable:*** *Cuando* Ω *tiene infinitos elementos, y además siempre se puede intercalar uno entre dos cualquiera de ellos.*

  Ej.: Tiempo de espera hasta que un paciente que acude a urgencias es atendido.

  Esta clasificación permite denominar a los espacios muestrales de la siguiente manera:

- ***Espacios muestrales discretos:*** *Si tienen un número finito o infinito numerable de elementos.*

- ***Espacios muestrales continuos:*** *Si tiene un número no numerable de elementos.*

Definición 4.4 ***Evento*** *es el subconjunto del espacio muestral.*

Ejemplo 4.1

En el experimento aleatorio "lanzar cuatro monedas simultáneamente". Se definen los eventos siguientes:

A: "Todas las monedas muestran el mismo lado"

B: "Ocurren por lo menos dos caras"

C: "Ocurre sello en el tercer lanzamiento"

Definir el espacio muestral de cada evento.

Solución:

El espacio muestral de los eventos A, B y C son respectivamente:

Ω (A) = (CCCC, SSSS)

Ω (B) = (CCCC, CCCS, CCSC, CSCC, SCCC, CSCS, CSSC, SCCS, SCSC, SSCC)

Ω (C) = (CCSC, CSSC, CSSS, SCSC, SCSS, SSSS, CCSS, SSSC)

Resumiendo:

El experimento aleatorio "lanzar una moneda cuatro veces"

Los espacios muestrales son los conjuntos Ω (A), Ω (B) y Ω (C).

Los eventos son A, B, C.

4.2.2 Conceptos

Para tener un concepto claro y amplio de lo que es la probabilidad se debe analizar tres definiciones las cuales son complementarias entre sí. Cada cual se ha de aplicar dependiendo la naturaleza del problema específico que se trate de resolver, las definiciones o conceptos son:

4.2.2.1 Según la definición clásica o a priori

La definición clásica de probabilidad fue dada por Laplace y consiste básicamente en determinar la probabilidad de que ocurra un evento dado que no ha sucedido, basándose en el supuesto de que todos los resultados posibles de un experimento aleatorio son *igualmente probables*; es decir, cada uno de los elementos del espacio muestral (Ω) tienen la misma posibilidad de ocurrencia. Se concluye, por tanto, que la definición clásica es:

Definición 4.5 ***La probabilidad*** *de un evento es la razón entre el número de casos (sucesos) favorables y el número total de casos (sucesos) posibles, siempre que nada obligue a creer que alguno de estos sucesos debe tener preferencia a los demás, lo que hace que todos sean igualmente posibles.*

$$\mathrm{P(A)} = \frac{\text{nº de casos favorables}}{\text{nº de casos posibles}} \qquad (4.1)$$

$$P(A) = \frac{n_A}{n} \qquad (4.2)$$

4.2.2.2 Según la definición frecuentista o a posteriori

Consiste en determinar la probabilidad de ocurrencia de un evento que ya sucedió, debido a esto su obtención se basa en la experiencia. La definición frecuentista de probabilidad es la siguiente:

Definición 4.6 *Si un experimento bien definido se repite n veces (n tan grande que tiende a infinito); sea $n_A < n$ el número de veces que el evento A ocurre en los n ensayos, entonces la frecuencia relativa de veces que ocurre el evento A, es la* *estimación de **la probabilidad** que ocurra el evento A, es decir:*

$$P(A) = \frac{n_A}{n} \tag{4.3}$$

Teniendo en cuenta que $n \to \infty$, aunque en la practica sólo se puede buscar una estimación máxima de *P(A)* basada en *n* grande.

4.2.2.3 Según la definición subjetiva

El enfoque subjetivo de la probabilidad se aplica a casos en que hay sólo una oportunidad de ocurrencia de un determinado evento, pudiendo o no ocurrir esa sola vez.

Definición 4.7 ***La probabilidad** de un evento A, es el grado de creencia asignado a la ocurrencia de este evento por un individuo particular, basado en toda la evidencia a su disposición, tomando en cuenta:*

- *P(A) = 0, representa la certeza que el evento A, no ocurrirá*
- *P(A) = 1, representa la certeza que el evento A, si ocurrirá*
- *0 < P(A) < 1; representa el grado de certeza que el evento A, ocurrirá*

El grado de creencia es un número que define según la evidencia que posee la persona que efectúa la asignación, por tanto este número puede variar.

En los ejemplos siguientes indistintamente si tienen o no solución, se quiere diferenciar cada una de las definiciones anteriores.

Ejemplo 4.2

Ejemplo del enfoque a priori o clásico

Se elige una carta aleatoriamente de una baraja de 52 cartas.

a) ¿Cuál es la probabilidad que sea un palo negro (espadas o tréboles)?

b) ¿Cuál que sea un diez?

c) ¿Cuál que sea una figura (Rey, Reina, Sota)?

d) ¿Cuál es la probabilidad que sea un cuatro o menos?

Solución:

El experimento aleatorio es "extraer una carta de 52"

El espacio muestral Ω tiene 52 elementos. Osea $n = 52$

a) Sea el evento A: "Obtener un palo negro".

A tiene 26 elementos (13 espadas + 13 tréboles). Osea $n_A = 26$. Luego,

$$P(A) = \frac{26}{52} = \frac{1}{2} = 0.5$$

b) Sea el evento B: "Obtener un diez"

B tiene 4 elementos (pues hay 4 dieses). Osea $n_B = 4$. Entonces,

$$P(B) = \frac{4}{52} = \frac{1}{13} = 0.077$$

c) Sea el evento C: "Obtener una figura"

C tiene 12 elementos (4 Reyes + 4 Reinas y 4 Sotas). Osea $n_C = 12$. Luego,

$$P(C) = \frac{12}{52} = \frac{3}{13} = 0.33$$

d) Sea el evento D: "Obtener un cuatro o menos"

D tiene 16 elementos (4 Ases + 4 Dos + 4 Tres + 4 Cuatros). Osea $n_D = 16$. Luego,

$$P(D) = \frac{16}{52} = \frac{4}{13} = 0.39$$

Ejemplo 4.3

Ejemplo del enfoque a posteriori o frecuentista.

Una muestra aleatoria de 10 fábricas que emplean un total de 10000 personas, demostró que ocurrieron 500 accidentes de trabajo durante un periodo reciente de 12 meses. Hallar la probabilidad de un accidente de trabajo en una industria determinada.

Solución:

n = 10000, numero de veces que se repite el experimento

Sea el evento A: "Un accidente de trabajo en la industria determinada"

Entonces $n(A) = 500$ y:

$$P(A) = \frac{500}{10000} = 0.05$$

Por definición de frecuencia relativa, ya que este valor de la probabilidad se basa en una muestra, por lo tanto es una estimación del valor real desconocido. Obsérvese, aquí se supone implícitamente que las formas de seguridad no han cambiado desde que se realizó el muestreo.

Ejemplo 4.4

Ejemplo de enfoque subjetivo:

Considere a una persona que se plantea la hipótesis “el dado está balanceado” y no sabe nada de probabilidades. Esta persona ejecuta 10 lanzamientos y obtiene puros unos, entonces el jugador infiere, haciendo uso de su sentimiento intuitivo, que el dado está cargado, dado que sabe que ocurra tal resultado es altamente improbable utilizando un dado normal. ¿Cuál sería su decisión si se hubieran obtenido cinco unos, dos veces el tres y una vez el 2, 4 y 6?

En ésta situación no es tan fácil tomar una decisión.

¿Cuál sería su decisión sí el resultado fuese cuatro unos, dos veces el tres y una vez el 2, 4,

5 y 6? El decidir sí el dado estaría cargado, tomando como base los resultados anteriores, prácticamente se estaría adivinando.

4.2 AXIOMAS REFERENTES A LA PROBABILIDAD

Independientemente de la forma como se define la probabilidad, ésta debe cumplir los axiomas siguientes, donde Ω, es el espacio muestral de un experimento ε, y A y B dos eventos cualesquiera:

Axioma 1: La probabilidad es un número real, en el intervalo de 0 a 1

$$P:\varepsilon \to [0,1] \subset \mathbb{R} \tag{4.4}$$

Axioma 2: El espacio muestral (Ω es cierto), como un todo, tiene probabilidad 1.

$$P(\Omega) = 1 \tag{4.5}$$

Axioma 3: Funciones de probabilidad son aditivas. Siendo A y B mutuamente excluyentes.

$$Si\ A, B \in \varepsilon, A \cap B = \varnothing \Rightarrow P(A \cup B) = P(A) + P(B) \tag{4.6}$$

Generalizando:

$$Si\ \{A_i\}_{i=1}^{n} \in \varepsilon, A_i \cap A_j = \varnothing \Rightarrow P(\cup A_i) = \sum_i P(A_i) \tag{4.7}$$

4.4 PROBABILIDAD CON EVENTOS EXCLUYENTES Y NO EXCLUYENTES (REGLA DE LA ADICIÓN)

Sean dos eventos A y B no excluyentes; si *(A + B)* denota el evento de que "ocurra *A* ó *B* ó ambos", se tiene:

$$P(A \cup B) = P(A) + P(B) - P(A\,B) \tag{4.8}$$

Dos o más eventos son mutuamente excluyentes si no pueden ocurrir juntos, es decir, la ocurrencia de uno imposibilita la ocurrencia de los otros. Así, si A y B son mutuamente excluyentes, $P(A\,B) = 0$ y la ecuación 5.8 se convierte en:

$$P(A \cup B) = P(A) + P(B) \tag{4.9}$$

Ejemplo 4.5

Sea el evento A "extracción de un as de una baraja de cartas" y el evento B "extracción de un rey" ¿Cuál es la probabilidad de extracción de un as o un rey en una sola extracción?

Solución:

Como as y rey no pueden extraerse al mismo tiempo en una sola extracción, son eventos mutuamente excluyentes:

$$P(A) = \frac{4}{52} = \frac{1}{13}$$

$$P(B) = \frac{4}{52} = \frac{1}{13}$$

$$P(A \cup B) = P(A) + P(B) = \frac{1}{13} + \frac{1}{13} = \frac{2}{13}$$

Ejemplo 4.6

Sea el evento A "extracción de un as de una baraja de cartas" y el evento B "extracción de una espada" ¿Cuál es la probabilidad de extracción de un as de espadas?

Solución:

En este ejemplo A y B no son mutuamente excluyentes, puesto que debe ser extraído el as de espadas:

$$P(A)=\frac{4}{52}$$
$$P(B)=\frac{13}{52}$$
$$P(A\,B)=\frac{1}{52}$$
$$P(A\cup B)=P(A)+P(B)-P(A\,B)$$
$$P(A\cup B)=\frac{4}{52}+\frac{13}{52}+\frac{1}{52}=\frac{4}{13}$$

4.5 PROBABILIDAD CONDICIONAL, EVENTOS INDEPENDIENTES Y DEPENDIENTES (REGLA DE LA MULTIPLICACIÓN)

Sean dos eventos *A* y *B*, la probabilidad de que ocurra el evento *A*, dado que ha ocurrido *B*, denotado *P [A/B]* ó *P(B dado A)* se llama *probabilidad condicional* de *B* dado que *A* se ha presentado. Esto implica que las probabilidades de *A* y *B*, son *dependientes* entre sí. En otras palabras, la información con respecto a la ocurrencia de *B* afectará la probabilidad de *A*.

En el caso de que la ocurrencia de *B* no tenga ningún efecto sobre la probabilidad de *A*, se origina el concepto de *independencia estadística*.
De los conceptos anteriores se tiene:

$$P(A\cap B)=P(B)\times P(A/B) \tag{4.10}$$

$$P(A\cap B)=P(A)\times P(B/A) \tag{4.11}$$

Para sucesos independientes:

$$P(A\,B) = P(A) \times P(B) \qquad (4.12)$$

Para el caso particular en que existan tres eventos A , B y C:

$$P(A\,B\,C) = P(A) \times P(B/A) \times P(C/AB) \qquad (4.13)$$

Para sucesos independientes:

$$P(A\,B\,C) = P(A) \times P(B) \times P(C) \qquad (4.14)$$

Ejemplo 4.7

De una urna que contiene 9 bolas rojas y 5 negras, se extraen sucesivamente 2 bolas. Calcular la probabilidad de los siguientes eventos:

a) Que las dos sean negras
b) Que las dos sean rojas
c) Que la primera sea roja y la segunda negra
d) Que la segunda sea roja sabiendo que la primera fue negra

Solución:

a) Sea N_1: Sacar la 1ª Negra
N_2: Sacar la 2ª Negra

$P(N_1 \cap N_2) = P(N_1)$ x $P(N_2/N_1) =$ 5/14 x 4/13 = 0.109

b) Sea R_1: Sacar la 1ª Roja

R_2: Sacar la 2ª Roja

$$P(R_1 \cap R_2) = P(R1) \text{ x } P(R_2/R_1) = 9/14 \text{ x } 8/13 = 0.39$$

c) Sea R_1: Sacar la 1ª Roja

N_2: Sacar la 2ª Negra

$$P(R_1 \cap N_2) = P(R_1) \text{ x } P(N_2/R_1) = 9/14 \text{ x } 5/13 = 0.247$$

d) Sea N_1: La 1ª es Negra

R_2: La 2ª es Roja

$$P(R_2/N_1) = 9/13 = 0.692$$

Ejemplo 4.7

Considerar el experimento de extraer cartas de una baraja ¿Cuál es la probabilidad de extraer dos reyes?

a) Sin devolver la 1ª carta.
b) Con devolución.

Solución:

a) R_1: "conseguir rey en la 1ª extracción"

R_2: "conseguir rey en la 2ª extracción"

$$P(R_1 \cap R_2) = P(R_1) \text{ x } P(R_2/R_1) = 4/40 \text{ x } 3/39 = 0.0077$$

b) $P(R_1 \cap R_2) = P(R_1) \text{ x } P(R_2) = 4/40 \text{ x } 4/40 = 0.01$

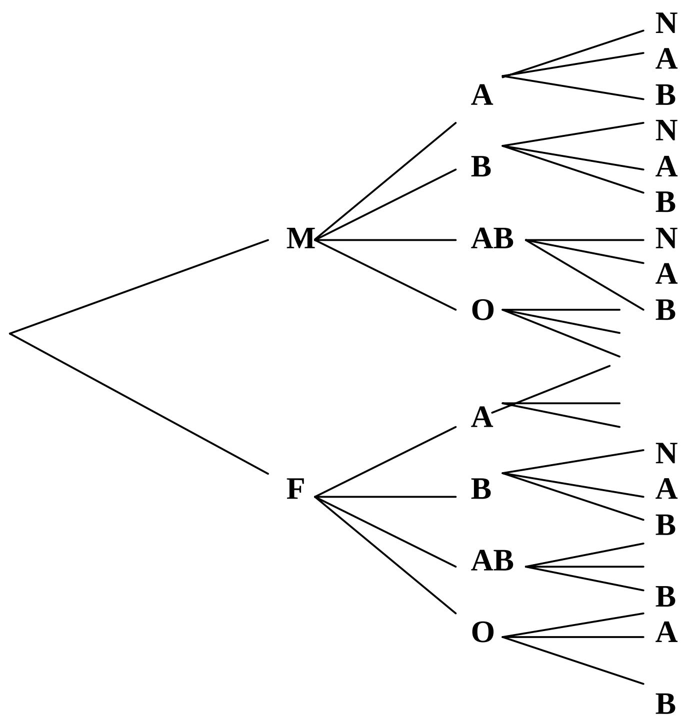
4.6 DIAGRAMA DEL ÁRBOL

Es un diagrama de líneas compuestas, en las cuales se anotan las probabilidades que se presentan, habiendo tantas líneas como probabilidades haya. Se emplea cuando se tiene un experimento aleatorio con sucesos compuestos de dos o más sucesos simples.

Ejemplo 4.9

Un médico general clasifica a sus pacientes de acuerdo a su sexo (masculino o femenino), tipo de sangre (A, B, AB u O) y en cuanto a la presión sanguínea (Normal, Alta o Baja). Mediante un diagrama de árbol diga en cuantas clasificaciones pueden encontrarse los pacientes de este médico?

Solución:

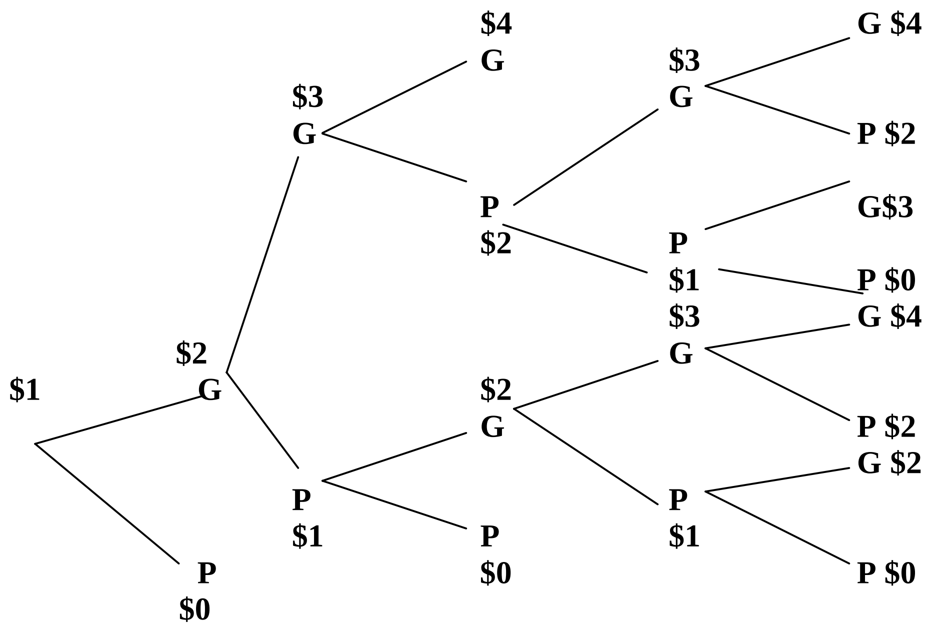


Al contar todas las ramas terminales, se ve que el número de clasificaciones son 2 x 4 x 3 = 24, mismas que se puede enumerar: MAN, MAA, MAB, MBN, MBA, MBB, etc.

Ejemplo 4.10

Un hombre tiene tiempo de jugar ruleta cinco veces como máximo, él empieza a jugar con un dólar, apuesta cada vez un dólar y puede ganar o perder en cada juego un dólar, él se va a retirar de jugar si pierde todo su dinero, si gana tres dólares (esto es si completa un total de cuatro dólares) o si completa los cinco juegos, mediante un diagrama de árbol, ¿cuántas maneras hay de que se efectué el juego de este hombre?

Solución:



Si se cuentan las ramas terminales hay 11 maneras de que este hombre lleve a cabo sus apuestas, en este diagrama se han representado los cinco juegos o apuestas que este hombre tiene tiempo de jugar.

4.7 TEOREMA DE LA PROBABILIDAD TOTAL

Sea $A_1, A2, \ldots, A_n$ un sistema completo de sucesos tales que la probabilidad de cada uno de ellos es distinta de cero, y sea B un suceso para el que se conocen las probabilidades $P(B/A_i)$, entonces la probabilidad del suceso B viene dada por :

$$P(B) = \sum_{i=1}^{n} P(A_i) \times P(B/A_i) \tag{4.15}$$

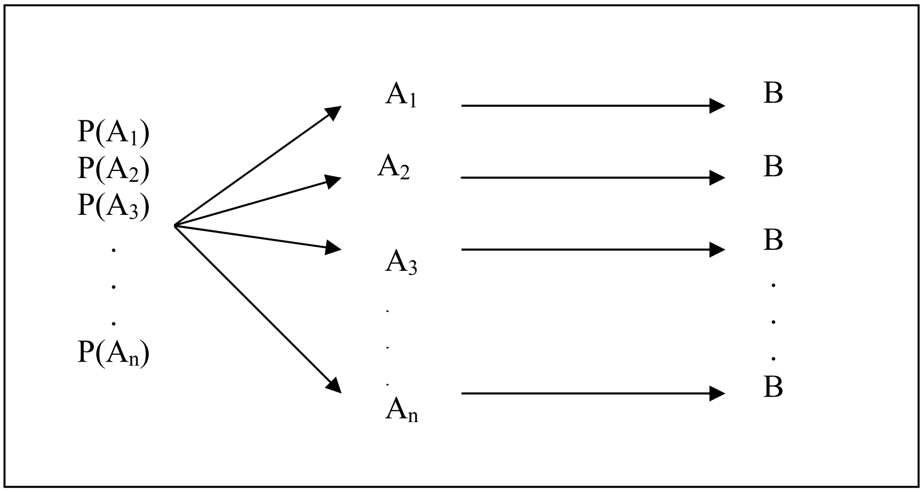


Figura 4.1 Diagrama de árbol de probabilidad total

4.8 TEOREMA DE BAYES

Sea $A_1, A_2, \ldots, A_n$ un sistema completo de sucesos tal que la probabilidad de cada uno de ellos es distinta de cero, y sea *B* un suceso cualquiera para el que se conocen las probabilidades $P(B/A_i)$, entonces :

$$P(A_i / B) = \frac{P(A_i) \times P(B / A_i)}{\sum_{i=1}^{n} P(A_i) \times P(B / A_i)} \qquad ;\ i = 1, 2, \ldots, n \tag{4.16}$$

Ejemplo 4.11

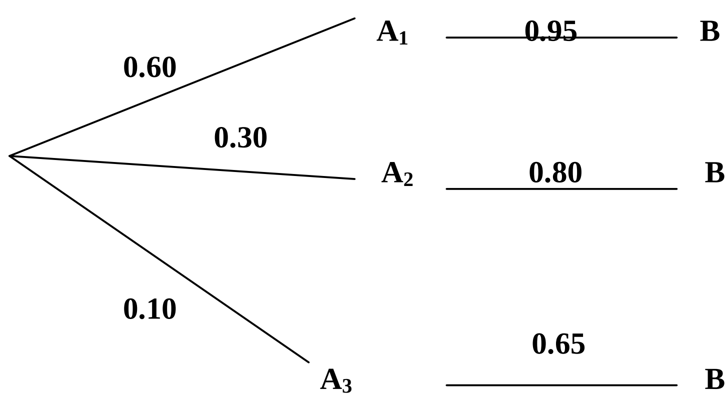
Una planta de ensamblaje recibe reguladores de voltaje de 3 diferentes proveedores, 60% para proveedor A_1, 30% para A_2, 10% para A_3. Si 95% de los reguladores de voltaje de A_1, 80% de los A_2, y 65% de los A_3 cumplen las especificaciones, calcular la probabilidad de que cualquier regulador de voltaje recibido por la planta cumplirá las especificaciones

Solución:

Aplicando el teorema de la probabilidad total:

$$P(B) = (0.60)(0.95) + (0.30)(0.80) + (0.10)(0.65)$$
$$P(B) = 0.875 \ \acute{o} \ 87.5\%$$

El problema se ve graficado según el diagrama de árbol:



Si además, se quiere que un regulador de voltaje específico, que se sabe funcione según especificaciones, venga del proveedor A_3:

$$P(A_3 / B) = \frac{(0.10)(0.65)}{(0.60)(0.95) + (0.30)(0.80) + (0.10)(0.65)}$$

$$P(A_3 / B) = 0.074 \text{ ó } 7.4\%$$

Se observa que la probabilidad de que el regulador de voltaje es proveído por A_3 disminuye, una vez que se conoce que trabaja de acuerdo a especificaciones.

4.9 TÉCNICAS DE CONTEO

4.9.1 Principio de multiplicación

Si un experimento aleatorio, ε_1 ocurre de n_1 formas y si para cada una de éstas, un experimento ε_2 ocurre de n_2 formas, y el *r*-ésimo experimento ε_r paso de n_r formas, entonces esta actividad puede ocurrir de $n_1 \times n_2 \times \ldots \times n_r$ *formas*.

El principio multiplicativo implica que cada uno de los pasos de la actividad deben ser llevados a efecto, uno tras otro.

Ejemplo 4.12

Una persona desea construir su casa, para lo cual considera que puede construir los cimientos de su casa de cualquiera de dos maneras (concreto o block de cemento), mientras que las paredes las puede hacer de adobe, adobón o ladrillo, el techo puede ser de concreto o lámina galvanizada y por último los acabados los puede realizar de una sola manera ¿cuántas maneras tiene esta persona de construir su casa?

Solución:

Considerando que $r = 4$ pasos

$n_1 =$ maneras de hacer cimientos = 2

$n_2 =$ maneras de construir paredes = 3

$n_3 =$ maneras de hacer techos = 2

$n_4 =$ maneras de hacer acabados = 1

$n_1 \times n_2 \times n_3 \times n_4 = 2 \times 3 \times 2 \times 1 = 12$

Lo que quiere decir que existen 12 maneras de construir la casa.

Ejemplo 4.13

¿Cuántas placas para automóvil pueden ser diseñadas si deben constar de tres letras seguidas de cuatro números, si las letras deben ser tomadas del abecedario y los números de entre los dígitos del 0 al 9? Resolver considerando que:

a) Resolver considerando que es posible repetir letras y números.

b) Resolver considerando que no es posible repetir letras y números.

c) Cuántas de las placas diseñadas en el inciso b empiezan por la letra D y además empiezan por el cero.

d) Cuantas de las placas diseñadas en el inciso b empiezan por la letra D seguida de la G.

Solución:

a) Considerando 26 letras del abecedario y los dígitos del 0 al 9
26 x 26 x 26 x 10 x 10 x 10 x 10 = 75 760 000

b) 26 x 25 x 24 x 10 x 9 x 8 x 7 = 78 624 000

c) 1 x 25 x 24 x 1 x 9 x 8 x 7 = 302 400

d) 1 x 1 x 24 x 10 x 9 x 8 x 7 = 120 960

4.9.2 Permutaciones

Una permutación es un arreglo de todos o parte de un conjunto de objetos distintos, en la que una permutación difiere de otra si el orden del arreglo o el contenido difieren. Si se selecciona permutaciones de r objetos a partir de n objetos distintos, se tiene:

$$_nP_r = (n)(n-1)(n-2)(n-3)...(n-r+1) \tag{4.17}$$

O bien:

$$_nP_r = \frac{n!}{(n-r)!} \tag{4.18}$$

Ejemplo 4.14

a) ¿Cuántas maneras diferentes hay de asignar las posiciones de salida de 8 autos que participan en una carrera de fórmula uno? (Considere que las posiciones de salida de los autos participantes en la carrera son dadas totalmente al azar).

b) ¿Cuántas maneras diferentes hay de asignar los primeros tres premios de esta carrera de fórmula uno?

Solución:

a) Utilizando el principio multiplicativo:

8 x 7 x 6 x 5 x 4 x 3 x 2 x 1 = 40 320

Utilizando la ecuación 5.18: $n = 8$, $r = 8$

${}_8P_8 = 8! = 40\ 320$

b) Utilizando el principio multiplicativo:

8 x 7 x 6 = 336

Utilizando la ecuación 5.18: $n = 8$, $r = 3$

${}_8P_3 = 8! / (8 - 3)! = 8! / 5! = 336$

4.9.3 Combinaciones

Una combinación es un arreglo de objetos distintos, sin importar el orden. Si se desea determinar el número de combinaciones cuando en n objetos distintos deben seleccionarse r a la vez, se expresa:

$$ {}_nC_r = \frac{n(n-1)(n-2)...(n-r+1)}{r!} \tag{4.19}$$

$$ {}_nC_r = \frac{n!}{(n-r)!r!} \tag{4.20}$$

De la ecuación 4.20 se deriva:

$$ {}_nC_r = \frac{{}_nP_r}{r!} \tag{4.21} $$

Ejemplo 4.15

Para contestar un examen un alumno debe contestar 9 de 12 preguntas.

a) ¿Cuántas maneras tiene el alumno de seleccionar las 9 preguntas?
b) ¿Cuántas maneras tiene si forzosamente debe contestar las 2 primeras preguntas?
c) ¿Cuántas maneras tiene si debe contestar una de las 3 primeras preguntas?
d) ¿Cuántas maneras tiene si debe contestar como máximo una de las 3 primeras preguntas?

Solución:

a) $n = 12$, $r = 9$

$${}_{12}C_9 = 12! / (12 - 9)!9!$$
$$= 220$$

Existen 220 maneras de seleccionar las nueve preguntas o dicho de otra manera, el alumno puede seleccionar cualquiera de 220 grupos de 9 preguntas para contestar el examen.

b) ${}_2C_2 \text{ x } {}_{10}C_7 = 1 \text{ x } 120 = 120$

Existen 120 maneras de seleccionar las 9 preguntas entre las que están las dos primeras preguntas

c) ${}_3C_1 \times {}_9C_8 = 3 \times 9 = 27$

Existen 27 maneras de seleccionar las 9 preguntas entre las que está una de las tres primeras preguntas

d) En este caso debe seleccionar 0 ó 1 de las tres primeras preguntas

${}_3C_0 \times {}_9C_9 + {}_3C_1 \times {}_9C_8 = (1 \times 1) + (3 \times 9) = 28$

Existen 28 maneras de seleccionar las preguntas a contestar.

BIBLIOGRAFÍA

Moya C. Rufino – Saravia A. Gregorio. Probabilidad e inferencia estadística; Editorial Limusa.

Chungara Castro Víctor. Estadística y Probabilidades; Editorial Leonardo.

Romero Mauricio. Texto de Probabilidad y Estadística; U.M.S.S.

Manchego C. Roberto. Estadística II; U.M.S.S.

CAPITULO V

VARIABLE ALEATORIA

CAPÍTULO V

VARIABLE ALEATORIA

5.1 VARIABLES ALEATORIAS

En este capítulo se define los dos tipos de variables aleatorias que existen, las variables aleatorias discretas y las variables aleatorias continuas, a continuación se detallan primero las funciones de distribución y las funciones de densidad para variables aleatorias discretas y continuas respectivamente, además se muestra la diferencia que existen entre ambas variables para las funciones de distribución, sus propiedades, sus características, propiedades del valor esperado, momentos de orden superior y asimetría. Al poder comparar todo esto para una y otra variable, se distingue la diferencia específica que existe entre ambas variables aleatorias, las discretas y las continuas.

Definición 5.1 *Dado un experimento aleatorio ε y* **Ω** *el espacio muestral asociado a ε; una función X que asigna a cada elemento ω en* **Ω** *uno y solamente un número real x = X (ω), se llama* ***variable aleatoria****. A una variable aleatoria, se le conoce también como variable estocástica, sus valores son números reales, que no pueden predecirse con certeza antes de ocurrir el fenómeno, es decir, ocurren al azar. Siendo X una función real se tiene que;* $X : \Omega \longrightarrow \Re$.

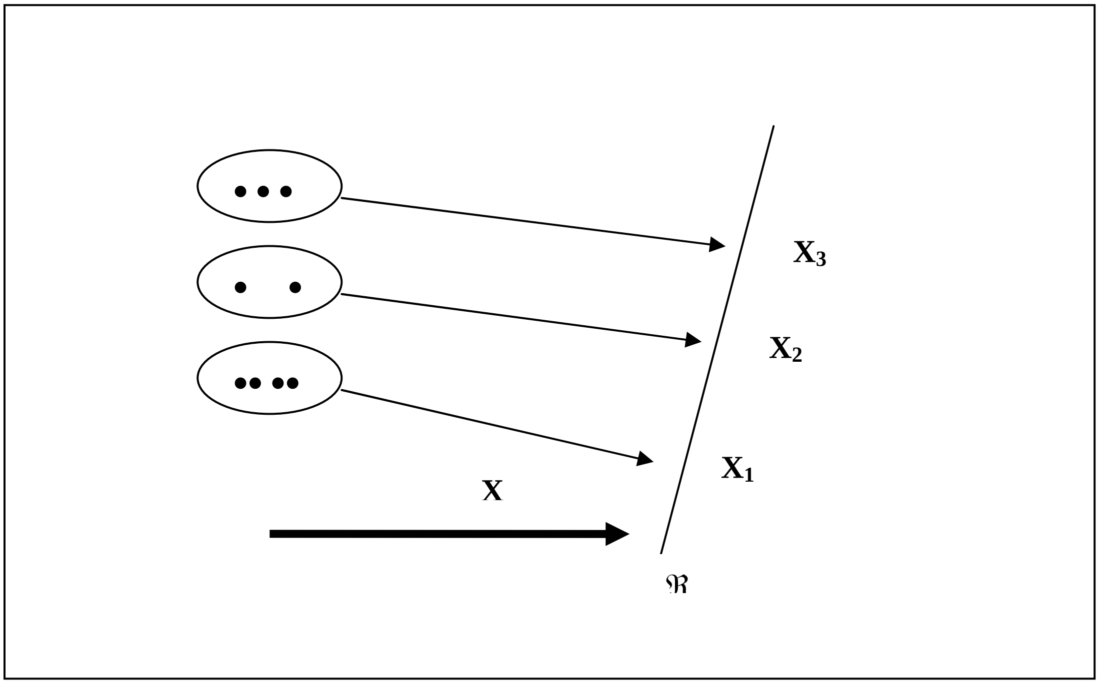


Figura 5.1 Dominio de la variable aleatoria X

Ejemplo 5.1

En el experimento aleatorio "estado civil de un docente de la universidad", se define la variable aleatoria, X de modo que:

X *(soltero)* = 1 X *(casado) = 2* X *(divorciado)* = 3 X *(viudo)* = 4

Se dice, entonces, que X es una variable aleatoria que toma los valores 1, 2, 3 y 4.

Las variables aleatorias se clasifican usualmente de acuerdo al número de valores que puede asumir, es así, que su clasificación es: Variable aleatoria discreta y variable aleatoria continua.

5.2 VARIABLE ALEATORIA DISCRETA

Definición 5.2 *Si el rango de la variable aleatoria X, es un conjunto finito o infinito numerable, se llama* ***variable aleatoria discreta.***

Es así que:

$$R_x = \{x_1, x_2, x_3 ...\} \tag{5.1}$$

Un ejemplo de variable aleatoria discreta es el número de días de lluvias ocurridas en los meses de un año cualquiera.

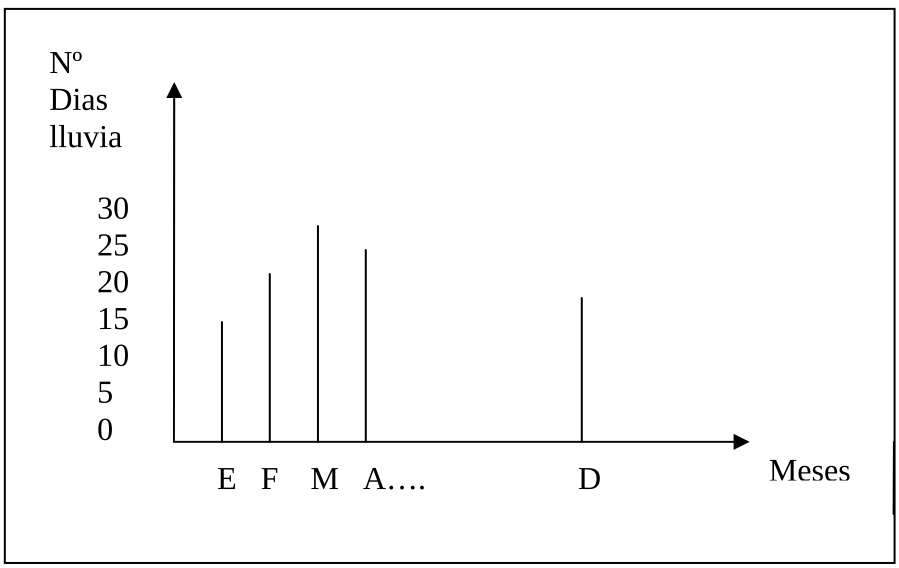


Figura 5.2 Ejemplo de variable aleatoria discreta

Ejemplo 5.2

Suponiendo que el número de días de trabajo en un año particular es 280 y los records de los empleados se marcan cada día que ellos están ausentes del trabajo. Se selecciona aleatoriamente un record y se observa los días marcados. La variable aleatoria X se define como el número de días ausentes del trabajo, entonces $R_x = \{0,1,2,...,280\}$. Luego, X es una variable aleatoria discreta con un número finito de posibles valores.

5.2.1 Función de probabilidad de una variable aleatoria discreta

Definición 5.3 *Si X es una variable aleatoria discreta, se llama* ***Función de probabilidad*** *de X a la función $f(x_i)$ definida por:*

$$f(x_i) = P[X = x_i] \tag{5.2}$$

Donde $f(x_i)$ debe satisfacer las siguientes condiciones:

a. $\sum_{i=1}^{n} f(x_i) = 1 \qquad ; \forall x_i \in R_X \qquad (5.3)$

b. $f(x_i) \geq 0 \qquad ; \forall x_i \in R_X \qquad (5.4)$

Definición 5.4 ***Distribución de probabilidad*** *es el resultado de asignar valores de probabilidad a todos los valores numéricos posibles de dicha variable aleatoria, ya sea mediante un listado o a través de una función matemática.*

Una distribución de probabilidad se representa por medio de tablas (tabla 5.1) y gráficamente (figura 5.3).

x_i	$f(x)=P[X=x]$
x_1	$p(x_1)$
x_2	$p(x_2)$
x_3	$p(x_3)$
.	.

Tabla 5.1 Representación tabular de la distribución de probabilidad

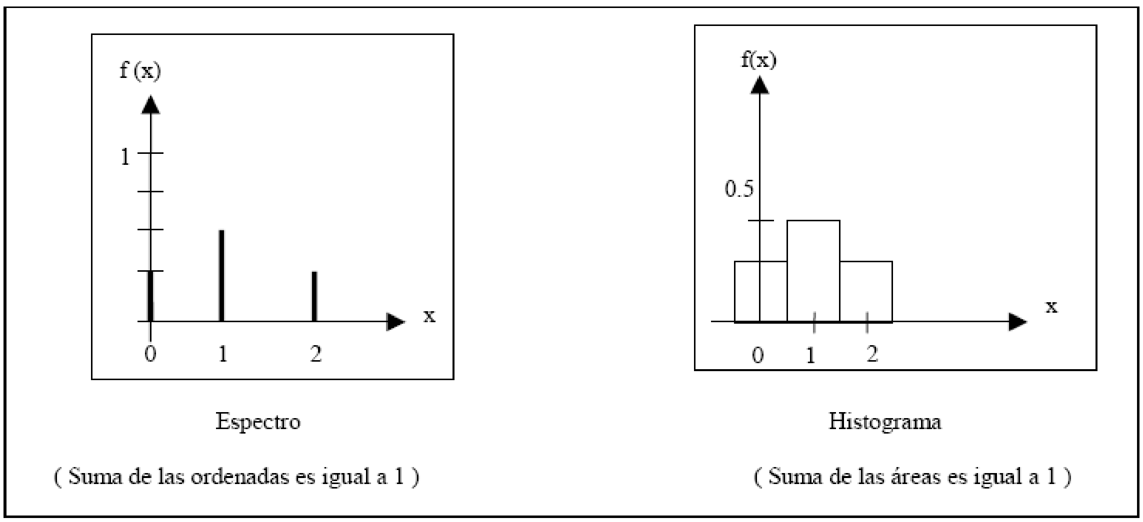


Figura 5.3 Representación grafica de una Distribución de probabilidades Del ejemplo explicativo Diagrama de barras (espectro) y un histograma

5.2.2 Función de distribución de una variable aleatoria discreta

La función de distribución o función de distribución acumulada de la variable aleatoria X denotada por *F(x)* está definida por:

$$F(x) = P(X \leq x) = \sum_{i=1}^{\forall i \to x} p(x_i) \ ; \qquad \forall x_i \in R_x \tag{5.5}$$

5.2.3 Propiedades de la función de distribución para una variable aleatoria discreta

1. $0 \leq F(x) \leq 1; \ \forall \ x \ \in \ \mathbb{R}_x$ (5.6)

2. *F(x)* es una función no decreciente:

$x_1, x_2 \in \ \mathbb{R}_x$ tales que $x_1 \leq x_2 \Rightarrow F(x_1) \leq F(x_2)1; \ \forall \ x \ \in \ \mathbb{R}_x$ (5.7)

3.

a) $\lim_{x \to \infty} F(x) = P[x / X < \infty] = P[X < \infty] = 1$ (5.8)

b) $\lim_{x \to -\infty} F(x) = P[x / X < -\infty] = P[X < -\infty] = 0$ (5.9)

Ejemplo 5.3

Para cada una de las siguientes funciones, determinar la constante *k* para que *f(x)* sea una función de probabilidad de una variable aleatoria *X*.

a) $f(x) = xk, \qquad x = 1, 2, 3..., 10$

b) $f(x) = k\left(\frac{1}{3}\right)^x, \qquad x = 1, 2, 3, ...$

Solución:

a) Para que *f(x)* sea una función de probabilidad debe cumplir la definición de ésta

1. $f(x) = kx > 0, \qquad \forall x = 1, 2, \ldots, 10, \quad si, solo\ si\ k > 0$
2. $\sum_{x=1}^{10} kx = k[1+2+3+4+5+6+7+8+9+10] = 1$

De donde k = 1/55. Entonces,

$$f(x) = \frac{x}{55}, \qquad x = 1, 2, \ldots, 10$$

es una función de probabilidad.

b)

1. $f(x) = k\left(\frac{1}{3}\right)^x > 0, \qquad \forall x = 1, 2, 3, \ldots, si, solo\ si\ k > 0$
2. $\sum_{x=1}^{\infty} k\left(\frac{1}{3}\right)^x = \frac{k}{3}\left[1 + \frac{1}{3} + \left(\frac{1}{3}\right)^2 + \ldots\right]$

$$= \frac{k}{3}\left[\frac{1}{1-\frac{1}{3}}\right] = k\left(\frac{1}{3}\right)\left(\frac{3}{2}\right) = 1$$

De donde, k = 2. Luego,

$$f(x) = 2\left(\frac{1}{3}\right)^x, \qquad x = 1, 2, 3, \ldots$$

es una función de probabilidad.

Ejemplo 5.4

Se tiene 6 cajas numeradas 1, 2, 3, 4, 5, y 6; se tiene también 6 cartas numeradas 1, 2, 3, 4, 5, y 6. Se coloca al azar, (aleatoriamente) una carta en cada caja. Sea X la variable aleatoria el lugar que ocupa la primera carta con el número par. Determinar la distribución de probabilidad de X**.** Representar las respuestas en una tabla de la siguiente forma:

X	$p(x)$

Solución:

La variable aleatoria X está definida por

$X(\omega)$ = número de la caja que indica el lugar que ocupa la primera carta con el número par.

Cartas $\{1, 2, 3, 4, 5, 6\}$ y $\{2, 4, 6\}$ son pares

$\mathbb{R}_x = \{1, 2, 3, 4\}$

Sea F_i : "En la caja i se colocó una carta impar, $i = 1, 2, 3$"

E_i : "En la caja i se colocó una carta par, $i = 1, 2, 3, 4$"

Entonces el dominio de X es:

$$\Omega = \{E_1, F_1E_2, F_1F_2E_3, F_1F_2F_3E_4\}$$

$$p(1) = P[X = 1] = P[E_1] = \frac{3}{6} = \frac{1}{2}$$

$$p(2) = P[X = 2] = P[F_1E_2] = P[F_1]P[E_2 / F_1] = \frac{3}{6} \times \frac{3}{5} = \frac{3}{10}$$

$$p(3) = P[X = 3] = P[F_1F_2F_3] = P[F_1]P[F_2 / F_1]P[E_3 / F_1F_2] = \frac{3}{6} \times \frac{2}{5} \times \frac{3}{4} = \frac{3}{20}$$

$$p(4) = P[X = 4] = P[F_1F_2F_3E_4] = P[F_1]P[F_2 / F_1]P[F_3 / F_1F_2]P[E_4 / F_1F_2F_3]$$

$$= \frac{3}{6} \times \frac{2}{5} \times \frac{1}{4} \times \frac{3}{3} = \frac{1}{20}$$

La representación tabular es:

X	1	2	3	4
p(x)	1/2	3/10	3/20	1/20

5.2.4 Características de una variable aleatoria discreta

1. El valor medio o media de una distribución, también conocido como esperanza matemática (μ), está definida por:

$$\mu = E(X) = \sum x \times f(x) \tag{5.10}$$

2. La Varianza de una distribución, se denota por Var(X) o por la letra griega σ_X^2 (o simplemente σ^2) y está definida por:

$$Var(X) = \sigma^2 = E[(X-\mu)^2] = \sum_{x \in \mathbb{R}_x} (x-\mu)^2 f(x) \qquad (5.11)$$

3. La Desviación Típica de la distribución se define como la raíz cuadrada de la varianza y se denota por σ , su valor se obtiene con la ecuación 5.12:

$$\sigma = +\sqrt{\sigma^2} \qquad (5.12)$$

4. La Moda es el valor de ocurrencia mas frecuente. Así la moda, es el valor de X que maximiza *f(x)* y que satisface la ecuación 5.13:

$$p(x_0) \geq p(x) \; ; \qquad \forall \; x \; \in \; \mathbb{R}_x \qquad (5.13)$$

5. La mediana de la variable aleatoria X (o *de la distribución)* se representa por:

$$F(x) = \sum f(x) = 0.5 \qquad (5.14)$$

6. El 100 *k*-ésimo **percentil** de la variable aleatoria X (denotado por x_k) es el número más pequeño posible, tal que, la probabilidad de no excederlo es cuando menos *k*, con $0 < k < 1$.

Es decir, x_k es el valor más pequeño, tal que:

$$F(x_k) = P(X \leq x_k) \geq k \tag{5.15}$$

$$P(X \geq x_k) \geq 1 - k \tag{5.16}$$

Los percentiles que cortan a la función de probabilidad o función de densidad en cuatro probabilidades iguales se llaman *cuartiles* de la distribución, y son: $x_{0.25}$, $x_{0.50}$ y $x_{0.75}$. Al cuartil x_{50} se le llama también *mediana* de la variable aleatoria. La diferencia $x_{0.75}$ - $x_{0.25}$ es el *rango intercuartil* que es la longitud de un intervalo que encierra el 50% de la distribución de probabilidad de X.

Ejemplo 5.5

Hallar la esperanza matemática de la variable aleatoria X con distribución de probabilidad dada por:

x	0	1	2	3	4
$p(x)$	0.2	0.4	0.3	0.08	0.02

Solución:

Por la ecuación 5.26 se encuentra la esperanza matemática

$$E(X) = 0(0.2) + 1(0.4) + 2(0.3) + 3(0.08) + 4(0.02) = 1.32$$

Ejemplo 5.6

Hallar la moda y la mediana de la variable aleatoria X, cuya distribución de probabilidad está definida por:

x	-2	-1	1	2
P(x)	1/3	1/6	1/6	1/3

Solución:

a) Desde que $p(x)$ es mayor, cuando x = -2, y x = 2, entonces x_o = 2 y -2 son dos modas de variable aleatoria X.

b) La mediana se calcula usando la definición:

$$F(x_o) = P[X < x_o] = \frac{1}{2}$$

se cumple para x_o = -1, también para cualquier $x \in [-1,1\rangle$. En particular se puede tomar x_{me} = 0.

5.2.5 Propiedades de los valores esperados para una variable aleatoria discreta

1. Si X es una variable aleatoria, a y b constantes. Entonces:

$(i) \quad E(a) = a \qquad (5.17)$

$(ii) \quad E[ah(X)] = aE[h(X)] \qquad (5.18)$

$(iii) \quad E[ah(X) + bg(X)] = aE[h(X)] + bE[g(X)] \qquad (5.19)$

2. Dada una función de probabilidad *f(x),* la Esperanza Matemática *E(X)* de una función cualquiera *g(x)* es:

$$E[g(X)] = \sum_{i=1}^{n} g(x_i) f(x_i) \qquad (5.20)$$

3. Dada cualquier colección de k variables aleatorias, la función $y = g(x_1, x_2, ..., x_k)$, es también una función aleatoria, y su valor esperado o media es:

$$E(y) = \sum_{x_1} \sum_{x_2} ... \sum_{x_k} g(x_1, x_2, ..., x_k) f(x_1, x_2, ..., x_k) \qquad (5.21)$$

4.

$$E[g(x_1, x_2)] = E[(X_1 - \mu_1)(X_2 - \mu_2)] \qquad (5.22)$$

Que es positivo cuando x_1 y x_2 toman simultáneamente valores muy grandes o muy pequeños y con gran probabilidad.

5. El valor esperado de una combinación linear de 2 variables aleatorias independientes $y = a_1 X_1 + a_2 X_2$ es:

$$E(y) = \mu_y = E(a_1 X_1 + a_2 X_2) = a_1 E(X_1) + a_2 E(X_2) \qquad (5.23)$$

La ecuación 5.23 se cumple tanto para variables independientes y no independientes.

5.1. Para el caso de la varianza:

$$Var(Y) = E[(Y-\mu_y)^2] = a_1^2\ Var(x_1) + a_2^2\ Var(x_2) \tag{5.24}$$

5.2.6 Momentos de orden superior y asimetría de una variable aleatoria discreta

El *momento de orden k alrededor del punto a de la variable aleatoria* **X** es la esperanza matemática de $(X-a)^k$, y se denota por $\mu_{k,a}$

$$\mu_{k,a} = E[(X-a)^k] = \sum_{x\in\mathbb{R}x} (x-a)^k\, p(x) \tag{5.25}$$

Si *a = 0*, se llaman *momentos iniciales o momentos alrededor del origen de orden k de una variable aleatoria a* $E(X^k)$. Es decir:

$$\mu_{k,0} = \mu_k^{'} = E(X^k) = \sum_{x\in\mathbb{R}x} x^k\, p(x) \tag{5.26}$$

Si *a = E(X)*, se llaman momentos alrededor de la media o momento central de orden *k* de una variable aleatoria a $E[(X-\mu)^k]$. Es decir:

$$\mu_{k,\mu} = \mu_k = E[(X-\mu)^k] = \sum_{x \in \mathbb{R}_x} (x-\mu)^k \, p(x) \tag{5.27}$$

Al segundo momento alrededor de la media de una variable aleatoria X se ha llamado varianza de la variable aleatoria, o sea $\sigma^2 = \mu_2, \mu$.

El primer momento central (o alrededor de la media) para cualquier variable aleatoria es igual a cero, es decir $\mu_1 = 0$.

Los momentos iniciales y centrales de primer, segundo, tercer y cuarto orden están vinculados por las relaciones:

$$\mu_1 = 0 \tag{5.28}$$

$$\mu_1 = \mu'_2 - \mu'^2_1 \tag{5.29}$$

$$\mu_3 = \mu'_3 - 3\mu'_1 \, \mu'_2 + 2\mu'^3_1 \tag{5.30}$$

$$\mu_4 = \mu'_4 - 4\mu'_1 \, \mu'_3 + 6\mu'^2_1 \, \mu'_2 - 3\mu'^4_1 \tag{5.31}$$

Si la distribución de probabilidad es simétrica respecto a la media, entonces todos los momentos centrales de orden impar son iguales a cero, o sea:

$$\mu_1 = \mu_3 = \mu_5 = \ldots = 0 \tag{5.32}$$

La Asimetría o Sesgo es la relación:

$$S_k = \frac{E[(X-\mu)^3]}{\sigma^3} = \frac{\mu_3}{\sigma^3} \tag{5.33}$$

Si $S_k > 0$, la distribución esta sesgada a la derecha (asimétrica positiva)

Si $S_k < 0$, la distribución esta sesgada a la izquierda (asimétrica negativa)

Si $S_k = 0$, la distribución es simétrica con respecto a la media

La Curtosis o coeficiente de curtosis de una variable aleatoria X mide el grado de agudeza de una distribución y esta dado por:

$$C_X = \frac{E[(X-\mu)^4]}{\sigma^4} = \frac{\mu_4}{\sigma^4} \tag{5.34}$$

Ejemplo 5.7

Sea X una variable aleatoria con distribución de probabilidad

X	2	4	6	8
$p(x)$	0.4	0.3	0.2	0.1

Hallar:

a) Los cuatro primeros momentos iniciales

b) Los cuatro primeros momentos centrales

c) La asimetría.

Solución:

a) El momento inicial de primer orden

$\mu'_1 = 2(0.4) + 4(0.3) + 6(0.2) + 8(0.1) = 4$

El momento inicial de segundo orden

$$\mu_2 = 4(0.4) + 16(0.3) + 36(0.2) + 64(0.1) = 20$$

El momento inicial de tercer orden

$\mu_3 = 8(0.4) + 64(0.3) + 216(0.2) + 512(0.1) = 116.8$

El momento inicial de cuarto orden

$\mu'_4 = 16(0.4) + 256(0.3) + 1296(0.2) + 4096(0.1) = 752$

b) El primer momento central

$\mu_1 = 0$

El segundo momento central

$\mu_2 = \mu_2 - \mu_1^2 = 20 - 4^2 = 20 - 16 = 4$

El tercer momento central

$\mu_3 = \mu'_3 - 3\mu'_1\, \mu'_2 + 2\mu'^3_1 = 116.8 - 3\times 4\times 20 + 2\times 4^3 = 4.8$

El cuarto momento central

$$\mu_4 = \mu'_4 - 4\mu'_1\, \mu'_3 + 6\mu'^2_1\, \mu'_2 - 3\mu'^4_1$$

$$= 752 - 4\times 4(116.8) + 6\times 4^2 \times 20 - 3\times 4^4 = 35.2$$

c) $\sigma^2 = \mu_2 = 4$, luego $\sigma^2 = 2$. La asimetría es:

$$S_k = \frac{\mu_3}{\sigma^3} = \frac{4.8}{2^3} = \frac{4.8}{8} = 0.60$$

5.3 VARIABLE ALEATORIA CONTINUA

Definición 5.5 *Si el rango* R_x *, de una variable aleatoria X es un intervalo sobre la recta de los números reales se llama* ***variable aleatoria continua.***

Esto significa que una variable aleatoria continua puede tomar valores enteros o decimales.

Un ejemplo de esta variable es el caudal diario registrado en una estación de aforo, representado en la figura 5.4.

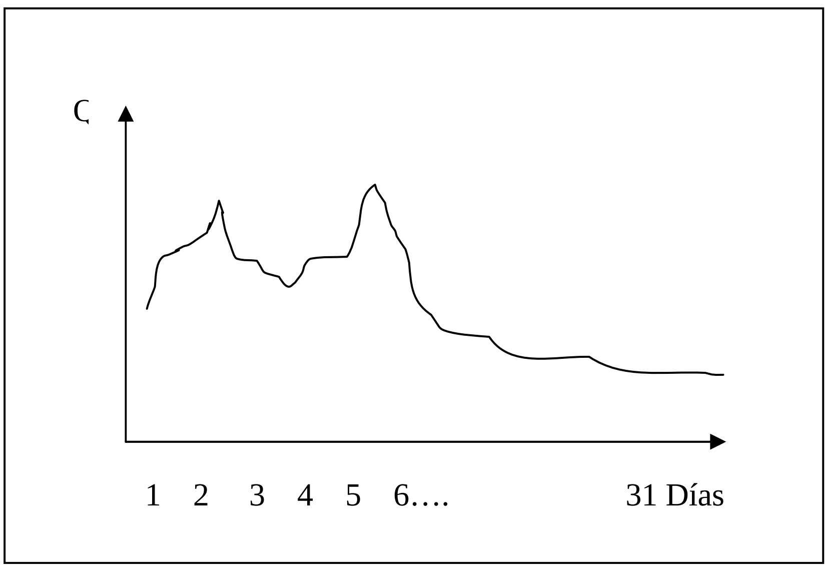


Figura 5.4 Ejemplo de Variable Aleatoria Continua

Ejemplo 5.8

Sea X la variable aleatoria que representa el número de kilogramos que pierde una persona al seguir una dieta específica durante cierto período.

Es una variable aleatoria continua, pues su rango (los valores que puede tomar) son todos los puntos de un intervalo, por ejemplo [1.5].

5.3.1 Función de Densidad

Sea X una variable aleatoria continua de rango R_X. La función de densidad de probabilidad asociada a la variable aleatoria, es una función *f(x)* integrable que satisface las siguientes condiciones:

a. $f(x) \geq 0 \qquad ; \forall x$ (5.35)

b. $\int_a^b f(x)dx = 1 \qquad ; \forall x$ (5.36)

En la ecuación 5.36 se tiene que *x* es cualquier número real, $a \rightarrow \infty$ *y* $b \rightarrow -\infty$, lo cual se representa en la figura 5.6. Entonces la probabilidad de que X se encuentre entre a y b es:

$$P(a < X < b) = \int_a^b f(x)dx \tag{5.37}$$

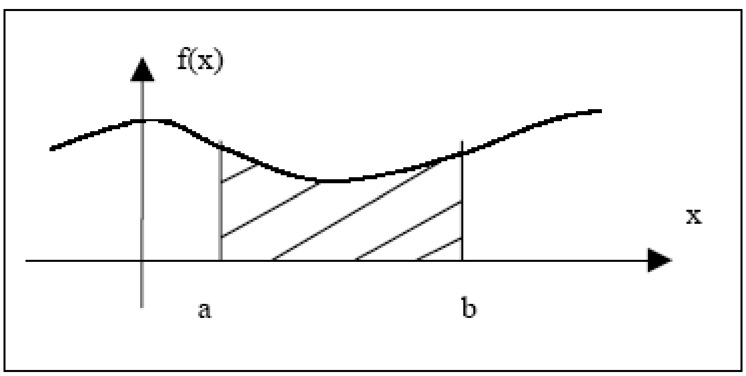


Figura 5.5 Curva de probabilidad de una función de densidad

De aquí que a *f(x)* se le conoce como función de distribución de probabilidad para una variable aleatoria continua. Pero con mayor frecuencia se le conoce como función de densidad.

5.3.2 Función de distribución de una variable aleatoria continua

La función de distribución o función de distribución acumulada de la variable aleatoria X, denotada por *F(x)* , se define como:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dx; \qquad \forall x_i \in R_x \tag{5.38}$$

5.3.3 Propiedades de la función de distribución para una variable aleatoria continua

1. $0 \leq F(x) \leq 1 \ , \ \forall \ x \ \in \ \mathbb{R}_x$ (5.39)

2.

a) $\lim\limits_{x \to -\infty} F(x) = \lim\limits_{x \to -\infty} \int_{-\infty}^{x} f(t)dt = 0$ (5.40)

b) $\lim\limits_{x \to \infty} F(x) = \lim\limits_{x \to \infty} \int_{-\infty}^{x} f(t)dt = 1$ (5.41)

3. La función de distribución es no decreciente, esto es si:

$$a \leq b \Rightarrow F(a) \leq F(b) \tag{5.42}$$

4. $\lim\limits_{h \to 0} F(x+h) = F(x) \ , \ \forall \ x \ \in \ \mathbb{R}_x$, con $h > 0$ (5.43)

Ósea F es continua por la derecha, en todos los puntos.

5. Del segundo teorema fundamental del cálculo se tiene que si *F(x)* es una función derivable, entonces:

$$f(x) = \frac{dF(x)}{dx} \tag{5.44}$$

Ejemplo 5.9

Sea:

$$f(x) = \begin{cases} kx^2 & , Si \ |x-2| \le 1 \\ 0 & , Si \ |x-2| > 1 \end{cases}$$

a) Hallar el valor de k tal que *f(x)* sea una función de densidad de una variable aleatoria continua X.

b) Calcular la probabilidad que la variable aleatoria X se encuentre en el intervalo $\langle 2, 3 \rangle$.

Solución:

Por la propiedad de valor absoluto:

| x-2 | ≤ 1, si solo si -1 ≤ -2 ≤ 1, de aquí 1 ≤ x ≤ 3

| x-2 | > 1, si solo si x-2 >1 ∨ x-2 < -1 o sea x > 3 ∨ x < 1

Por lo tanto la función *f(x)* se escribe:

$$f(x)=\begin{cases} 0 & ,x<1 \\ kx^2 & ,1\le x\le 3 \\ 0 & ,x>3 \end{cases}$$

a) Como todos los valores de la variable aleatoria se hallan en el intervalo $\langle 1,3\rangle$. Por la ecuación 5.37 se obtiene:

$$\int_1^3 kx^2dx = k\,\frac{x^3}{3}\Big|_1^3 = k\left(9-\frac{1}{3}\right)=k\,\frac{26}{3}=1$$

De la ecuación anterior: $k=\dfrac{3}{26}$

b) Por la ecuación 5.38:

$$P(2<X<3)=\int_2^3\frac{3x^2}{26}dx=\frac{x^3}{26}\Bigg|_2^3=\frac{27}{26}-\frac{8}{26}=\frac{19}{26}$$

Ejemplo 5.10

La variable aleatoria y tiene por función de densidad

$$f(y)=\begin{cases} 1/5\ , & 0\le y\le 5 \\ 0\ , & en\ otro\ lugar \end{cases}$$

Calcular la probabilidad que sean reales las raíces de la ecuación:

$$4x^2+4xy+(y+2)=0$$

Solución:

La solución de la ecuación dada es:

$$x = \frac{-4y \pm \sqrt{16y^2 - 4(4)(y+2)}}{8} = \frac{-y \pm \sqrt{y^2 - y - 2}}{2}$$

Las raíces serán reales, esto es $x \in \mathbb{R}$, sólo si, $y^2 - y - 2 \geq 0$

$$y^2 - y - 2 = (y-2)(y+1) \geq 0 \Rightarrow \begin{cases} y-2 \geq 0 \;\wedge\; y+1 \geq 0 \\ \vee \\ y-2 \leq 0 \;\wedge\; y+1 \leq 0 \end{cases}$$

$$\Rightarrow \begin{cases} y \geq 2 \;\wedge\; y \geq -1 \\ \vee \\ y \leq 2 \;\wedge\; y \leq -1 \end{cases}$$

Luego, la solución a la desigualdad es el intervalo $[2, \infty\rangle \cup \langle -\infty, -1]$ y como el rango de la variable aleatoria es [0,5]. Por lo tanto, los valores de y que hacen que los valores de la ecuación sean reales es equivalente a calcular:

$$P[2 \leq y \leq 5] = \int_2^5 \frac{1}{5} dy = \frac{1}{5}[5-2] = \frac{3}{5}$$

5.3.4 Características de la variable aleatoria continua

1. El valor medio o media de una distribución μ, también conocido como valor esperado $E(X)$ esta definido por:

$$\mu = E(X) = \int_{-\infty}^{\infty} x \times f(x)dx \tag{5.45}$$

2. La Varianza de la distribución, esta definida por:

$$Var(X) = \sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \tag{5.46}$$

3. La Desviación Típica de la distribución se define como la raíz cuadrada de la varianza y se denota por σ , su valor se obtiene con la ecuación 5.46:

$$\sigma = +\sqrt{\sigma^2} \tag{5.47}$$

4. La Moda está definida por:

$$f(x_0) \geq f(x) \;; \qquad \forall\ x \in \mathbb{R}_x \tag{5.48}$$

5. La Mediana se representa por:

$$F(x) = \int_{-\infty}^{x} f(x)dx = 0.5 \tag{5.49}$$

6. El 100 *k*-ésimo **percentil** de la variable aleatoria X es ídem que para la variable aleatoria discreta, en el subtítulo 5.2.4.

Ejemplo 5.11

Sea X es una variable aleatoria cuya función de densidad esta definida por:

$$f(x)=\begin{cases}\dfrac{3}{4}x(2-x), & 0\le x\le 2\\ 0, & en\ otro\ caso\end{cases}$$

Calcular la esperanza de X

Solución:

Por la ecuación 5.45 se encuentra la esperanza matemática

$$E(X)=\int_{-\infty}^{\infty}\frac{3}{4}x^2(2-x)dx=\int_0^2\frac{3}{4}x^2(2-x)dx$$

$$=\frac{3}{4}\left[\frac{2x^3}{3}-\frac{x^4}{4}\right]_0^2=1$$

Ejemplo 5.12

Se da la función de densidad de probabilidad de la variable aleatoria X

$$f(x)=\begin{cases}x-x^3/4, & 0\le x\le 2\\ 0, & en\ otros\ casos\end{cases}$$

Hallar:

a) La moda

b) La mediana de esta variable aleatoria

Solución:

a) Hallar el máximo de la función *f(x)*. Para esto se encuentra la primera y la segunda derivada

$$f'(x) = 1 - \frac{3}{4}x^2; \qquad f''(x) = -\frac{3}{2}x$$

De la ecuación $f'(x) = 1 - \frac{3}{4}x^2 = 0$, se obtiene $x = \pm\frac{2}{\sqrt{3}}$

Entre las dos raíces de esta ecuación se escoge solo la que esta comprendida entre [0,2]. Por lo tanto, $x = \frac{2}{\sqrt{3}}$

Puesto que $f''\left(\frac{2}{\sqrt{3}}\right) = -\frac{3}{2}\left(\frac{2}{\sqrt{3}}\right) = -\sqrt{3} < 0$, entonces en $x = \frac{2}{\sqrt{3}}$

la función *f(x)* presenta su máximo, o sea:

$$x_{md} = \frac{2}{\sqrt{3}}$$

b) Por la definición de la mediana se tiene

$$P[X < x_o] = \int_0^{x_o}\left(x - \frac{x^3}{4}\right)dx = \left(\frac{x^2}{2} - \frac{x^4}{16}\right)_0^{x_o} = \frac{x_o^2}{2} - \frac{x_o^4}{16} = \frac{1}{2}$$

Así, de la ecuación $\frac{x_o^2}{2} - \frac{x_o^4}{16} = \frac{1}{2}$ ó $x_o^4 - 8x_o^2 + 8 = 0$

Se obtiene: $x_o = \pm\sqrt{4 \pm \sqrt{8}}$

Entre las cuatro raíces de esta ecuación se escoge la que está comprendida en el intervalo [0,2]. Por lo tanto:

$$x_{me} = \sqrt{4 - \sqrt{8}} \simeq 1.09$$

Ejemplo 5.13

Si X es una variable aleatoria tal que:

$$F(x) = \begin{cases} 0 & , x < 0 \\ x/100 & , 0 \leq x \leq 100 \\ 1 & , x > 100 \end{cases}$$

Hallar:

a) $x_{0.10.}$

b) $x_{0.20.}$

c) La mediana.

d) El rango intercuartil.

Solución:

a) $F(x_{0.10}) = x_{0.10}/100 = 0.10$; es decir, $x_{0.10} = 10$

b) $F(x_{0.20}) = x_{0.20}/100 = 0.20$; es decir, $x_{0.20} = 20$

c) La mediana está definida por

$F(x_{0.50}) = x_{0.50}/100 = 0.50$; es decir, $x_{0.50} = 50$

d) En forma análoga se obtiene:

$x_{0.25} = 100\ (0.25) = 25$

$x_{0.75} = 100\ (0.75) = 75$

El rango intercuartil es $x_{0.75} - x_{0.25} = 50$

5.3.5 Propiedades de los valores esperados para variables aleatorias continuas

1. Dada una función de probabilidad *f(x),* la Esperanza Matemática *E(X)* de una función cualquiera *g(x)* es:

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) f(x) dx \qquad (5.50)$$

2. Cuando $g(x) = ax + b$ para a y b constantes, el valor esperado de $g(x)$ es:

$$E(aX + b) = aE(X) + b \qquad (5.51)$$

2.1. Para el mismo caso, su varianza:

$$Var(aX + b) = a^2 Var(X) \qquad (5.52)$$

3. Dada cualquier colección de *k* variables aleatorias, valor esperado o media de la función $y = g(x_1, x_2, ..., x_k)$ es:

$$E(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} .....\int_{-\infty}^{\infty} g(x_1, x_2, ..., x_k) f(x_1, x_2, ..., x_k) dx_1, ..., dx_k \qquad (5.53)$$

4. En general, tanto para variables continuas o discretas y con cualquier número de variables aleatorias:

Si X_i tiene una media μ_i y varianza σ_i^2 para:

$i = 1, 2, ..., k$, la combinación linear de $Y = a_1 X_1 + a_2 X_2 + ... + a_k X_k$ tiene:

$$E(a_1 X_1 + a_2 X_2 + ... + a_k X_k) = a_1 E(X_1) + a_2 E(X_2) + ... + a_k E(X_k) \qquad (5.54)$$

ó:

$$\mu_y = \sum_{i=1}^{k} a_i \mu_i \quad , \textit{para las variables aleatorias independientes} \qquad (5.55)$$

La ecuación 5.55 es aplicable a las variables aleatorias independientes.

La varianza esta dada por las ecuaciones 5.56 y 5.57:

$$Var(a_1 X_1 + a_2 X_2 + ... + a_k X_k) = a_1^2 \ Var(x_1) + a_2^2 \ Var(x_2) + ... + a_k^2 \ Var(X_k) \qquad (5.56)$$

ó:

$$\sigma_y^2 = \sum_{i=1}^{k} a_i^2 \sigma_i^2 \qquad (5.57)$$

5.3.6 Momentos de orden superior y asimetría de una variable aleatoria continua

El momento de orden k alrededor del punto a de la variable aleatoria X se denota por $\mu_{k,a}$.

$$\mu_{k,a} = E[(X-a)^k] = \int_{\mathbb{R}_x} (x-a)^k f(x)dx \qquad (5.58)$$

Los momentos iniciales o momentos alrededor del origen de orden k de una variable aleatoria a $E(X^k)$, son:

$$\mu_{k,0} = \mu_k^{'} = E(X^k) = \int_{\mathbb{R}_x} x^k f(x)dx \qquad (5.59)$$

El momento alrededor de la media o momento central de orden k de una variable aleatoria a $E[(X-\mu)^k]$ es:

$$\mu_{k,\mu} = \mu_k = E[(X-\mu)^k] = \int_{\mathbb{R}_x} (x-\mu)^k f(x)dx \tag{5.60}$$

Los momentos iniciales y centrales de primer, segundo, tercer y cuarto orden así como la asimetría y curtosis para una variable aleatoria continua son ídem que para la variable aleatoria discreta en el subtítulo 5.2.5.

BIBLIOGRAFÍA

Moya C. Rufino – Saravia A. Gregorio. Probabilidad e inferencia estadística; Editorial Limusa.

Chungara Castro Víctor. Estadística y Probabilidades; Editorial Leonardo.

Romero Mauricio. Texto de Probabilidad y Estadística; U.M.S.S.

Manchego C. Roberto. Estadística II; U.M.S.S.

CAPITULO VI

FUNCIONES DE DISTRIBUCIÓN DE VARIABLES ALEATORIAS

CAPÍTULO VI

FUNCIONES DE DISTRIBUCIÓN DE VARIABLES ALEATORIAS

En el capítulo anterior se especifica la diferencia entre las variables aleatorias discretas y continuas, en este capítulo se habla específicamente de las funciones de distribución que existen para ambas variables, llámese así a las distribuciones discretas de probabilidad, funciones de distribución de las variables aleatorias discretas y a las distribuciones continuas de probabilidad, funciones de distribución de las variables aleatorias continuas.

6.1 DISTRIBUCIONES DISCRETAS DE PROBABILIDAD

6.1.1 Distribución Binomial

Las características la distribución binomial son:

a) Ésta distribución arroja dos tipos de resultados, Ej.: Defectuoso, no defectuoso, cara o cruz, etc., denominados "éxito" (que es lo que se espera que ocurra) o "fracaso" (lo contrario del éxito).

b) Las probabilidades asociadas a cada uno de estos resultados son constantes, es decir, no cambian.

c) Cada uno de los ensayos o repeticiones del experimento son independientes entre sí.

d) El número de repeticiones *n* es constante.

Definición 6.1 *La distribución de probabilidad de la variable aleatoria binomial X, se llama* ***distribución binomial*** *y se denota por* $P[X = x \mid B; n, p]$ *ó* $b(x; n, p)$ *que se lee: La probabilidad de obtener exactamente x éxitos y* $n - x$ *fracasos.*

La ecuación de la distribución binomial es:

$$b(x;n,p) = \binom{n}{x} p^x q^{n-x}, \qquad x = 0,1,2,...,n \tag{6.1}$$

Si *n* es muy larga, el cálculo de probabilidades binomiales puede ser tedioso, para esto existe el uso de tablas.

La distribución binomial acumulada es:

$$B(x,n,p) = \sum_{k=0}^{x} b(k,n,p) \tag{6.2}$$

Los valores de *B(x,n,p)* a menudo tienen mas aplicación estadística y están tabulados.

Para el uso de tablas se usa la ecuación 6.3:

$$b(x,n,p) = B(x,n,p) - B(x-1,n,p) \tag{6.3}$$

6.1.1.1 Características de la Distribución Binomial

La media de la variable aleatoria Binomial esta definida por:

$$\mu = E(X) = np \tag{6.4}$$

La varianza está definida por:

$$\sigma_X^2 = Var(X) = npq \tag{6.5}$$

Y la desviación estándar de X es:

$$\sigma = \sqrt{npq} \tag{6.6}$$

Ejemplo 6.1

En la materia de Probabilidad y Estadística la probabilidad de que un estudiante apruebe es de 30 %. Hallar la probabilidad de que entre 10 estudiantes elegidos al azar aprueben:

a) 5 estudiantes.
b) Ninguno.
c) Al menos 8.

Solución:

a) Aplicando la distribución binomial

$n = 10$

$x = 5$ = variable que define el número de estudiantes aprobados.

$p = p(\acute{e}xito) = p$(estudiantes aprobados) = 0.3

$q = p(fracaso) = p$(estudiantes reprobados) = 1- p = 0.7

$$P(X = 5) = b(x = 5, n = 10, p = 0.3) = \binom{10}{5}(0.3)^5(0.7)^{10-5} = 0.1029$$

b) Ninguno significa la probabilidad de que ninguno de los 10 estudiantes aprueben

$$P(X=0)=b(x=0,n=10,p=0.3)=\binom{10}{0}(0.3)^0(0.7)^{10-0}=0.0282$$

c) Al menos 8 significa que de los 10 pueden aprobar 8, 9 o 10. Distribución acumulada.

$$\begin{aligned}P(X\geq 8)&=P(X=8)+P(X=9)+P(X=10)\\&=\binom{10}{8}(0.3)^8(0.7)^{10-8}+\binom{10}{9}(0.3)^9(0.7)^{10-9}+\binom{10}{10}(0.3)^{10}(0.7)^{10-10}\\&=0.0014+0.0001+0.000=0.0015\end{aligned}$$

6.1.2 Distribución Multinomial

Las características de esta distribución son:

a) La distribución multinomial arroja más de dos tipos de resultados.

b) Las probabilidades asociadas a cada uno de los resultados son constantes.

c) Cada uno de los ensayos o repeticiones del experimento son independientes, y permiten k resultados mutuamente excluyentes con probabilidad $p_1, p_2, p_3, \ldots ,p_k$ $\left(\sum_{k=i}^{k} p_i = 1)\right)$. Además el número de repeticiones n es constante.

d) El tipo de categoría dependen del número de resultados como ser:

1º resultado → 1º tipo de categoría

2º resultado → 2º tipo de categoría

. .

. .

. .

*k*º resultado → *k*º tipo de categoría

Definición 6.2 *La distribución de probabilidad conjunta de las variables aleatorias* $X_1, X_2, \ldots, X_k$*, es decir, la distribución de probabilidad del vector aleatorio*$(X_1, X_2, \ldots, X_k)$ *se llama* ***Distribución Multinomial****; y se lee: Probabilidad de que en n ensayos aparezcan* x_1 *objetos del primer tipo,* x_2 *objetos del segundo tipo ... y* x_k *objetos del último tipo.*

La distribución multinomial es obtenida por la ecuación 6.7:

$$p(x_1, x_2, \ldots, x_k) = P[X_1 = x_1, \ldots, X_k = x_k] = \frac{n!}{x_1!, x_2!, \ldots, x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}$$

(6.7)

Se debe tomar en cuenta que $\frac{n!}{x_1! x_2! \ldots x_k!}$ es una permutación por que es importante el orden, además $\sum_{i=1}^{k} x_i = n$.

Se observa también que $X_1, X_2, \ldots, X_k$ no son variables aleatorias independientes, puesto que $\sum_{i=1}^{k} x_i = n$. Es decir, conocidos los valores de $(k - 1)$ variables aleatorias cualesquiera, se conoce la que falta.

6.1.2.1 Características de la Distribución Multinomial

La media de esta distribución es:

$$E(X_i) = np_i; \qquad i = 1, 2, ..., k \tag{6.8}$$

Y la varianza es:

$$Var(X_i) = np_i(1 - p_i); \qquad i = 1, 2, ..., k \tag{6.9}$$

Ejemplo 6.2

En cierta población grande el 70% de las personas son derechos; el 20% izquierdos y 10% ambidiestros. Se escogen 10 personas aleatoriamente de la población. ¿Cual es la probabilidad que:

a) Todos sean derechos?

b) 7 sean derechos, 2 zurdos y 1 ambidiestro?

Solución:

a) $n = 10$; $x_1 = 10$; $p_1 = 0.7$

$$p(x_1 = 10; n = 10) = \frac{10!}{10!0!}(0.7)^{10} = 0.028$$

b) $n = 10$; $x_1 = 7$; $x_2 = 2$; $x_3 = 1$

$p_1 = 0.7$; $p_2 = 0.2$; $p_3 = 0.1$

$$p(x_1 = 7, x_2 = 2, x_3 = 1; n = 10) = \frac{10!}{7!2!1!}(0.7)^7(0.2)^2(0.1)^1 \simeq 0.119$$

6.1.3 Distribución Hipergeométrica

Las características de esta distribución son:

a) Al realizar un experimento con este tipo de distribución, se esperan dos tipos de resultados.

b) Las probabilidades asociadas a cada uno de los resultados no son constantes.

c) Cada ensayo o repetición del experimento no es independiente de los demás.

d) El número de repeticiones del experimento *n* es constante.

Definición 6.3 *La distribución de probabilidad de la variable aleatoria hipergeométrica X(ω) = el número de éxitos en la muestra de tamaño n sin reemplazo, tal que* $\mathrm{R_x} = [0,1,2,...,\min(n,M)]$ *; se llama distribución hipergeométrica y se denota por* $h(x;N,n,M)$.

La distribución hipergeométrica se obtiene por la ecuación 6.10:

$$\mathrm{p}(x) = h(x;N,n,M) = \sum_{k=0}^{x} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}; \qquad x = 0,1,2,...,\min(n,M) \qquad (6.10)$$

La función de distribución acumulada de la variable aleatoria hipergeométrica es:

$$F(x) = P[X \leq x] = \sum_{k=0}^{x} \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}; \qquad x = 0,1,2,...,\min(M,n) \qquad (6.11)$$

6.1.3.1 Características de la Distribución Hipergeométrica

La media de esta distribución es:

$$E(X) = n\left(\frac{M}{N}\right) \qquad (6.12)$$

Y la varianza es:

$$Var(X) = n\left(\frac{M}{N}\right)\left(1-\frac{M}{N}\right)\left(\frac{N-n}{N-1}\right) \qquad (6.13)$$

Ejemplo 6.3

De una urna que contiene 5 bolas negras y 2 blancas, se extraen 3 bolas sin reposición. Obtener:

a) La función de distribución acumulada de probabilidad.

b) La probabilidad de extraer 2 negras.

Solución:

a) Reemplazando en la distribución hipergeométrica la variable aleatoria X es la obtención de una bola negra.

$N = 7$; $n = 2$; $M = 5$

$$P(X = k) = \frac{\binom{5}{k}\binom{7-5}{2-k}}{\binom{7}{2}} = \frac{\binom{5}{k}\binom{2}{2-k}}{\binom{7}{2}}$$

Se debe notar que los valores posibles de k son 0, 1, 2 únicamente.

b) Calculando luego para $x = 2$

$$P(X = 2) = \frac{\binom{5}{2}\binom{7-5}{2-2}}{\binom{7}{2}} = \frac{10}{21} = 0.47619$$

6.1.4 Distribución de Poisson

La distribución de Poisson mide la probabilidad de un suceso aleatorio a lo largo de un intervalo espacial o temporal.

Esta distribución tiene las siguientes características:

a) La ocurrencia de los eventos en unidades de medida contiguas son independientes.

b) El número promedio de ocurrencias de eventos en una unidad de medida es conocido e igual a λ.

c) La probabilidad de ocurrencia del evento es constante para dos intervalos de unidad de medida cualesquiera.

Definición 6.4 *Se dice que la variable aleatoria X, cuyos valores posibles son x = 0, 1, 2… tiene distribución de Poisson con parámetro* λ *(* $\lambda > 0$ *), si su función de probabilidad es la ecuación 6.14.*

$$p(x,\lambda) = P[X = x \mid P:\lambda] = \frac{\lambda^x \varepsilon^{-\lambda}}{x!}; \qquad x = 0,1,2,... \qquad (6.14)$$

En la ecuación 6.14, se tiene:

$\lambda = np =$ Es el número promedio de ocurrencias de los eventos en las p unidades de medida.

$p(x,\lambda)$ = probabilidad de que ocurran x éxitos, cuando el número promedio de ocurrencia de ellos es λ

λ = media o promedio de éxitos por unidad de tiempo, área o producto

$\varepsilon = 2.718$

x = variable que denota el número de éxitos que se desea que ocurra

Según el análisis a efectuar es necesario adecuar el promedio o valor de λ al periodo con el que se trabaja (ver ej. 6.4), para lo cual la formula es:

$$p(x,\lambda) = \frac{(\lambda t)^x \varepsilon^{-\lambda t}}{x!}; \qquad x = 0,1,2,... \qquad (6.15)$$

La función de distribución acumulada está dada por:

$$F(x,\lambda) = P[X = x \mid P:\lambda] = \sum_{k=0}^{x} \frac{\lambda^k \varepsilon^{-\lambda}}{k!}; \qquad x = 0,1,2,... \qquad (6.16)$$

6.1.4.1 Características de la distribución de Poisson

La media de la distribución de Poisson está dado por:

$$E(X) = \mu = \lambda \qquad (6.17)$$

La varianza es:

$$\sigma^2 = \lambda \qquad (6.18)$$

Ejemplo 6.4

En una central telefónica que recibe dos llamadas cada 3 minutos; calcular la probabilidad de que en el periodo de 6 minutos se presenten:

a) 5 llamadas.

b) No más de dos llamadas.

Solución:

a)

x = variable que define el número de llamadas en el periodo de 6 minutos = 5

λ = (2 llam/ 3 min)6 min. = 4 llamadas

$$p(x = 5, \lambda = 4) = \frac{(4)^5 (2.718)^{-4}}{5!} = 0.1562$$

b) Al indicar no más de dos llamadas se precisa de la suma de las posibilidades para 0, 1 y 2 llamadas.

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$P(X \leq 2) = \frac{(4)^0 (2.718)^{-4}}{0!} + \frac{(4)^1 (2.718)^{-4}}{1!} + \frac{(4)^2 (2.718)^{-4}}{2!} = 0.2381$$

c)

$$P(X \geq 4) = 1 - P(X \leq 3) = 1 - \left[P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \right]$$

$$P(X \geq 4) = 1 - \left[\frac{(4)^0 (2.718)^{-4}}{0!} + \frac{(4)^1 (2.718)^{-4}}{1!} + \frac{(4)^2 (2.718)^{-4}}{2!} + \frac{(4)^2 (2.718)^{-4}}{2!} \right] = 0.2381$$

6.1.5 Aproximación de Poisson a la Binomial

En este caso se determinan probabilidades de experimentos Binomiales, pero que dadas sus características, es posible aproximarlas con la distribución de Poisson, estas características son, $n \rightarrow \infty$ (n es muy grande) y $p \rightarrow 0$ (p es muy pequeña), por lo que:

$$p(x, n, p) = \binom{n}{x} p^x q^{n-x} \cong \frac{\lambda^x \varepsilon^{-\lambda}}{x!} \tag{6.19}$$

En el caso que no se cumpla $n \rightarrow \infty$ y $p \rightarrow 0$, la aproximación no se puede llevar a efecto, por lo que la ecuación a utilizar en este caso es:

$$p(x, \lambda) = \frac{\lambda^x \varepsilon^{-\lambda}}{x!} \tag{6.20}$$

En la ecuación 6.20:

$\lambda = np$ = número esperado de éxitos = tasa promedio de éxitos
n = número de repeticiones del experimento
p = probabilidad de éxito = p(éxito)

En la práctica es aceptable usar $n \geq 20$ y $p \leq 0.05$. Sí $n \geq 100$, la aproximación es generalmente excelente siempre y cuando $np \leq 10$. Para algunos autores es aconsejable usar $p < 0.1$ y $np \leq 5$.

Ejemplo 6.5

Se sabe que el 5% de los libros encuadernados en cierto taller tienen encuadernaciones defectuosas. Determinar la probabilidad de que 2 de 100 libros encuadernados en ese taller, tengan encuadernaciones defectuosas, usando:
a) La fórmula de la distribución Binomial.
b) La aproximación de Poisson a la distribución Binomial.

Solución:

a) $n = 100$
$p = 0.05 = p$(encuadernación defectuosa) $= p(\textit{éxito})$
$q = 0.95 = p$(encuadernación no defectuosa) $= p(\textit{fracaso})$
$x = 0,1, \quad 2, 3,...,100$ encuadernaciones defectuosas

$$p(x = 2, n = 100, p = 0.05) = \binom{100}{2}(0.05)^2(0.95)^{100-2}$$

$$= (4950)(0.05)^2(0.95)^{98} = 0.0812$$

b) $n = 100$ encuadernaciones

p = 0.05

$\lambda = np$ = (100)(0.05) = 5

x = 0, 1, 2, 3,...,100 encuadernaciones defectuosas

$$p(x=2,\lambda=5)=\frac{\lambda^x\varepsilon^{-\lambda}}{x!}=\frac{(5)^2(2.718)^{-5}}{2!}=0.0843$$

Al comparar los resultados de las probabilidades con una y otra distribución, se ve que la diferencia entre un cálculo y otro es de tan solo 0.0031, por lo que la aproximación de Poisson es una buena opción para calcular probabilidades Binomiales.

6.1.6 Distribución Geométrica

Esta distribución es un caso especial de la Binomial, en la cual se desea que ocurra un éxito por primera y única vez en el último ensayo que se realiza del experimento.

La distribución de probabilidad geométrica esta definida por la ecuación 6.21.

$$g(x)=P[X=x]=pq^{x-1}; \qquad x=1,2,3,... \tag{6.21}$$

En la ecuación 6.21:

g(x) = probabilidad de que ocurra un éxito en el evento *x* por primera y única vez

p = probabilidad de éxito

q = probabilidad de fracaso

Para hallar la función de distribución acumulada se emplea la ecuación 6.22.

$$F(x) = P[X \leq x] = \sum_{k=1}^{x} pq^{k-1} = p\sum_{k=1}^{x} q^{k-1} = p\left[\frac{1-q^x}{1-q}\right] \qquad x = 1,2,3,... \tag{6.22}$$

6.1.6.1 Aplicación de la distribución geométrica

De la distribución geométrica se llega a un concepto muy importante conocido como el periodo de retorno.

$$P(X < x) = 1 - P(X \geq x) \tag{6.23}$$

De donde:

$$P(X < x) = 1 - \frac{1}{T} \tag{6.24}$$

ó

$$T = \frac{1}{1-P} \tag{6.25}$$

Donde:

T	Periodo de retorno
$P(X \geq x)$	Probabilidad de excedencia
$P(X < x)$	Probabilidad de no excedencia

6.1.6.2 Características de la distribución Geométrica

La media de la distribución geométrica esta dada por:

$$E(X) = \mu = \frac{1}{p} \tag{6.26}$$

La varianza y la desviación estándar son:

$$\sigma^2 = \frac{q}{p^2} \tag{6.27}$$

$$\sigma = \frac{\sqrt{q}}{p} \tag{6.28}$$

Ejemplo 6.6

Sí la probabilidad de que un cierto dispositivo de medición muestre una desviación excesiva es de 0.05, ¿cuál es la probabilidad de que:

a) El sexto de estos dispositivos de medición sometidos a prueba sea el primero en mostrar una desviación excesiva?

b) El séptimo de estos dispositivos de medición sometidos a prueba, sea el primero que no muestre una desviación excesiva?.

Solución:

a) $x = 6 =$ sexto dispositivo de medición probado que es el primero que muestra una variación excesiva.

$p = 0.05 =$ probabilidad de que un dispositivo de medición muestre una variación excesiva.

$q = 0.95 =$ probabilidad de que un dispositivo de medición no muestre una variación excesiva.

$$g(x = 6) = (0.95)^{6-1}(0.05) = 0.03869$$

b) x = 5 = el quinto dispositivo de medición probado, es el primero que no muestra una desviación excesiva.

p = 0.95 = probabilidad de que un dispositivo de medición no muestre una variación excesiva.

q = 0.05 = probabilidad de que un dispositivo de medición muestre una variación excesiva.

$$g(x = 5) = (0.05)^{5-1}(0.95) = 0.0000059$$

6.2 DISTRIBUCIONES CONTINUAS DE PROBABILIDAD

6.2.1 Distribución Uniforme

Usa parámetros α y β y tiene una densidad de probabilidad:

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha}, & \alpha < x < \beta \\ 0, & para\ el\ resto \end{cases} \qquad (6.29)$$

Gráficamente:

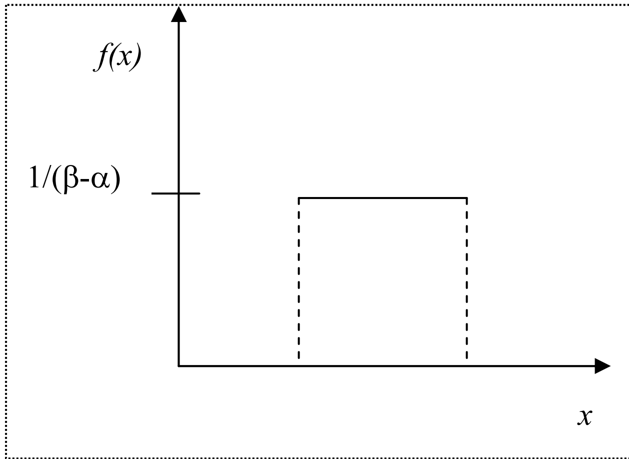


Figura 6.1 Grafica de distribución uniforme

Se debe notar que todos los valores entre α y β son "igualmente probables" en el sentido en que la probabilidad de que "*x*" caiga en un intervalo de ancho *Δx* (enteramente contenido en *α y β*) = *Δx / (β - α),* sin considerar la posición exacta del intervalo.

6.2.1.1 Características de la Distribución Uniforme

La media de la distribución uniforme se define por:

$$\mu = \int_{\alpha}^{\beta} x.\frac{1}{\beta - \alpha} dx \qquad (6.30)$$

$$\mu = \frac{\alpha + \beta}{2} \qquad (6.31)$$

Y la varianza, calculando primero:

$$\mu'_2 = \int_{\alpha}^{\beta} x^2 \frac{1}{\beta - \alpha} dx$$

Resolviendo se tiene : $\mu'_2 = \dfrac{\alpha^2 + \alpha\beta + \beta^2}{3}$

Luego : $\sigma^2 = \mu'_2 - \mu^2$

$$\sigma^2 = \frac{1}{12}(\beta + \alpha)^2 \qquad (6.32)$$

Ejemplo 6.7

La variable aleatoria X esta uniformemente distribuida en el intervalo [0,4]. ¿Cuál es la probabilidad que las raíces de la ecuación $y^2 + 4Xy + X + 1 = 0$ son reales?

Solución:

Desde que la variable aleatoria X, esta uniformemente distribuido en el intervalo [0,4], su función de densidad es:

$$f(x) = \begin{cases} 1/4, & 0 \le x \le 4 \\ 0, & \textit{en otros casos} \end{cases}$$

Resolviendo la ecuación cuadrática se tiene:

$$y = \frac{-4X \pm \sqrt{16X^2 - 4(X+1)}}{2}$$

$y \in \mathbb{R}, si\ solo\ si \qquad 16X^2 - 4X - 4 \geq 0 \qquad ó \qquad 4X^2 - X - 1 \geq 0$

Completando cuadrado se obtiene $\left(X - \frac{1}{8}\right)^2 \geq \frac{17}{64}$, de donde:

$$X \geq \frac{\sqrt{17}+1}{8} \qquad ó \qquad X \leq \frac{1-\sqrt{17}}{8}$$

La variable aleatoria **X** no puede tomar valores negativos, y

$$\frac{1-\sqrt{17}}{8} < 0$$

Por lo tanto, la ecuación tendrá raíces reales solamente en el caso

$$X \geq \frac{\sqrt{17}+1}{8}$$

y la probabilidad que esto ocurra es ,

$$P[\frac{\sqrt{17}+1}{8} \leq X] = \int_{\frac{\sqrt{17}+1}{8}}^{4} \frac{1}{4} dx = 1 - \frac{\sqrt{17}+1}{32} = 0.8399$$

6.2.2 Distribución Normal

Características:

a) Es generada por una variable de tipo continuo, denominada x; $-\infty < x < \infty$

b) La función que nos define esta distribución es:

$$f(x,\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}\varepsilon^{-(x-\mu)^2/2\sigma^2} \quad ; \ -\infty < x < \infty \qquad (6.33)$$

Al dar a la función los valores de μ, σ^2 y valores a x, obtendremos la distribución en cuestión, la que tiene forma de campana, por lo que también se le conoce como campana de Gauss. Hay un número infinito de funciones de densidad normal, una para cada combinación de μ y σ. La media μ mide la ubicación de la distribución y la desviación estándar σ mide su dispersión.

c) Es simétrica con respecto a su eje vertical.

d) Es asintótica con respecto a su eje horizontal; esto quiere decir que jamás va a tocar el eje de las equis.

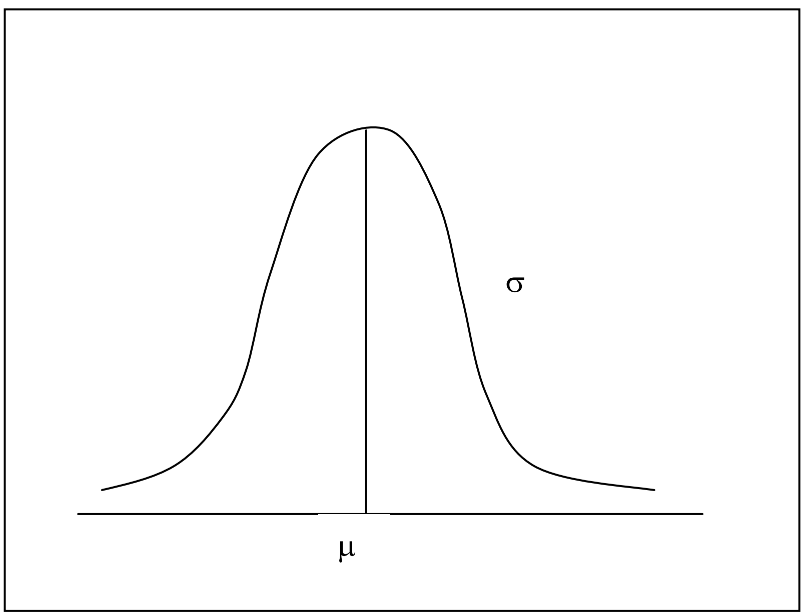


Figura 6.2 Grafica de distribución normal

e) El área total bajo la curva es 1.

f) Sí se suma a $\mu \pm \sigma$, se observa que aproximadamente el 68.26% de los datos se encuentran bajo la curva, si se suma a $\mu \pm 2\sigma$, el 95.44% de los datos estará entre esos límites y si se suma a $\mu \pm 3\sigma$, entonces el 99.74% de los datos caerá dentro de esos límites. Esta característica es a la vez una forma empírica y rápida de demostrar si los datos que se analizan tienen una distribución normal; ya que para trabajar los datos con esta distribución, debe verificarse que efectivamente así se distribuyen, ya que de no hacerlo, las decisiones que en un momento dado se tomarán de un análisis de los datos con la distribución normal, serían erróneas.

La función $f(x, \mu, \sigma^2)$, se integra entre los límites de la variable x; esto es,

$$p(a \geq x < b) = \int_a^b f(x, \mu, \sigma^2)dx \tag{6.34}$$

La integral anterior da el área bajo la curva de la función, desde a hasta b, que corresponde o es igual a la probabilidad buscada.

Debido a la dificultad que se presenta para integrar esta función cada vez que sea necesario, lo que se hace es tipificar el valor de la variable x, esto es, x se transforma en un valor de z, de la siguiente manera:

$$z = \frac{x - \mu}{\sigma} = valor \tag{6.35}$$

Este valor de z es buscado en una tabla donde vienen áreas asociadas a este valor, y haciendo uso de los valores tabulados, se determina la probabilidad requerida. La tabla que es usada para calcular las probabilidades es la que nos da el área que se muestra a continuación:

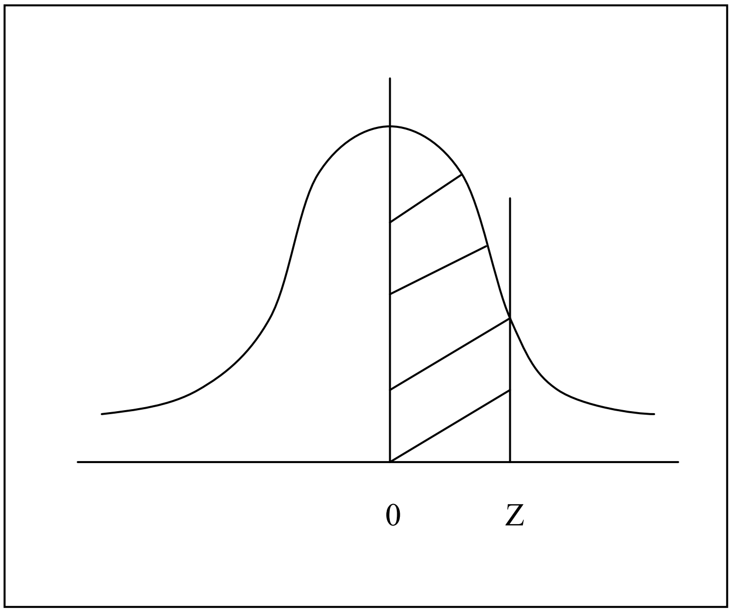


Figura 6.3 Area asociada a Z

6.2.2.1 Características de la Distribución Normal

La media y al varianza de la Distribución Normal están definidas por las ecuaciones 6.36 y 6.37 respectivamente:

$$E(X) = \mu \tag{6.36}$$

$$Var(X) = \sigma^2 \tag{6.37}$$

Ejemplo 6.8

El acero que se utiliza para tuberías de agua a menudo se recubre internamente con un mortero de cemento para evitar la corrosión. En un estudio de los recubrimientos de mortero de una tubería empleada en un proyecto de transmisión de agua en California

(Transportation Engineering Journal, Noviembre de 1979) se especificó un espesor de 7/16 pulgadas para el mortero. Un gran número de mediciones de espesor dieron una media de 0.635 pulgadas y una desviación estándar de 0.082 pulgadas. Si las mediciones de espesor, tenían una distribución Normal, ¿qué porcentaje aproximado fue inferior a 7/16 de pulgada?

Solución:

x = variable que define el espesor del mortero en pulgadas

μ = 0.635 pulgadas

σ = 0.082 pulgadas

$$Z = \frac{7/16 - 0.635}{0.082} = \frac{0.4375 - 0.635}{0.082} = -2.4085 \approx -2.41$$

$$p(z = -2.41) = 0.492$$

$p(x < 7/16\ pulgadas) = 0.5 - p(z = -2.41) = 0.5 - 0.492 = 0.008$

Por tanto, 0.008 x 100% = 0.8% de los recubrimientos de mortero tienen un espesor menor de 7/16 pulgadas

6.2.3 Distribución Exponencial

Sea X una variable aleatoria continua. Se dice que X tiene una distribución exponencial con parámetro real λ, si su función de densidad esta dado por la ecuación 6.38.

$$f(x)=\begin{cases}\lambda\varepsilon^{-\lambda x}, & x\geq 0\\ 0, & en\ otros\ casos\end{cases} \tag{6.38}$$

En la ecuación 6.38 el parámetro real λ es una constante positiva. Además $\varepsilon = 2.718$.

La función de distribución de la variable aleatoria X con Distribución exponencial es:

$$F(x)=\begin{cases}0, & x<0\\ 1-\varepsilon^{-\lambda x}, & x\geq 0\end{cases} \tag{6.39}$$

6.2.3.1 Características de la Distribución Exponencial

La media de la distribución exponencial es:

$$E(X)=\mu=\frac{1}{\lambda} \tag{6.40}$$

Y la varianza:

$$\sigma^2=\frac{1}{\lambda^2} \tag{6.41}$$

Ejemplo 6.9

Suponer que un sistema contiene cierto tipo de componente cuyo tiempo de falla en años está dado por la variable aleatoria T, distribuida exponencialmente con tiempo promedio de falla $\beta = 5$. Sí 5 de estos componentes se instalan en diferentes sistemas, ¿cuál es la probabilidad de que al menos 2 continúen funcionando después de 8 años?

Solución:

La probabilidad de que un determinado componente esté funcionando aún después de 8 años es:

$$P(>8) = \frac{1}{5}\int_{8}^{\infty} \varepsilon^{-\frac{t}{5}} dt = \varepsilon^{-\frac{8}{5}} \Big| \cong 0.2$$

la | indica que la integral se va a evaluar desde 8 hasta ∞.

Sea x el número de componentes funcionando después de 8 años. Entonces mediante la distribución Binomial,

$n = 5$

$p = 0.20$ = probabilidad de que un componente esté funcionando después de 8 años

$q = 1 - p = 0.80$ = probabilidad de que un componente no funcione después de 8 años

$P(x \geq 2) = p(x = 2) + p(x = 3) + p(x = 4) + p(x = 5) = 1 - p(x = 0, 1)$

$$= 1 - \left[\binom{5}{0}(0.2)^0(0.8)^5 + \binom{5}{1}(0.2)^1(0.8)^4\right] = 1 - 0.7373 = 0.2627$$

6.2.4 Distribución Log-Normal

Ocurre cuando se encuentra que los logaritmos de una variable aleatoria tienen una distribución normal. La densidad de probabilidad es:

$$f(x)=\begin{cases}\dfrac{1}{\sqrt{2\pi}\beta}x^{-1}e^{-(\ln x-\alpha)^2/2\beta^2}, & \alpha>0 \ \ y \ \ \beta>0 \\ 0, & para\ el\ resto\end{cases} \tag{6.42}$$

Donde:

lnx = logaritmo natural de x.

Gráficamente para α = 0 y β = 1:

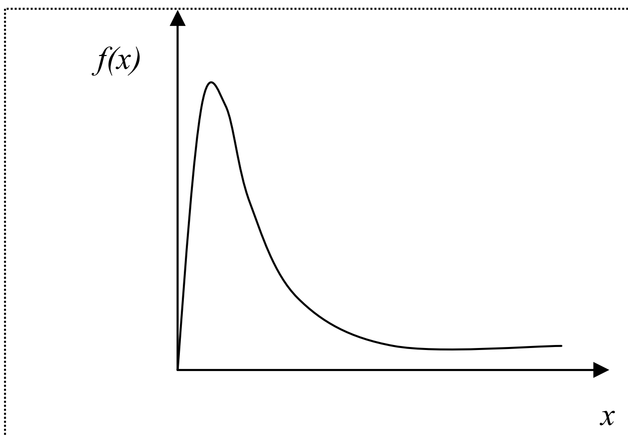


Figura 6.4 Grafica de distribución Log-Normal (Sesgada positivamente)

Para encontrar la probabilidad de que una variable aleatoria tome un valor entre "*a*" y "*b*" (*0 < a < b*) se debe evaluar:

$$\int_a^b \frac{1}{\sqrt{2\pi}\beta}x^{-1}e^{-(\ln x-\alpha)^2/2\beta^2}\,dx \tag{6.43}$$

Si se sustituye $y = ln\ x$, el integrando se convierte en la distribución normal con $\alpha=\mu$ y $\sigma=\beta$:

$$\int_{\ln a}^{\ln b} \frac{1}{\sqrt{2\pi}\beta} e^{-(y-\alpha)^2/2\beta^2} dy \qquad (6.44)$$

La ecuación 6.44 es igual a:

$$F\left(\frac{\ln b-\alpha}{\beta}\right)-F\left(\frac{\ln a-\alpha}{\beta}\right) \qquad (6.45)$$

Donde F es la función de distribución acumulada de la distribución normal estándar.

6.2.4.1 Características de la Distribución Log-Normal

La media de la Distribución Log-Normal es:

$$\mu = e^{\alpha+\frac{\beta^2}{2}} \qquad (6.46)$$

Y la varianza es:

$$\sigma^2 = e^{2\alpha+\beta^2}\left(e^{\beta^2}-1\right) \qquad (6.47)$$

6.2.5 Distribución Log-normal de 3 parámetros

La función de densidad de la distribución log-normal de 3 parámetros es:

$$f(x) = \frac{1}{(x - x_0)\sqrt{2\pi}\beta} \varepsilon^{-[\ln(x - x_0) - \alpha]^2 / 2\beta^2}; \qquad para\ x_0 \leq x < \infty \tag{6.48}$$

La función de distribución acumulada de log-normal de 3 parámetros es:

$$F(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z} \varepsilon^{(-Z^2/2)} dZ \tag{6.49}$$

Donde:

$$Z = \frac{\ln(x - x_0) - \alpha}{\beta} \tag{6.50}$$

6.2.6 Distribución Gamma

La densidad de distribución probabilística es:

$$f(x) = \begin{cases} \dfrac{1}{\beta^2 \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & para\ x > 0, \alpha > 0, \beta > 0 \\ 0, & para\ el\ resto \end{cases} \tag{6.51}$$

En la ecuación 7.45 $\Gamma(\alpha)$ es una valor de la función Gamma definida como:

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \tag{6.52}$$

La integración por partes muestra que:

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1); \ para\ \alpha > 1 \tag{6.53}$$

Por consiguiente:

$$\Gamma(\alpha) = (\alpha - 1)! \quad ; \text{ cuando } \alpha \text{ es entero positivo} \tag{6.54}$$

Gráficamente:

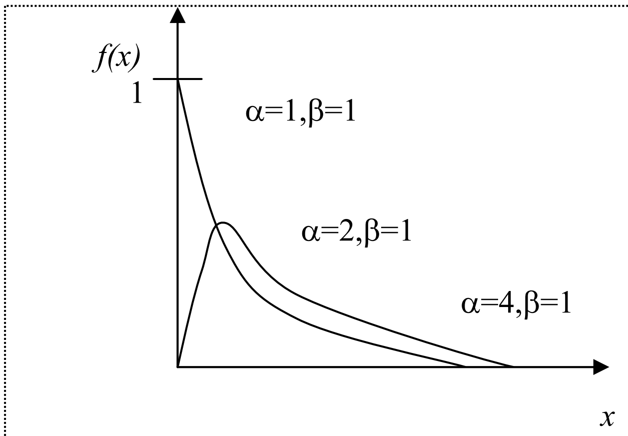


Figura 6.5 Grafica de distribución Gamma

6.2.6.1 Características de la Distribución Gamma

Para hallar la media:

$$\mu = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}\int_0^{\infty} x x^{\alpha-1} e^{\frac{-x}{\beta}} dx$$

Haciendo: $y = \dfrac{x}{\beta}$; $\mu = \dfrac{\beta}{\Gamma(\alpha)}\displaystyle\int_0^{\infty} y^{\alpha} e^{-y} dy$

$$\mu = \frac{\beta\Gamma(\alpha+1)}{\Gamma(\alpha)}$$

haciendo uso de: $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$

$$\mu = \alpha\beta \tag{6.55}$$

La Varianza se halla haciendo uso de las mismas propiedades:

$$\sigma^2 = \alpha\beta^2 \tag{6.56}$$

Para el caso especial de $\alpha = 1 \rightarrow$ distribución exponencial cuya propiedad de densidad es:

$$f(x) = \begin{cases} \dfrac{1}{\beta} e^{-x/\beta}, & para\ \alpha > 0\ \ y\ \beta > 0 \\ 0, & para\ el\ resto \end{cases} \tag{6.57}$$

De la ecuación 6.57 la media y la varianza son las ecuaciones 6.58 y 6.59 respectivamente:

$$\mu = \beta \tag{6.58}$$
$$\sigma^2 = \beta^2 \tag{6.59}$$

6.2.7 Distribución Gamma de tres parámetros o Pearson tipo III

Se dice que una variable aleatoria X tiene una distribución Gamma de tres parámetros o distribución Pearson tipo III, si su función de densidad es:

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}(x - x_0)^{\alpha-1} e^{-(x-x_0)/\beta} \tag{6.60}$$

Para:

$$x_0 \le x < \infty$$
$$-\infty < x_0 < \infty$$
$$0 < \beta < \infty$$
$$0 < \alpha < \infty$$

La función acumulada es:

$$F(x) = \int_{x_0}^{x} \frac{(x - x_0)^{\alpha-1} \varepsilon^{-\frac{(x-x_0)}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)} dx \qquad (6.61)$$

Donde la variable reducida y gamma de 3 parámetros o Pearson tipo III, es:

$$y = \frac{x - x_0}{\beta} \qquad (6.62)$$

6.2.8 Distribución Log-Pearson tipo III

Se dice que una variable aleatoria X, tiene una distribución log-Pearson tipo III, si su función de densidad de probabilidad es:

$$f(x) = \frac{(\ln x - x_0)^{\alpha-1} \varepsilon^{-\frac{\ln x - x_0}{\beta}}}{x\beta^{\alpha}\Gamma(\alpha)} \qquad (6.63)$$

Para:

$x_0 \le x < \infty$
$-\infty < x_0 < \infty$
$0 < \beta < \infty$
$0 < \alpha < \infty$

Su función de distribución acumulada es:

$$F(x) = \int_{x_0}^{x} \frac{(\ln x - x_0)^{\alpha - 1} \varepsilon^{-\frac{\ln x - x_0}{\beta}}}{x\beta^{\alpha}\Gamma(\alpha)} dx \qquad (6.64)$$

La variable reducida y log-pearson tipo III, es:

$$y = \frac{\ln x - x_0}{\beta} \qquad (6.65)$$

6.2.9 Distribución Gumbel

La distribución Gumbel, es una de las distribuciones de valor extremo, es llamada también Valor Extremo Tipo I, Fisher-Tippett tipo I ó distribución doble exponencial.

Su función de densidad es:

$$f(x) = \frac{1}{\alpha} \varepsilon^{-\frac{x-\beta}{\alpha} - \varepsilon^{-\frac{x-\beta}{\alpha}}} \qquad (6.66)$$

Su función de distribución acumulada es:

$$F(x) = \varepsilon^{-\varepsilon^{-\frac{x-\beta}{\alpha}}}; \qquad para \ -\infty < x < \infty \tag{6.67}$$

Donde la variable aleatoria reducida Gumbel, se define como:

$$y = \frac{x-\beta}{\alpha} \tag{6.68}$$

6.2.10 Distribución Weibull

Esta distribución está muy cerca de la distribución exponencial. La densidad de probabilidad de la distribución Weibull es:

$$f(x) = \begin{cases} \alpha\beta x^{\beta-1}e^{-\alpha x^{\beta}}, & para\ x > 0, \alpha > 0, \beta > 0 \\ 0, & para\ el\ resto \end{cases} \tag{6.69}$$

6.2.10.1. Características de la Distribución Weibull

La media de la Distribución Weibull es:

$$\mu = \alpha^{\frac{-1}{\beta}}\Gamma\left(1+\frac{1}{\beta}\right) \tag{6.70}$$

La varianza es:

$$\sigma^2 = \alpha^{\frac{-2}{\beta}}\left\{\Gamma\left(1+\frac{2}{\beta}\right) - \left[\Gamma\left(1+\frac{1}{\beta}\right)\right]^2\right\} \tag{6.71}$$

Gráficamente:

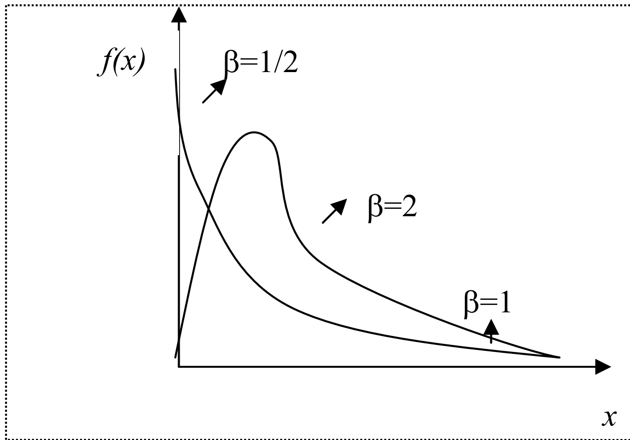


Figura 6.6 Grafica de distribución Weibull

Donde:

$\alpha = 1$, $\beta = 1, 2, ½$

6.2.11 Distribución Chi Cuadrado

Es un caso particular de la distribución Gamma y se obtiene haciendo $\alpha = 1/2;\ \ r = n/2$.

La función de densidad de la distribución chi-cuadrado es:

$$f(x) = \begin{cases} \dfrac{2^{-n/2} x^{n/2-1}}{\Gamma(n/2)} \varepsilon^{-x/2} & ; x > 0 \\ 0 & ; x < 0 \end{cases} \tag{6.72}$$

Gráficamente:

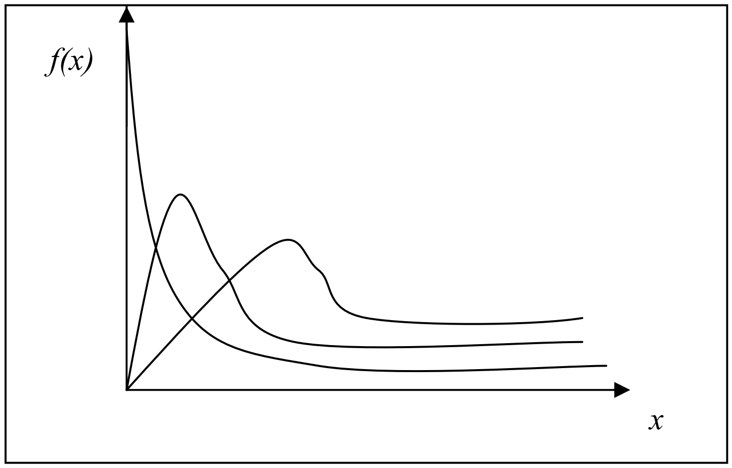


Figura 6.7 Grafica de distribución Chi-cuadrado

La media y la varianza son:

$$E(X) = \mu = n \tag{6.73}$$

$$\sigma^2 = 2n \tag{6.74}$$

6.2.12 Distribución t de Student

La función de densidad de la distribución t de Student es:

$$f(t) = T_0 \left(1 + \frac{t^2}{r}\right)^{-\frac{r+1}{2}} \qquad para \;\; -\infty < t < \infty \;\; ; \;\; r \in \mathbb{R}^+ \tag{6.75}$$

Es una curva simétrica respecto de 0. Se la compara con la curva normal, sólo que es ligeramente mas achatada.

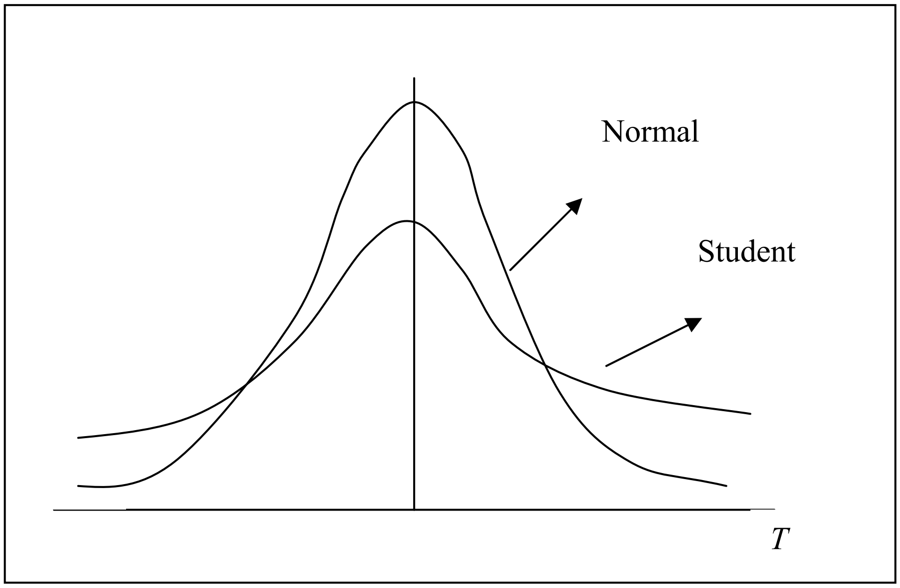


Figura 6.8 Gráfica de Distribución t de Student

El valor de la constante se define en términos de la función Gamma

$$T_0 = \frac{\Gamma[(r+1)/2]}{\Gamma(r/2)\sqrt{\pi r}} \tag{6.76}$$

La media y la varianza son:

$$E(X) = \mu = 0 \tag{6.77}$$

$$\sigma^2 = \frac{r}{r-2} \tag{6.78}$$

Grados de libertad: Se entiende por grados de libertad a las opciones que se presentan en la estimación de un cálculo.

Por ejemplo si se conoce que $a + b = 12$ y se deben calcular los valores de *a, b.* Una solución es elegir a = 7; entonces necesariamente debe cumplirse b = 5. Esto muestra que se tiene una opción para elegir un valor, por lo tanto el ejemplo tiene un grado de libertad.

Cuando se trata de calcular la probabilidad según la distribución de Student, usualmente se emplean tablas, donde para determinado valor de r (grados de libertad), se encuentra tabulada la probabilidad.

6.3 METODOLOGÍA DE AJUSTE A UNA DISTRIBUCIÓN Y EJEMPLOS APLICATIVOS

El ajuste a una distribución consiste en comprobar gráfica y estadísticamente si la frecuencia empírica de la serie analizada, se ajusta a una determinada función de probabilidades seleccionada a priori, con los parámetros estimados con base en los valores muestrales.

Las pruebas estadísticas tienen por objeto medir la certidumbre que se obtiene al hacer una hipótesis estadística sobre una población, es decir, calificar el hecho de suponer que una variable aleatoria, se distribuya según una cierta función de probabilidades.

6.3.1 Prueba de Smirnov – Kolmogorov

Esta prueba consiste en comparar las diferencias existentes, entre la probabilidad empírica de los datos de la muestra y la probabilidad teórica, tomando el valor máximo de valor absoluto, de la diferencia entre el valor observado *i* el valor de la recta teórica del modelo, es decir:

$$\Delta = \max\left|F(x) - P(x)\right| \tag{6.79}$$

Donde:

Δ = estadístico de Smirnov-Kolmogorov

$F(x)$ = Probabilidad de la distribución teórica

$P(x)$ = Probabilidad experimental o empírica de los datos, denominada también frecuencia acumulada.

El estadístico Δ tiene su función de distribución de probabilidades. Si Δ_0 es un valor crítico para un nivel de significación α , se tiene que:

$$P[\max |F(x) - P(x)| \geq \Delta_0] = \alpha \tag{6.80}$$

ó

$$P(\Delta \geq \Delta_0) = \alpha \tag{6.81}$$

también:

$$P(\Delta < \Delta_0) = 1 - \alpha \tag{6.82}$$

El procedimiento para efectuar el ajuste, mediante este método es el siguiente:

1º Calcular *P(x)* usando la formula de Weibull:

$$P(x) = \frac{M}{N+1} \tag{6.83}$$

Donde:

M = número de orden

N = número de datos

2º Calcular *F(x)* usando la función acumulada.

3º Calcular $F(x) - P(x)$, para todos los valores de *x*.

4º Seleccionar la máxima diferencia:

$$\Delta = \max\left|F(x) - P(x)\right| \tag{6.84}$$

5º Calcular el valor crítico del estadístico Δ, es decir Δ_0, para $\alpha = 0.05$ y N igual al número de datos. Los valores de Δ_0 se muestran en la tabla 6.1.

6º Comparar Δ con Δ_0:

Si $\Delta < \Delta_0 \Rightarrow$ el ajuste es bueno, al nivel de significación seleccionado.
Si $\Delta \geq \Delta_0 \Rightarrow$ el ajuste no es bueno, siendo necesario probar con otra distribución.

6.3.1.1 Ventajas y limitaciones

- No requiere un conocimiento a priori de la función de distribución teórica.
- Es aplicable a distribuciones de datos no agrupados, es decir, no se requiere hacer intervalos de clase.
- Es aplicable a cualquier distribución teórica.
- Se aplica en la función de distribución acumulada y no en la función de densidad.
- No es una prueba exacta sino aproximada.

Tamaño Muestral N	Nivel de significacion α				
1	0,900	0,925	0,950	0,975	0,995
2	0,684	0,726	0,776	0,842	0,929
3	0,565	0,597	0,642	0,708	0,828
4	0,494	0,525	0,564	0,624	0,733
5	0,446	0,474	0,510	0,565	0,669
6	0,41	0,436	0,470	0,521	0,618
7	0,381	0,405	0,438	0,486	0,577
8	0,358	0,381	0,411	0,457	0,543
9	0,339	0,360	0,388	0,432	0,514
10	0,322	0,342	0,368	0,410	0,490
11	0,307	0,326	0,352	0,391	0,468
12	0,295	0,313	0,338	0,375	0,450
13	0,284	0,302	0,325	0,361	0,433
14	0,274	0,292	0,314	0,349	0,418
15	0,266	0,283	0,304	0,338	0,404
16	0,258	0,274	0,295	0,328	0,392
17	0,25	0,266	0,286	0,318	0,381
18	0,244	0,259	0,278	0,309	0,371
19	0,237	0,252	0,272	0,301	0,363
20	0,231	0,246	0,264	0,294	0,356
25	0,21	0,22	0,24	0,27	0,32
30	0,19	0,20	0,22	0,24	0,29
35	0,18	0,19	0,21	0,23	0,27
N > 35	1,07/√N	1,14/√N	1,22/√N	1,36/√N	1,63/√N

Tabla 6.1 Valores críticos de Δ_0 para varios valores de N y niveles de significación α

Ejemplo 6.10

Dada la serie histórica de caudales medios anuales en m^3/s, que corresponde a un registro de 38 años, realizar la prueba de ajuste Smirnov – Kolmogorov, para ver si se ajustan a una distribución normal. El cálculo de valores de *F(x)* para todos los valores de *x* (donde *x* representa el caudal)

121,3	26,7	110,1	63,4	122,4	64,2	59,6
144,9	92,8	95,6	76,3	162,1	110,2	40,3
142,4	58,8	48,8	52,3	97,2	144,7	112,2
205,8	57,4	148,3	36,3	52,5	109,2	137,1
114,5	79	67,5	88	165,6	48,5	32,9
72,5	76,9	70				

Solución:

Calculo de *P(x)*

Ordenando los datos de caudales en forma creciente y calculando la probabilidad empírica $P(x)=\dfrac{m}{n+1}$ *;* se obtiene las columnas (2) y (3) de la tabla 6.2.

Luego se calcula:

$$\overline{X}=\frac{1}{n}\sum x_i=92.32$$

$$\sigma=\sqrt{\frac{1}{n-1}\sum(X-\overline{X})^2}=42.80$$

Calculando la variable estandarizada $Z=\dfrac{X-\overline{X}}{\sigma}$, se obtiene la columna (4) de la tabla 6.2.

Siendo *F(Z) = F(X)* se calcula *F(Z).*

A partir de las columnas (3) y (5) de la tabla 6.2, se obtiene $\Delta=\left|F(x)-P(x)\right|$, en la columna (6).

Se obtiene $\Delta=\Delta_{\max}=0.1089$, de la tabla 6.2.

Se calcula Δ_0; con valores de la tabla 6.1 para $\alpha = 0.05$, entonces se obtiene:

m 1	Q = X m^3/s 2	P(X) m/n+1 3	Z 4	F(Z) 5	Δ 6
1	26,7	0,0256	-1,53	0,0630	0,0374
2	32,9	0,0503	-1,39	0,0823	0,0310
3	36,3	0,0769	-1,31	0,0951	0,0182
4	40,3	0,1026	-1,22	0,1112	0,0086
5	48,5	0,1282	-1,02	0,1539	0,0257
6	48,8	0,1538	-1,02	0,1539	0,0001
7	52,3	0,1795	-0,94	0,1736	0,0059
8	52,5	0,2051	-0,93	0,1762	0,0289
9	57,4	0,2308	-0,82	0,2061	0,0247
10	58,8	0,2564	-0,78	0,2177	0,0387
11	59,6	0,2821	-0,76	0,2236	0,0585
12	63,4	0,3077	-0,68	0,2483	0,0594
13	64,2	0,3333	-0,66	0,2546	0,0787
14	67,5	0,359	-0,58	0,2810	0,0780
15	70	0,3846	-0,52	0,3015	0,0831
16	72,5	0,4103	-0,46	0,3228	0,0875
17	76,3	0,4359	-0,37	0,3557	0,0802
18	76,9	0,4615	-0,36	0,3594	0,1021
19	79	0,4872	-0,31	0,3783	0,1089
20	88	0,5128	-0,10	0,4602	0,0526
21	92,8	0,5385	0,01	0,5040	0,0345
22	95,6	0,5641	0,08	0,5319	0,0322
23	97,2	0,5897	0,11	0,5438	0,0459
24	109,2	0,6154	0,39	0,6517	0,0363
25	110,1	0,6410	0,42	0,6628	0,0218
26	110,2	0,6667	0,42	0,6628	0,0039
27	112,2	0,6923	0,46	0,6772	0,0151
28	114,5	0,7179	0,52	0,6985	0,0194
29	121,3	0,7436	0,60	0,7517	0,0081
30	122,4	0,7692	0,70	0,7580	0,0112
31	137,1	0,7949	1,05	0,8531	0,0582
32	142,4	0,8205	1,17	0,8790	0,0585
33	144,7	0,8462	1,22	0,8888	0,0424
34	144,9	0,8718	1,23	0,8907	0,0189
35	148,3	0,8974	1,31	0,9049	0,0075
36	162,1	0,9231	1,63	0,9484	0,0253
37	165,5	0,9487	1,71	0,9564	0,0077
38	205,8	0,9744	2,65	0,9960	0,0216

Tabla 6.2 Cálculo de resultados para ejercicio 6.10

$$\Delta_0 = \frac{1.36}{\sqrt{n}} = \frac{1.36}{\sqrt{38}} = 0.22$$

Una vez calculados todos los valores necesarios se sigue el criterio de decisión:

Como $\Delta = 0.1089 < \Delta_0 = 0.22$

Se concluye que los datos de caudales se ajustan a la distribución normal, con un nivel de significación del 5% o una probabilidad del 95%.

6.3.2 Ejemplos aplicativos

Ejemplo 6.11

Dada la serie histórica de caudales medios anuales, en m^3/s, que corresponde a un registro de 50 años para un río cualquiera. Calcular:

a) $P(Q \leq 180\ m^3/s)$; $P(Q \geq 100\ m^3/s)$;

b) El periodo de retorno para una caudal de $210\ m^3/s$

c) El caudal para un periodo de retorno de 50 años.

95.05	98.13	100.18	101.66	101.76
105.21	105.81	106.40	107.43	107.62
108.75	110.77	114.31	116.69	119.52
123.00	123.22	124.31	127.82	128.15
132.49	134.10	136.22	144.22	145.79
146.08	153.64	153.97	154.80	156.80
158.48	162.29	164.35	169.18	169.64
177.00	182.53	183.11	183.49	184.98
193.78	193.88	197.58	207.78	208.18
212.48	217.52	239.07	256.62	266.54

Solución:

a) Primeramente se calcula los parámetros (media y distribución estándar) por el método de los momentos y máxima verosimilitud:

$$\bar{Q} = \frac{1}{N}\sum_{i=1}^{N} Q_i = 152.2476 \ m^3/s$$

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(Q_i - \bar{Q})^2} = 43.6124$$

Realizando la prueba de ajuste se tiene:

$$\Delta = \max\left|F(x) - P(x)\right| = 0.1019$$

$$\Delta_0 = 0.1923, \ \ para \ \alpha = 0.05$$

Criterio de decisión:

Como: $\Delta = 0.1019 < \Delta_0 = 0.1923$

Se concluye que los datos se ajustan a la distribución normal con un nivel de significación del 0.05 o una probabilidad del 95%

Una vez realizado el ajuste se puede proseguir con el cálculo.

Calculo de: $P(Q \le 180) = F(Q = 180)$

Calculando *Z*, para *Q* = 180:

$$Z = \frac{Q - \overline{Q}}{\sigma} = \frac{180 - 152.2476}{43.6124} = 0.6363$$

$F(Q = 180) = F(Z = 0.6363) = 73.77$ *(valor interpolado)*

$\therefore P(Q \le 180) = 0.7377 = 73.77\%$

Calculo de: $P(Q \ge 100\ m^3/s)$

$$P(Q \ge 100) = 1 - P(Q < 100)$$
$$= 1 - F(Q = 100)$$

Calculando *Z*, para *Q* = 100

$$Z = \frac{100 - 152.2476}{43.6124} = -1.20$$

$F(Q = 100) = F(Z = -1.20) = 1 - 0.8849$ *(tablas)*

De donde; $P(Q \ge 100) = 1 - (1 - 0.8849)$

$\therefore P(Q \ge 100) = 0.8849$

b) El periodo de retorno para un caudal de 210 m^3/s

$$T = \frac{1}{1 - P(Q \le 210)}$$

Pero: $P(Q \leq 210) = F(Q = 210)$

$Si\ Q = 210 \Rightarrow Z = \dfrac{210 - 152.2476}{43.6124} = 1.3242$

$F(Q = 210) = F(Z = 1.32) = 0.9066$

De donde: $T = \dfrac{1}{1 - 0.9066}$

$\therefore T = 10.7\ \ años$

Esto significa que cada 10 años el caudal de 210 m^3/s será igualado o excedido.

c) Caudal para un periodo de retorno de 50 años

$P(Q \leq q) = 1 - \dfrac{1}{T} = 1 - \dfrac{1}{50}$

$P(Q \leq q) = 0.98 = F(Z)$

$Si\ F(Z) = 0.98 \Rightarrow Z = 2.055$ *(valor interpolado)*

Pero: $Z = \dfrac{Q - \overline{Q}}{\sigma} \Rightarrow Q = \overline{Q} + \sigma Z$

$Q = 152.2476 + 43.6124 \times 2.055$

$Q = 241.87\ \ m^3/s$

$\therefore$ El Q para un periodo de retorno de 50 años es 241.87 m^3/s

Ejemplo 6.12

Dada la serie histórica de caudales medios anuales, del ejemplo 8.1, calcular el caudal para un período de retorno de 75 años, utilizando la distribución log-normal.

Solución:

Se realiza primero el ajuste a la distribución concluyendo que los datos se ajustan a la distribución log-normal de dos parámetros con un nivel de significación del 5%, o una probabilidad del 95%. Un resumen de los resultados obtenidos es:

$\Delta = 0.0815$

$\Delta_0 = 0.1923$, para un nivel de significación del 5%

Se tiene: $\Delta = 0.0815 < \Delta_0 = 0.1923$

Como los datos se ajustan a la distribución se puede proseguir con los cálculos.

$$F(Q = q) = P(Q \leq q) = 1 - \frac{1}{T}$$

$$F(Q = q) = 1 - \frac{1}{75} = 0.986666 = F(Z)$$

De tablas, se obtiene por interpolación:

$Z = 2.215$

Pero para una distribución log-normal de 2 parámetros, la variable estandarizada es:

$$Z = \frac{\ln Q - \alpha}{\beta}$$

Por el método de la máxima verosimilitud, se obtiene:

$\alpha = 4.9872 \;\; y \;\; \beta = 0.2742$

Sustituyendo:

$$\frac{\ln Q - 4.9872}{0.2742} = 2.215$$

De donde:

$\ln Q = 0.2742 \times 2.215 + 4.9872$
$\ln Q = 5.594553$

Luego:

$Q = \varepsilon^{5.594553}$
$Q = 268.96\ m^3/s$

$\therefore$ El caudal para un periodo de retorno de 75 años, es $Q = 268.96\ m^3/s$

Ejemplo 6.13

Se tiene el registro de caudales máximos de 29 años, para la estación de un río, en el que se desea construir una presa de almacenamiento, calcular el caudal de diseño para el vertedor de demasías, con un período de retorno de 50 años. Usar la distribución Gumbel.

1660	917	3800	1410	2280
618	683	934	779	921
876	740	1120	610	1150
563	520	360	367	658
824	824	1230	522	581
557	818	1030	418	

Solución:

La media = 957.59
La desviación estándar = 682.72

Calculando los parámetros obtenemos:

$$\alpha = 532.3182$$
$$\beta = 650.3238$$

Calculando el caudal de diseño para un periodo de retorno de 50 años:

$$F(Q = q) = P(Q \leq q) = 1 - \frac{1}{T}$$
$$F(Q = q) = 1 - \frac{1}{50} = 0.98 = F(y)$$
$$F(y) = \varepsilon^{-\varepsilon^{-y}} = 0.98$$
$$-\varepsilon^{-y} = \ln 0.98$$
$$\varepsilon^{-y} = 0.020202707$$
$$-y = \ln(0.020202707)$$
$$y = 3.9019$$

Pero:

$$y = \frac{Q - \beta}{\alpha} = 3.9019$$
$$Q = \beta + 3.9019\alpha = 650.3238 + 3.9019 \times 532.3182$$
$$Q = 2727.38 \ m^3 / s$$

$\therefore$ El caudal de diseño para un periodo de retorno de 50 años es $Q = 2727.38 \ m^3 / s$

BIBLIOGRAFÍA

Moya C. Rufino – Saravia A. Gregorio. Probabilidad e inferencia estadística; Editorial Limusa.

Chungara Castro Víctor. Estadística y Probabilidades; Editorial Leonardo.

Romero Mauricio. Texto de Probabilidad y Estadística; U.M.S.S.

Manchego C. Roberto. Estadística II; U.M.S.S.

Máximo Villón Béjar. Hidrología Estadistica; Lima Perú.

CAPÍTULO VII

TEORÍA DE LA ESTIMACIÓN ESTADÍSTICA

CAPÍTULO VII

TEORIA DE LA ESTIMACIÓN ESTADISTICA

7.1 INTRODUCCIÓN

Definición 7.1 ***Los Parámetros*** *de una distribución teórica, son variables que para cada conjunto de datos tienen un valor definido. Una vez que los parámetros quedan definidos, también queda definida la distribución teórica.*

Definición 7.2 *Dada una función de distribución con parámetros* $\alpha, \beta, \gamma, \ldots,$ *se llaman* ***estimadores*** *a los valores a, b, c, ..., obtenidos a partir de los estadísticos de la muestra que se supone pertenece a la población que se pretende caracterizar.*

El proceso de estimación básicamente cosiste en los siguientes pasos:

i. Seleccionar un estimador para inferir el parámetro deseado del conjunto o universo bajo estudio.
ii. Seleccionar una muestra de este conjunto.
iii. Valorar al estimador de la muestra seleccionada.
iv. Inferir, de este valor, el parámetro buscado de ese universo.

La estimación de un parámetro puede adoptar la forma de un solo punto, es decir la estimación de un solo punto del parámetro de interés (estimación puntual); o de un intervalo es decir de un rango de valores dentro del cual se espera el valor del parámetro (estimación por intervalo).

7.2 ESTIMACIÓN PUNTUAL

Definición 7.3 *Cualquier estadístico $Y = G(X_1, X_2, \ldots, X_n)$ cuyo valor $y = G(x_1, x_2, \ldots, x_n)$ se utiliza para estimar el parámetro θ, se dice que es un estimador de θ y se escribe así:*

$$\widehat{\theta} = G(X_1, X_2, \ldots, X_n) \tag{7.1}$$

En síntesis ***estimación puntual*** *del parámetro de la población, es un valor numérico que se obtiene a partir de alguna formula conocida mediante el uso de una muestra de la población.*

Por ejemplo, la muestra de la tabla 7.1 tiene una media de 55.82 m^3/s, la misma que representa una estimación puntual para la media μ de la población correspondiente.

Como ejemplo se puede citar: Dada una muestra de 20 datos de caudales anuales, en m^3/s, de una población (estación de aforo), que se indican en la tabla 7.1, al calcular sus parámetros estadísticos se obtienen:

Media:	55.82
Varianza:	666.84
Desviación estándar:	25.82
Coeficiente de sesgo:	0.42
Coeficiente de variación:	0.46
Coeficiente de curtosis:	2.27

30.1	26.5	80.4	50.3	60.2
27.8	90.6	34.1	17.7	48.3
50.4	90.1	50.6	44.4	33.3
100.1	85.3	95.2	40.2	60.8

Tabla 7.1. Muestra de 20 datos de caudales, en m^3/s de una población

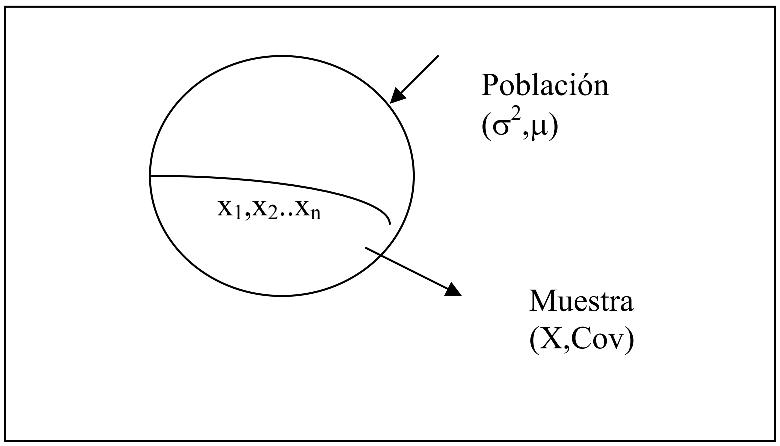


Figura 7.1 Grafica de población y muestra

7.2.1 Propiedades de un Estimador

7.2.1.1 Estimación insesgada

Un estimador $\widehat{\theta} = G(X_1, X_2, ..., X_n)$ se dice que es insesgado para el parámetro θ, si $E(\widehat{\theta}) = \theta$.

Si no se cumple con esta definición, el estimador se dice que es sesgado.

7.2.1.2. Estimadores eficientes

Si las distribuciones muestrales de dos estadísticos tienen la misma media, el estadístico que tenga menor varianza es el estimador eficiente de la media, el otro estimador es el estimador no eficiente.

Los valores correspondientes de estos estadísticos son estimaciones eficientes o no eficientes respectivamente.

Si se consideran todos los posibles estadísticos, cuyas distribuciones muestrales tiene la misma media, al que tiene menor varianza se llama el más eficiente o mejor estimador de esta media.

7.2.2 Estimación puntual por el Método de Máxima Verosimilitud

El método de máxima verosimilitud fue desarrollado por R. A. Fisher (1922).

El principio de máxima verosimilitud, en su aplicación a la obtención de estimadores puntuales, consiste en seleccionar aquel estimador que maximice la probabilidad de obtener la muestra realmente observada.

Definición 7.4 *Sea X la variable aleatoria cuya función de probabilidad,* $f(x;\theta)$ *depende del parámetro* θ. *Sea* $X_1, X_2, ..., X_n$ *una muestra aleatoria de X y sean* $x_1, x_2, ..., x_n$ *los valores observados de la muestra.* ***La función de verosimilitud*** *de la muestra se define así:*

$$V(\theta) = \prod_{i=1}^{n} f(x_i;\theta) \qquad (7.2)$$

$$V(\theta) = f(x_1, x_2, ..., x_n; \theta) = f(x_1; \theta) f(x_2; \theta) ... f(x_n; \theta) \tag{7.3}$$

El método de máxima verosimilitud consiste en tomar como valor estimado de θ, el valor que hace máxima la función $V(\theta)$.

Entonces, si $\widehat{\theta}$ hace máxima a $V(\theta)$, entonces también hace máxima a su logaritmo $\ln V(\theta)$. Para convertir el producto en suma, se suele tomar la función $L = \ln V(\theta)$ y determinar θ por la ecuación:

$$\frac{\partial L}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \ln f(x_i; \theta)}{\partial \theta} = 0 \tag{7.4}$$

Si la distribución de probabilidad tiene varios parámetros desconocidos, $\theta_1, \theta_2, ..., \theta_k$ entonces en lugar de una ecuación, tendremos *k* ecuaciones:

$$\frac{\partial L}{\partial \theta_1} = 0 \quad ; \frac{\partial L}{\partial \theta_2} = 0 \quad ; ...; \frac{\partial L}{\partial \theta_k} = 0 \tag{7.5}$$

A partir de los cuales se obtiene las estimaciones para estos parámetros.

Las propiedades de los estimadores calculados por el método de la máxima verosimilitud, son:

- Usualmente insesgado
- Si la eficiencia de estimadores existe para los parámetros, el método puede producirlos
- La solución de la ecuación de verosimilitud, proporciona un estimador que converge al valor poblacional, cuando el tamaño muestral tiende a infinito, por lo

que el estimador es consistente.

Ejemplo 7.1

Suponer que el número diario de ventas de autos nuevos que efectúa determinado concesionario, es una variable aleatoria de Poisson con parámetro λ. Dado que en 20 días la venta total de autos fue de 30, ¿Cuál es el estimador de máxima verosimilitud de λ?

Solución:

Se comienza determinando, el estimador de máxima verosimilitud del parámetro λ de la distribución de Poisson.

$$p(x,\lambda)=\frac{\lambda^{x}\varepsilon^{-\lambda}}{x!};\qquad x=0,1,2,...$$

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de tamaño n de la variable aleatoria de Poisson, con valores $x_1, x_2, \ldots, x_n$. Entonces:

$$p(x_i,\lambda)=P[X=x_i]=\frac{\lambda^{x_i}\varepsilon^{-\lambda}}{x_i!};\qquad i=1,2,...,n$$

La función de verosimilitud es:

$$V(\lambda)=\prod_{i=1}^{n}p(x_i,\lambda)=\prod_{i=1}^{n}\frac{\varepsilon^{-\lambda}\lambda^{x_i}}{x_i!}=\frac{\varepsilon^{-n\lambda}\lambda^{\Sigma x_i}}{\prod_{i=1}^{n}x_i!}$$

Tomando logaritmo a $V(\lambda)$ se obtiene:

$$L = \ln V(\lambda) = -n\lambda + \sum_{i=1}^{n} x_i \ln \lambda - \ln\left(\prod_{i=1}^{n} x_i!\right)$$

Derivando con respecto a λ se tiene:

$$\frac{\partial L}{\partial \lambda} = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} = 0$$

De donde $\quad \hat{\lambda} = \dfrac{\sum_{i=1}^{n} x_i}{n} = \overline{x}$

Expresado en palabras, la media muestral es el estimador de máxima verosimilitud del parámetro λ de la distribución de Poisson. Por lo tanto empleando los datos del problema, n = 20 días, $\sum_{i=1}^{20} x_i = 30$ autos, se tiene:

$$\hat{\lambda} = \frac{30}{20} = 1.5$$

7.2.3 Estimación puntual por el método de los momentos

El método de los momentos fue desarrollado por Karl Pearson (1902). El principio básico de la estimación por este método, es establecer para cada función de distribución, la relación entre los parámetros y los momentos centrales, de tal manera que:

$$\begin{aligned} \theta_1 &= f_1(\mu_i, \mu_{i+1}, \ldots) \\ \theta_2 &= f_2(\mu_j, \mu_{j+1}, \ldots) \\ \theta_3 &= f_3(\mu_k, \mu_{k+1}, \ldots) \end{aligned} \qquad (7.6)$$

Donde:

$\theta_1, \theta_2, \theta_3$ Son los parámetros de la función de distribución.

μ_i, μ_j, μ_k Son los momentos con respecto a la media, o momentos centrales de la población.

Cuando la distribución de probabilidad a la que se estiman los parámetros por este método es simétrica y particularmente si es normal, se puede demostrar que este es un método muy eficiente, pero cuando las distribuciones son asimétricas y por lo tanto sesgadas, como sucede muy a menudo con la mayoría de las variables hidrológicas, el utilizar este método representa una perdida de eficiencia en la estimación.

Ejemplo 7.2

Sea $X_1, X_2, \ldots, X_n$ una muestra de una población con distribución de Poisson con parámetro λ. Estimar λ por el método de los momentos.

Solución:

La distribución de Poisson tiene solamente un parámetro, en consecuencia se tendrá solo una ecuación.

El primer momento poblacional es $\mu_1^{'} = E(X) = \lambda$

El primer momento muestral es $M_1^{'} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{X}$

De los dos momentos se tiene la ecuación:

$\mu_1^{'} = \lambda = M_1^{'} = \overline{X}$

En consecuencia $\widehat{\lambda} = \overline{X}$

Nota: El método de máxima verosimilitud teóricamente es el más correcto para el cálculo de los parámetros de las distribuciones, en el sentido en que produce menos errores en la estimación de los parámetros de la población. Pero para algunas distribuciones de probabilidad no existe solución analítica para todos los parámetros en términos de los estadísticos de la muestra lo cual lleva a calcular estos parámetros por métodos numéricos, para esto el método de los momentos es el mas indicado.

7.3 ESTIMACIÓN POR INTERVALO

La precisión de la estimación puntual puede evaluarse en una muestra por estimación de un intervalo junto con una medida de la seguridad que tal intervalo contenga al parámetro desconocido de la población. Dichos intervalos se llaman intervalos de confianza o estimación por intervalo del parámetro desconocido.

Definición 7.5 Sea $X_1, X_2, \ldots, X_n$ *una muestra aleatoria de una población X, cuya distribución de probabilidad es* $f(x,\theta)$.

Sea $\theta_1 = G(X_1, X_2, ..., X_n)$ *y* $\theta_2 = G(X_1, X_2, ..., X_n)$ *dos estadísticos tales que* $\theta_1 < \theta_2$, *para los cuales se cumple:*

$$P[\theta_1 \leq \theta \leq \theta_2] = \gamma = 1 - \alpha \tag{7.7}$$

Donde γ no depende de θ; entonces el intervalo aleatorio $[\theta_1, \theta_2]$ se llama el intervalo del 100γ % *ó* $100(1-\alpha)\%$ de confianza para θ.

La ecuación 7.7 se interpreta como sigue: γ es igual a la probabilidad que el intervalo aleatorio $[\theta_1, \theta_2]$ contenga a θ.

θ_1, se llama el limite inferior de confianza para θ.

θ_2, se llama el limite superior de confianza para θ.

$\gamma = (1-\alpha)$, se llama el nivel (o coeficiente) de confianza.

Los valores mas utilizados para γ son:

$$\gamma = 0.90; \qquad \gamma = 0.95; \qquad \gamma = 0.975; \qquad \gamma = 0.98; \qquad \gamma = 0.99$$

Por ejemplo si se hacen los cálculos correspondientes la media de la población μ, de la muestra de la tabla 7.1, con una probabilidad del 95% (nivel de confianza), está en el intervalo: $43.7362 \leq \mu \leq 67.9038$.

7.3.1 Estimación por intervalos de confianza de parámetros poblacionales

Sean μ_s y σ_s la media y desviación estándar de la distribución muestral de un estadístico δ.

- Si la distribución muestral de δ es aproximadamente normal (suposición cierta para muchos estadísticos si $N \geq 30$); el estadístico δ.
- Se encuentran los intervalos $\mu_s - \sigma_s$ a $\mu_s + \sigma s$, $\mu_s - 2\sigma_s$ a $\mu_s + 2\sigma_s$, o $\mu_s - 3\sigma_s$ a $\mu_s + 3\sigma_s$, el 68.27 %, 95.45 % y 99.73 % respectivamente.
- Análogamente , se puede confiar en encontrar μ en los intervalos $\delta - \sigma_s$, a $\delta + \sigma_s$, $\delta - 2\sigma_s$, a $\delta + 2\sigma_s$, o $\delta - 3\sigma_s$, a $\delta + 3\sigma_s$, el 68.27 %, 95.45 % y 99.73 % de las veces respectivamente.

Análogamente se tienen:

$$\delta \pm 1.96\sigma_s \rightarrow 95\ \%$$

Para δ

$$\delta \pm 2.58\sigma_s \rightarrow 99\ \%$$

El porcentaje de confianza se llama también nivel de confianza

Los números 1.96, 2.88, etc. De los límites de confianza se llaman coeficientes de confianza o valores críticos, se denotan por *Zc*.

La siguiente tabla presenta los distintos niveles de confianza utilizados en la práctica.

Nivel de confianza	99,73%	99%	98%	96%	95,45%	95%	90%	80%	68,27%	50%
Zc	3,00	2,58	2,33	2,05	2,00	1,96	1,645	1,28	1,00	0,6745

Tabla 7.2 Niveles de Confianza utilizados en la práctica

7.3.2 Estimación de medias por intervalos de confianza

Si el estadístico δ es la media muestral $\overline{X}$, entonces los limites de confianza del 95 % y 99 % para la estimación de la media poblacional μ, vienen dados por $\overline{X} \pm 1.96\sigma_{\overline{X}}$ y $\overline{X} \pm 2.88\ \sigma_{\overline{X}}$ respectivamente.

Generalmente los limites de confianza están dados por $\overline{X} \pm z_c \sigma_{\overline{X}}$ donde z_c depende del nivel de confianza.

Suponiendo que son extraídas de una población finita todas las posibles muestras sin reemplazamiento de tamaño N, y el tamaño de la población el N_p $(N_p > N)$.

Si:

$\mu_{\overline{X}}$ = media de la distribución muestral de medias.

$\sigma_{\overline{X}}$ = desviación estándar de la distribución muestral de medias.

μ = Medias de la población.

σ = Desviación estándar de la población.

$$\mu_{\overline{X}} = \mu \tag{7.8}$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}}\sqrt{\frac{N_p - N}{N_p - 1}} \tag{7.9}$$

Si la población es infinita o si el muestreo es con reemplazamiento. La $\mu_{\overline{X}}$ es igual que en la ecuación 7.8.

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}} \tag{7.10}$$

Para una población finita los intervalos de confianza:

$$\bar{X} \pm Z_c \frac{\sigma}{\sqrt{N}} \sqrt{\frac{N_p - N}{N_p - 1}} \qquad (7.11)$$

Para una población infinita o muestreo con reemplazamiento:

$$\bar{X} \pm Z_c \frac{\sigma}{\sqrt{N}} \qquad (7.12)$$

En general σ, es desconocida para obtener limites de confianza anteriores, se utiliza la estima muestral $\hat{S}$ o $S \Rightarrow$ Aproximación satisfactoria para $N \geq 30$. Si $N < 30$, aproximación mala.

7.3.3 Intervalos de confianza para proporciones

Si S es estadístico es la proporción de éxitos en una muestra de tamaño N extraída de una población binomial, donde p es la proporción de éxito (probabilidad de éxito)

Los límites de confianza para $p \rightarrow P \pm z_c \sigma_p$, donde P es la proporción de éxitos en la muestra de tamaño N están dados por:

Los limites de confianza para la proporción poblacional són:

A. *Para caso de población infinita o con reemplazamiento en población finita*

$$P \pm z_c \sqrt{\frac{pq}{N}} = P \pm \sqrt{\frac{p(1-p)}{N}} \qquad (7.13)$$

B. *Para caso de muestreo sin reemplazamiento en población finita de tamaño N_p*

$$P \pm z_c \sqrt{\frac{pq}{N}} \sqrt{\frac{N_p - N}{N_p - 1}} \qquad (7.14)$$

Para calcular estos limites de confianza se puede utilizar la estimación muestral P para p (Satisfactorio para $N \geq 30$).

7.3.4 Intervalos de confianza para diferencias y sumas

Si S_1 y S_2 son dos estadísticos con distribuciones muestrales aproximadamente normales, los límites de confianza para la diferencia de los parámetros poblacionales correspondientes a S_1 y S_2 viene dados por:

$$S_1 - S_2 \pm z_c \sigma_{S_1 - S_2} = S_1 - S_2 \pm z_c \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2} \qquad (7.15)$$

Para la suma:

$$S_1 + S_2 \pm z_c \sigma_{S_1 - S_2} = S_1 + S_2 \pm z_c \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2} \qquad (7.16)$$

Nota: Ambas muestras deben ser independientes

7.3.5 Intervalos de confianza para desviaciones estándar

Los limites de confianza para desviación típica σ de una población que se distribuye normalmente, y que es estimada por una muestra con desviación típica *S*:

$$S \pm z_c \sigma_S = S \pm z_c \frac{\sigma}{\sqrt{2N}} \tag{7.17}$$

Para calcular estos limites de confianza, se utiliza S *o* $\hat{S}$ para estimar σ.

7.3.6 Error probable

Los limites de confianza del 50% de los parámetros de la población, correspondientes a un estadístico S están dados por $S \pm 0.6745\sigma_S$.

Nota: La cantidad $0.6745\sigma_S$ *es el error probable de estimación.*

BIBLIOGRAFÍA

Moya C. Rufino – Saravia A. Gregorio. Probabilidad e inferencia estadística; Editorial Limusa.

Romero Mauricio. Texto de Probabilidad y Estadística; U.M.S.S.

ANEXO A

MANUAL DE MINITAB

MANUAL DE INTRODUCCION AL MANEJO Y ANALISIS DE DATOS CON MINITAB v. 13

Fuente: Ing. M.Sc. Wilson Trujillo A.

I. INTRODUCCION

La literatura consagrada al paquete estadístico MINITAB v.13 es abundante y diversificada provocada precisamente por el uso cada vez más frecuente de esta potente herramienta de gestión y análisis de datos. Es muy fácil encontrar libros enteramente consagrados al MINITAB v.13, manuales de uso, documentos de curso y actualmente una abundante literatura en formato “electrónico” en el Internet.

Durante los dos últimos años en los que hemos estado involucrados en la formación y enseñanza de la estadística utilizando el MINITAB v.13, se ha realizado una revisión detallada de varios documentos.

Producto de esta revisión hemos constatado, la existencia de buenos manuales “informáticos” que carecen de un soporte teórico estadístico sólido representando un gran peligro para los utilizadores con bases estadísticas deficientes.

Así mismo hemos encontrado documentos que consagran demasiadas “hojas” a los conceptos estadísticos con una “tímida” presentación de las opciones del MINITAB v.13. Generalmente estos documentos son rápidamente abandonados precisamente por su carácter poco práctico o aplicable.

Un problema mayor que hemos podido constatar en esta literatura es la falta de un modelo conceptual claro de la gestión y el análisis de datos provocando que el utilizador de MINITAB v.13 interesado en manejar esta herramienta se sienta frustrado.

La intención de este manual es precisamente tratar de resolver estos problemas al menos de manera parcial. El propósito de este meanual es servir de documento de referencia a personas interesadas en utilizar el paquete estadístico MINITAB v.13 para manejar y analizar sus datos desde una perspectiva “utilizador”. Tratamos de mantener un equilibrio entre los aspectos informáticos y teóricos señalando cada vez que sea posible algunas alertas y cuidados que debe considerar el utilizador. Debe quedar completamente claro que este no es un manual del MINITAB v.13 sino un manual de gestión y análisis de datos con la ayuda del MINITAB v.13.

ALGUNAS CONSIDERACIONES TEORICAS

Como se ha mencionado en la introducción de este manual el propósito de este capitulo es ofrecer al utilizador un marco conceptual del proceso de análisis de datos. A menudo los utilizadores del MINITAB v.13, no saben como realizar el análisis de sus datos. Se recomienda al utilizador leer el articulo en formato “electrónico” relativo al protocolo de observación [Villarroel, 2003]. El artículo puede se telecargado en la dirección http://www.fcyt.umss/investiga/cesa/protocolo.html.

Es importante que el interesado en realizar el análisis de sus datos utilizando el MINITAB v.13 o cualquier herramienta informática consagrada al análisis estadístico comprenda de manera sencilla y clara lo que implica la estadística en su dimensión análisis de datos.

La estadística debe ser comprendida como un conjunto de herramientas orientadas a analizar (aprovechar) datos, su rol en las diferentes áreas de las ciencias es central.
De manera general la estadística tiene tres grandes “áreas”: la Estadística Descriptiva, la Inferencia Estadística y un “puente” entre ambas, la teoría de Probabilidades en cual área el MINITAB v.13 es una de las herramientas más potentes, noción de variable aleatoria y distribuciones de probabilidad teóricas.

La estadística descriptiva reúne un conjunto de herramientas orientadas a resumir los datos. Estas herramientas se clasifican según el número de variables consideradas simultáneamente y la naturaleza de la variable. Considerando el número de variables se distinguen herramientas de resumen unidimensional, cuando se considera una sola variable; herramientas de resumen bidimensional cuando se consideran dos variables al mismo tiempo; finalmente herramientas de resumen multidimensional cuando se consideran más de dos variables simultáneamente. Considerando la naturaleza de los datos, estos se clasifican en herramientas para datos cuantitativos y cualitativos. Las herramientas de resumen de datos deben seleccionarse dependiendo del número de variables que se consideran simultáneamente y la naturaleza de los datos.

La inferencia estadística reúne un conjunto de herramientas orientadas a generalizar resultados obtenidos en una muestra a nivel poblacional. Las herramientas se distinguen para resolver los problemas de estimación y los problemas de *test* de hipótesis. La selección de la herramienta también es función del número de variables manejadas simultáneamente y la naturaleza de los datos. En la fase inferencial es importante considerar las condiciones de aplicación de los métodos estadísticos.

Antes de realizar los análisis orientados a resumir los datos o inferir los mismos a nivel poblacional, es muy importante que el utilizador realice un análisis exploratorio de sus datos orientado a identificar datos influyente. *Exploring Data Analysis* – EDA reúne una serie de herramientas dirigidas a controlar la calidad de los datos. Es evidente que la selección de una herramienta de control de calidad de los datos debe ser seleccionada en función del número de variables que se manejan simultáneamente y la naturaleza de los datos.

El MINITAB v.13 ofrece una serie de estas herramientas de análisis de datos. El manejo de esta herramienta implica saber “hacer clic”. Un manejo responsable del MINITAB v.13 requiere de una sólida base teórica en estadística.

II. GESTION DE ARCHIVOS Y DATOS

1. Descripción de las ventanas del MINITAB.-

En MINITAB todos los datos asociados con un Estudio o Proyecto son contenidos en una Hoja de Trabajo (*worksheet*). Un proyecto puede tener muchas hojas de trabajo, el número de hojas esta limitado solo por la memoria del computador.

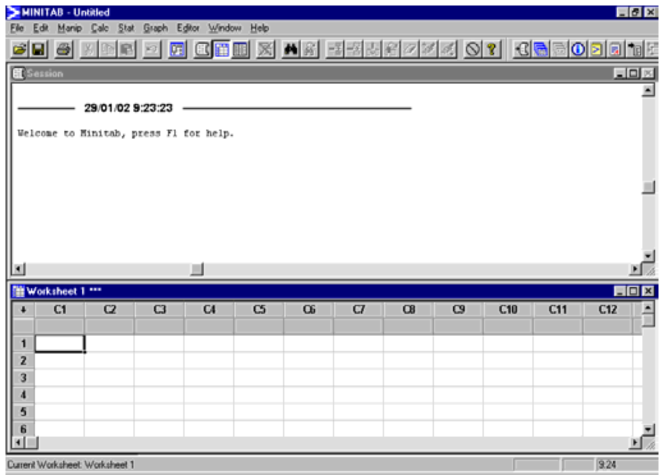


Figura 1: Hoja de Trabajo del MINITAB

Las hojas de trabajo no son visibles, pero se pueden apreciar los datos en la Sesión windows, y en la hoja fólder del Manejador de Proyectos. Cada hoja se contendrá dentro el DATA windows. Como se puede observar la pantalla se divide en dos hojas de trabajo una que permite registrar y gestionar las variables, y observaciones y la otra que muestra las salidas de los procedimientos estadísticos realizados con los datos, a diferencia de otros paquetes estadísticos.

Existe también un tercer tipo de ventana de solo gráficos, y dispone además de la opción de edición, esta opción se habilita haciendo un doble clic en el grafico generado.

2. Tipos y formas de Datos que se utilizan.-

Una hoja de trabajo puede contener tres tipos de datos, numéricos, texto y de tiempo / fechas, en tres formatos: columnas, constantes o matrices.

TIPOS:

- Los datos son numéricos ?, Manejo de Datos Numéricos.

  Los valores pueden ser positivos ó negativos y pueden contener un separador decimal, si los datos son de forma exponencial la notación adoptada es, 3.2E12. Por defecto, MINITAB automáticamente cambia el formato de columnas base en otros contenidos. Por ejemplo, si una columna contiene un número que es mas grande o muy pequeño, MINITAB cambia el formato de la columna y usa una notación exponencial y todos los números aparecen en este formato.

  Precisión Numérica; el MINITAB muestra y computa números a doble precisión es decir con 15 o 16 dígitos.

- Los datos son Nombres ?, Manejo de Datos de Texto.

  Los datos de texto pueden ser por ejemplo: Sucre, Empresa A, etc.
  En lo que son datos de texto sus características son:
  - Hasta 80 caracteres de largo
  - Elabora cualquier carácter
  - Se los puede colocar en columnas y constantes pero no en matrices.

- Los datos son Fechas ?, Manejo de Datos de Tiempo.

  Los datos de tiempo / fecha pueden ser de la forma: 17/3/02 08:25:22 am.

FORMAS:

Los datos pueden tener tres formas: Columnas, Constantes o Matriz, como se describe en el cuadro siguiente:

Cuadro 2.1: Formas de los Datos

Formato	Contenido	Referido por		Limite
		Numero	**Nombre**	
Columna	Numérico, texto o fecha / tiempo	C + un número, como c1 o c22	Por ejemplo tenemos "Años"	Depende de la memoria
Constante	Un simple número o una cadena de texto.	K + un número, como k1 o k93	Por ejemplo tenemos "Primero"	1000
Matriz	Un bloque rectangular o celdas contenidas por números	M + un número, como M1 o M44	Por ejemplo "Inversa de M1"	100

Fuente: Guía de uso del MINITAB v.13

3. Valores Perdidos (Missing Value).-

Un valor perdido se explica como aquel que no pudo codificarse por falta de información o falta de respuesta.

El MINITAB usa un asteristico (*) para identificar un valor perdido en columnas de datos numéricos y además en los de tiempo / fecha y un espacio en blanco para los datos de texto.

4. Escribiendo y editando datos.-

Los Datos entran automáticamente en formatos de columna. Cuando usted teclea una entrada en una columna vacía, MINITAB asigna un tipo de dato a la columna: numérico, texto, o fecha / tiempo. Si el tipo de dato no es numérico, MINITAB también agrega una identificación al lado del número de la columna: C1-D para los datos de fecha / tiempo y C1-T para los datos del texto.

Cada columna generalmente representa una variable. Más usted puede cambiar datos de un tipo a otro.

Entrada de datos por columnas, filas y bloques.-

Al hacer clic en esta **opción** se obtiene la posibilidad de darle dirección a la entrada de Datos.

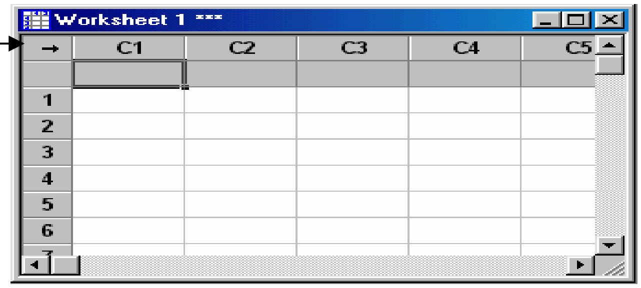


Figura 2: Hoja de Datos del MINITAB

La introducción de datos se la puede realizar de varias formas por columnas, por filas o por bloques.

Al estar activada en la **opción** descrita e introduciendo un dato en la hoja de trabajo el desplazamiento del cursor será hacia la próxima columna.

Sin embargo, basta hacer el clip en el indicador para cambiar el desplazamiento del cursor por filas en la introducción de datos.

En lo que respecta a la introducción de datos por bloques se hace un barrido de las celdas a introducir y se procede a la inserción.

5. Comandos del Menú.-

Menú de Atajo: Pulse el botón derecho en una ventana de MINITAB para abrir el menú del atajo.

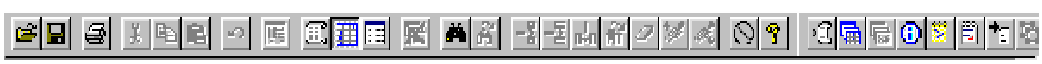


Figura 3: Opciones del menú MINITAB

Nota: Si un artículo del menú se oscurece, significa que en ese momento no esta indisponible.

6. Trabajando con Proyectos.-

Un proyecto de MINITAB contiene todo un trabajo: los datos, salidas, texto de las órdenes, los gráficos, etc. Cuando usted guarda el proyecto salva todo su trabajo. Se puede tener un solo proyecto abierto en un momento.

Haciendo una descripción completa podemos tener:

- Las columnas de datos en cada ventana de los Datos.
- Cada ventana del Gráfico.
- El texto completo en la ventana de la Sesión y el plegador de la Historia.
- Las constantes guardadas, matrices, y objetos del plan, eso se resume en el plegador de hojas de trabajo.
- Cualquier eslabón del archivo Guardado en el plegador de los documentos relacionados.

⇒ Para ver la descripción del Proyecto: Se procede con el comando, FILE -> Project Description.

⇒ La descripción de cada hoja de trabajo puede ser creada con: EDITOR -> Worksheet Description.

⇒ Para abrir un nuevo Proyecto, la opción es: FILE -> Open Project.

⇒ Para salvar el Proyecto, la opción es: FILE -> Save Project.

El Proyecto puede manejarse en muchas secciones individuales. El software puede crear datos, gráficos, y salidas dentro del MINITAB. La aplicación también puede agregar datos y gráficos al proyecto copiándolos de otros archivos. Pueden imprimirse separadamente los proyectos, en una variedad de formatos del archivo. También puede desechar una hoja de trabajo o gráfico.

7. Descripción General del Menú Principal

Entre las muchas opciones que se dan las mas frecuentes de uso son:

Figura 4: Opciones del menú MINITAB

(**FILE – ARCHIVOS**):
New (Nuevo): Ejecución de apertura de un nuevo Proyecto.
Open Proyect (Abrir el Proyecto): Ejecución de apertura de un proyecto ya elaborado anteriormente.
Save Proyect (Guardar Proyecto): Ejecución para salvar o guardar un proyecto en elaboración.
Save Proyect as (Guardar Proyecto como): Ejecución para salvar o guardar por primera vez un proyecto en elaboración.

(EDIT – EDICIÓN):
Undo (Retroceder): Esta opción nos permite retorcer en un paso de la secuencia lógica que adoptamos al realizar una tarea "x".
Clear Cells (Limpieza de Celdas): Borrado de los datos descritos en una o varias Celdas.

Delete Cells (Borrado de Celdas): Borrado de una fila o columna de la Base de Datos seleccionada.
Copy Cells (Copiado de Celdas): Copiado de Celdas seleccionadas para un uso posterior.
Cut Cells (Cortado de Celdas): "Cortado" de Celdas seleccionadas para un uso posterior.
Paste Cells (Pegado de Celdas): Pegado de las celdas copiadas o cortadas para su uso.

(MANIP - MANEJO):
Copy Columns (Copia de Columnas): Hace una copia de una Columna sobre otra, sobreponiendo los datos.
Transpose Columns (Transposición de Columnas): Realiza la transposición de todas las columnas en filas, es decir de una matriz m x n a una n x m por ejemplo.
Chance Data Type (Cambio en el tipo de Datos): Realiza el cambio de "Tipo de dato", por ejemplo tipo Texto por datos de tipo Numérico y de forma inversa.

(CALC - CALCULAR): Esta es la **Cuarta Opción del Menú Principal** del Minitab, las opciones más utilizadas son las de calculator y de Distribuciones de Probabilidad-Probability Distributions:

(STAT – HERRAMIENTAS ESTADÍSTICAS): Entre las muchas herramientas,

Boxplot (Grafico de Caja): Realiza el grafico de Dispersión para variables cuantitativas.
Stem and Leaf (Grafico de Hojas y Tallos): Hace la descripción de Datos mediante los caracteres que se repiten en las decenas.
Tally (Tabla de Frecuencias): Genera tabla de frecuencias para variables de tipo Cualitativo.
Cross Tabulation (Tabulación de Cruce de variables): Realiza la Estadística Descriptiva a dos Dimensiones entre variables cualitativas mostrando los porcentajes condicionadas por las filas, columnas y totales.
Display Descriptive Statistics (Estadística Descriptiva): Realiza la Estadística Descriptiva a una dimensión de una Variable Cuantitativa.
1–Sample Z (Test Z de Conformidad): Dentro de la Estadística Inferencial esta el test de Conformidad sobre la base de la Curva Normal.
1–Sample t (Test t de Conformidad): Dentro de la Estadística Inferencial esta el test de Conformidad sobre la base de la "t" de Student, como los demás test: **2–Sample t** (Test t de Comparación de dos promedios Indepen.); **Paired t** (Test t de Comparación de dos promedios Dependientes); **1 Proportion** (Test de Conformidad de una Proporción); **2 Proportion** (Test de Comparación de dos Proporciones).
Correlation (Correlación de dos variables Cuantitativas): Análisis de asociación lineal entre dos variables Cuantitativas.
Normalyti Test (Test de Normalidad): Esta opción nos permite medir el grado de aproximación a la distribución Normal que tiene los datos.

(EDITOR – EDICIÓN): trabajando sobre la hoja electrónica de celdas:
Find, opción que permite encontrar un valor especifico en todas las celdas.

Re place, reemplaza valores encontrados en la hoja de trabajo por un valor deseado.
Go To, posiciona el cursor en un numero de columna y fila deseado.
Go To (Next Column, Active Cell), posiciona el cursor en la próxima columna de su posición original.
Insert Columns, Inserta una nueva columna.
Insert Rows, Inserta una nueva fila.
Insert Cells, Inserta una nueva celda.

III. ANÁLISIS EXPLORATORIO DE DATOS Y ESTADÍSTICA DESCRIPTIVA

1. ANÁLISIS EXPLORATORIO DE DATOS UNIDIMENSIONAL.

El análisis exploratorio de datos, *Display Descriptive Statistics* reúne una serie de herramientas para realizar el control de calidad de los datos. Las herramientas se seleccionan en función del número de variables consideradas simultáneamente y el tipo de dato. Es importante notar que ningún procedimiento más elaborado se justifica si no se realiza una verificación de la calidad de los datos.

a) Herramientas de control de calidad de datos unidimensional-variable cuantitativa.

Una de las herramientas más utilizadas cuando se considera una sola variable de naturaleza cuantitativa es el Box Plot*. El* Box Plot *en su contexto de herramienta de control de calidad identifica dos tipos de datos influyentes:* outliers *y el* extrem value*. Estos valores son identificados en el gráfico por un pequeño circulo y un asterisco respectivamente. Por la tanto la presencia de este tipo de símbolos son indicadores que existen datos influyentes.*

Un dato es declarado como *outliers* cuando este se aleja por debajo del cuartil 1 o encima del cuartil 3 en 1.5 veces el rango intercuartil (RIQ). Un *extrem value* es aquel dato que supera los limites igual a 3 veces el rango intercuartil por encima del cuartil 3 o por debajo del cuartil 1. Se debe notar que el rango intercuartil corresponde gráficamente a la altura de la caja en el diagrama *Box Plot*.

La figura presenta un diagrama de *Box Plot*, en el cual se identifica la presencia de valores influyentes.

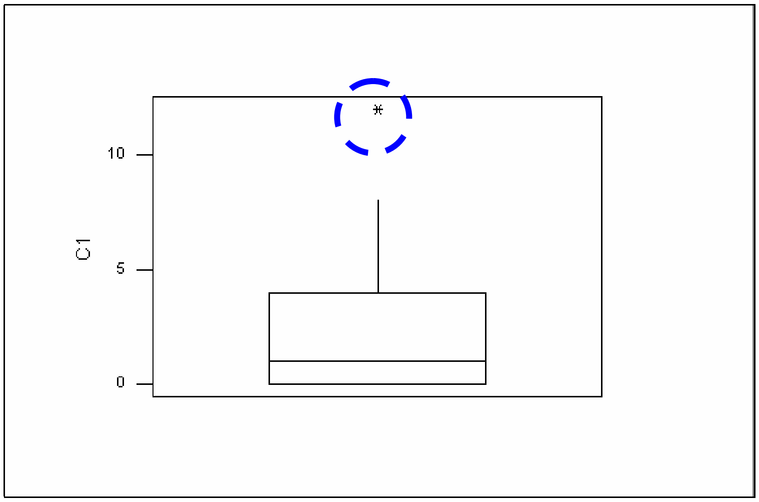


FIGURA 5: BOX PLOT – MINITAB V.13

b) Herramientas de control de calidad de datos unidimensional-variable cualitativa.

EN EL CASO DE VARIABLES DE NATURALEZA CUALITATIVA LA HERRAMIENTA QUE SE UTILIZA PARA CONTROLAR LA CALIDAD DE LOS DATOS ES LA *TABLA DE FRECUENCIAS*. LA PRESENCIA DE UN DATO QUE NO CORRESPONDE A LOS NIVELES DE LA VARIABLE CUALITATIVA ES RÁPIDAMENTE PUESTA EN EVIDENCIA EN UNA TABLA DE FRECUENCIA UNIDIMENSIONAL. EN LA FIGURA SE PRESENTA UNA TABLA DE FRECUENCIA EN LA CUAL SE IDENTIFICA LA PRESENCIA DE UN DATO MAL DIGITADO CON NIVEL IGUAL A 11, SIENDO LAS DOS ÚNICAS OPCIONES SECO O LLUVIOSO.

Tally for Discrete Variables: dia

dia	Count	CumCnt	Percent
11	1	1	4,76
Lluvioso	11	12	52,38
Seco	9	21	42,86
N=	21		

CUADRO 3.1. TABLA DE FRECUENCIA. HERRAMIENTA DE CONTROL DE CALIDAD UNIDIMENSIONAL PARA DATOS CUALITATIVOS.

c) *Herramientas de control de calidad bidimensionales – variables cuantitativas.*

PARA EL CONTROL DE CALIDAD DE DOS VARIABLES CUANTITATIVAS LA HERRAMIENTA ADECUADA ES EL *SCATTER PLOT,* ESTA HERRAMIENTA

NOS PRESENTA A LOS DATOS EN UNA GRÁFICA A DOS DIMENSIONES (UN PLANO), DONDE SE PUEDE APRECIAR LA DISPERSIÓN DE LOS DATOS EN UNA NUBE DE PUNTOS. LA FIGURA PRESENTA UN DIAGRAMA DE *SCATTER PLOT,* EN ESTE GRÁFICO SE OBSERVA LA PRESENCIA DE DATOS INFLUYENTES.

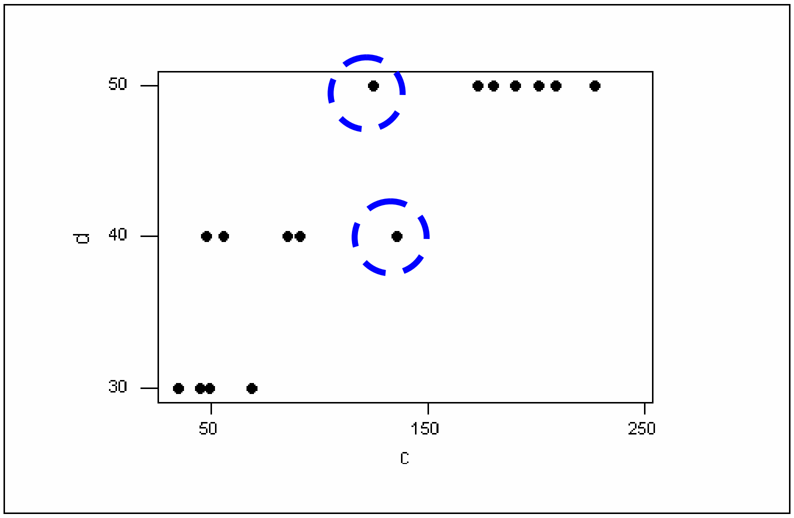


FIGURA 6. SCATTER PLOT. HERRAMIENTA DE CONTROL DE CALIDAD BIDIMENSIONAL PARA DATOS CUANTITATIVOS.

d) *Herramientas de control de calidad bidimensionales –variables cualitativas.*

Cuando se desea realizar el control de calidad de datos que corresponden a dos variables cualitativas se usa las tablas de frecuencia bidimensional, donde se evidencia un "cruce" de las variables, una de las variables se denomina "variable de control" cuyas categorías formaran las filas de la tabla del Crosstab. Esta tabla fundamentalmente permite identificar datos mal digitados. El cuadro siguiente muestra un ejemplo de una tabla de frecuencias a dos dimensiones. En esta se puede apreciar que existe una variable "mal digitada" que aparece con el número 11.

	jueves	lunes	martes	miercoles	viernes	All
11	0	1	0	0	0	1
	--	100,00	--	--	--	100,00
	--	20,00	--	--	--	4,76
	--	4,76	--	--	--	4,76
Lluvioso	3	2	2	1	3	11

	27,27	18,18	18,18	9,09	27,27	100,00
	75,00	40,00	50,00	25,00	75,00	52,38
	14,29	9,52	9,52	4,76	14,29	52,38
Seco	1	2	2	3	1	9
	11,11	22,22	22,22	33,33	11,11	100,00
	25,00	40,00	50,00	75,00	25,00	42,86
	4,76	9,52	9,52	14,29	4,76	42,86
ALL	4	5	4	4	4	21
	19,05	23,81	19,05	19,05	19,05	100,00
	100,00	100,00	100,00	100,00	100,00	100,00
	19,05	23,81	19,05	19,05	19,05	100,00

Cell Contents --
Count
% of Row
% of Col
% of Tbl

Cuadro 3.2. tabla de frecuencia bidimensional. herramienta de control de calidad bidimensional para datos cualitativos.

e) *Herramientas de control de calidad multidimensional.*

La presentación de herramientas de control de calidad multidimensional no es objeto de esta publicación, sin embargo se recomienda utilizar el gráfico de individuos del Análisis en Componentes Principales – ACP, en el caso de variables cuantitativas. por el contrario si las variables son de naturaleza cualitativa se recomienda realizar tablas de frecuencia por variable.

2. ANÁLISIS DESCRIPTIVO DE DATOS.

El análisis descriptivo de datos reúne una serie de herramientas de la estadística descriptiva orientadas a resumir y describir los datos. Las herramientas se seleccionan en función del número de variables consideradas simultáneamente y el tipo de dato. Por lo tanto, es importante diferenciar, según el número de variables, las herramientas de resumen unidimensional cuando se considera una sola variable, herramientas de resumen bidimensionales, cuando se consideran dos variables y finalmente, herramientas de resumen multidimensionales cuando se consideran mas de dos variables. En cada una de estas situaciones la naturaleza del dato constituye una condición de aplicación de una u otra herramienta.

De manera general las herramientas de resumen de información se clasifican en tres categorías:

- Distribuciones de frecuencia o también denominadas tablas de frecuencia
- Representaciones gráficas

- Parámetros o estadígrafos

a) Herramientas de resumen Unidimensional – Variable Cuantitativa.

Si la variable es cuantitativa el concepto de distribución de frecuencia no tiene sentido, tal que al limite uno puede imaginarse una tabla infinita.

Las representaciones gráficas más comunes para esta situación son el diagrama *box plot* y el histograma en su forma continua.

⇒ **Ejemplo**

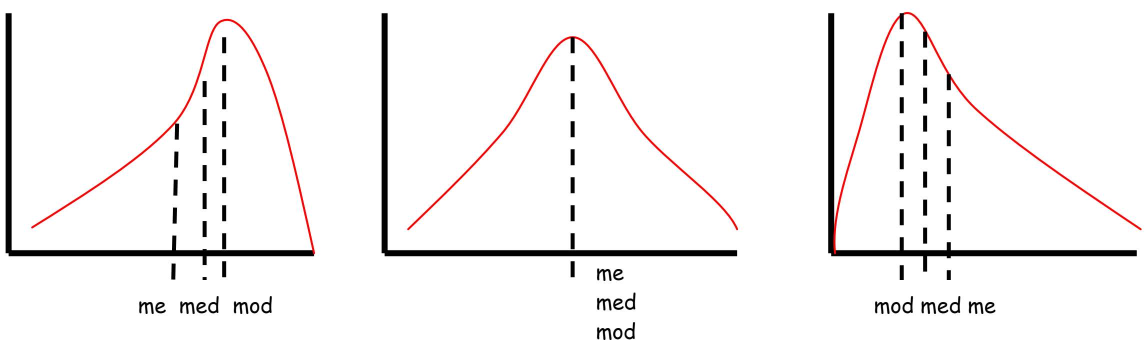
El numero de accidentes ocurrido en un mes dado en los 13 departamentos de manufactura de una planta industrial fue: 2,0,0,3,3,12,1,0,8,1,0,5,1. Realizar una Descripción de los datos.

⇒ **Concepto – definición.**

La estadística Descriptiva Unidimensional comprende las técnicas que se emplean para resumir y describir datos numéricos asociadas a una única variable. Estos métodos implican análisis gráficos y los resultados computacionales.

Relación entre Media, Mediana y Moda.-

En toda distribución simétrica, la media, mediana y moda coinciden en valor. En una distribución asimétrica positiva, el promedio siempre es mayor que la moda. En una distribución asimétrica negativa, el promedio siempre es menor que la moda. En ambos casos la mediana siempre se encontrara entre el promedio y la moda. Una medida de asimetría en estadística (basada en la diferencia entre los valores de la media y la mediana de un grupo de valores) es el coeficiente de asimetría de Fisher (Skewness).



Asimetría Negativa (g1<0) Simétrica Asimetría Positiva (g1>0)

Considerando el uso de las tres medidas de posición en relación con datos asociados a muestras. La moda no es una medida aceptable de tendencia central de datos muéstrales, por que su valor puede variar ampliamente de una muestra a otra, como también el promedio.

Herramientas - Niveles de Análisis:

Después de obtener las graficas para diferentes muestras del mismo tipo, se procede a los niveles de análisis y/o interpretación de los datos:

1. *Análisis de la Posición*, considerando el comportamiento de la mediana en el Box Plot, respecto del promedio.
   Para este análisis tenemos la opción, **stat / Basic Statistics / Display Descriptive Statistics**;
2. *Análisis de Dispersión y de Datos influyentes (aberrantes)*, respecto del valor del Rango Interquartilico.
   Este análisis se lo hace por medio de la herramienta grafica, **Graph / Boxplot**, y muestra la amplitud del rango interquartil, como la presencia de valores influyentes.
3. *Análisis de la forma de la distribución*, para identificar la naturaleza (Positiva o Negativa) de la asimetría o forma de la distribución se parte de la generación del Histograma, **Graph / Histogram;**

⇒ **Procedimiento MINITAB**

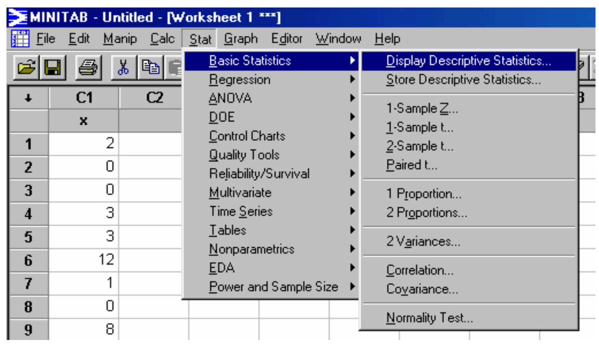


Figura 7: Opciones del menú Stat/ Estadística Descriptiva

Los Resultados son:

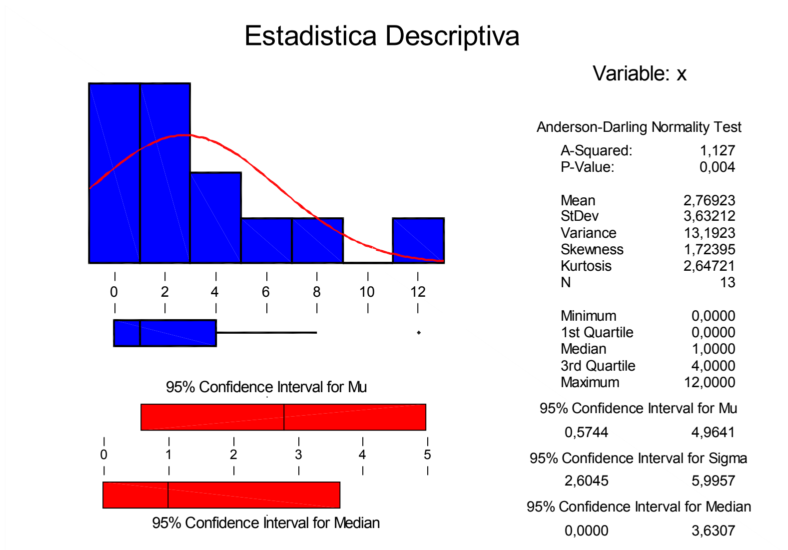


Figura 8: Salida de la opción Display Descriptive Statistic.

Interpretación

Análisis de Posición; Los datos muestran que existe un diferencia entre los valores del promedio (2.77 accidentes) y la mediana (1 accidente), por lo que, nos hace suponer que existe una influencia por valores extremos en la serie de datos.

Análisis de Dispersión y Datos Influyentes; La grafica del BoxPlot nos muestra la dispersión de los datos y la existencia de un valor aberrante, su valor es 12, el cual se presume crea la influencia sobre el valor de la media.

Análisis de Forma de la Distribución; La distribución tiene una asimétria (positiva) hacia la izquierda, lo que nos dice que la mayor parte de los accidentes ocurren por debajo del promedio muestral.

b) Herramientas de resumen unidimensional –variable cualitativa

La tabla de frecuencia es una herramienta que resume los datos de una variable cualitativa en una distribución de frecuencias. Esta tabla proporciona información sobre la frecuencia absoluta (numero de veces que se repite una valor), el porcentaje, porcentaje valido (valores que no toman en cuenta los *missing value* o datoEs perdidos), el porcentaje acumulado, esta información para variables cualitativas de naturaleza nominal. Para el caso de variables cualitativas ordinales eventualmente se puede calcular los parámetros de la mediana, moda y percentiles.

⇒ **Ejemplo**

Se realiza una medición de caudales en tres diferentes bombas la que deberá abastecer un mínimo requerido por el sistema de 300 m^3/día, basándose en los datos recabados de la toma de caudales por cada bomba de igual de potencia. ¿Existe la cantidad de observaciones iguales o suficientes para realizar la comparación, además realicé una descripción del tipo de día de medición con mayor frecuencia para la toma de mediciones?.

Datos de la medición:

Cuadro 3.1: Datos observados de medición

	Diámetro (mm)	**Caudal (m3/día)**	**Día de Medición (Variable: m)**
1	50	173.3	Lluvioso
2	40	48	Seco
3	40	90.9	Seco
4	50	226.7	Lluvioso
5	50	125.3	Seco
6	40	85.3	Seco

7	30	49.8	Lluvioso
8	40	48	Seco
9	50	180.2	Seco
10	40	90.9	Lluvioso
11	30	69.1	Seco
12	50	208.7	Lluvioso
13	50	201.3	Lluvioso
14	40	55.8	Lluvioso
15	40	135.8	Lluvioso
16	30	49.8	Lluvioso
17	50	190.1	Seco
18	30	35.1	Seco
19	30	44.9	Lluvioso
20	30	45.1	Lluvioso
21	30	69.1	Lluvioso

Fuente: Ejercicios de aplicación MINITAB

⇒ **Concepto – definición.**

Una distribución de frecuencias es una función que asigna a todo valor o conjunto de valores, de una variable su frecuencia de ocurrencia. Se la puede representar a través de una tabla, un grafico o un modelo. Los datos organizados en una distribución de frecuencias se llaman datos agrupados. En contraste con ello, en el caso de datos no agrupados se enlistan todos los valores observados de la variable aleatoria.

Los Intervalos de Clase de una distribución de Frecuencias están definidos por, los limites nominales de clase inferior y superior, la presentación de datos se lo hace con el siguiente formato:

Atributo	**Tarjas**	**Frecuencia Absoluta**	**Frecuencia Relativa**
A1	**N**	n_i	h_i
…	**…**	**…**	**…**
An	**N**	$\sum_{i=1}^{n} n_i = n$	$h_i = \frac{n_i}{n} * 100$

⇒ **Procedimiento MINITAB**

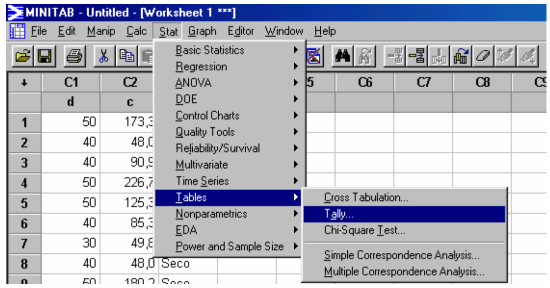


Figura 8: Opciones del menú Stat/ Tablas.

Los resultados son:

Tally for Discrete Variables (Tabla de Variables Discretas): m

m	**Count**	**CumCnt**	**Percent**	**CumPct**
Lluvioso	12	12	57,14	57,14
Seco	9	21	42,86	100,00
N=	21			

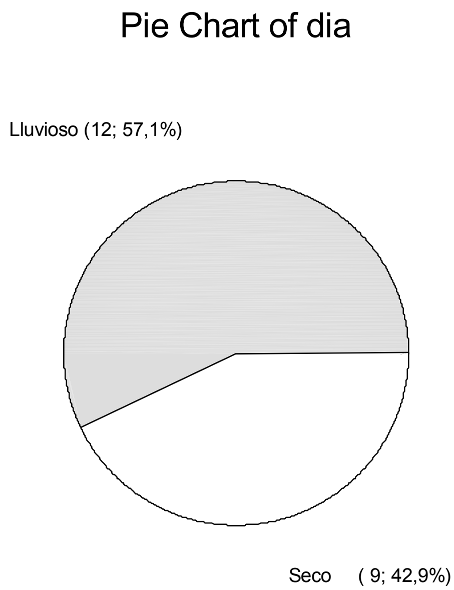
Interpretación:

Existen 12 mediciones en días lluviosos que representan un 57,14% de los 21 informes, los días secos corresponden a 9 mediciones hechas, que representan un 42,86% del total de mediciones.

La frecuencia de días lluviosos es mayor en un 13% que la toma de mediciones en días secos lo que nos hace suponer que no existe igual cantidad de registros, esta herramienta solo es descriptiva y nos explica un resumen de los datos obtenidos.

La herramienta grafica en este caso será el diagrama de torta (pie), puesto que esta herramienta se utiliza en variables de tipo Nominal y no así ordinal.

Para el diagrama de torta se tiene la opción, **Graph / Pie Chart**, la grafica es:



El grafico hace también la misma descripción de la tabla de frecuencias, pero el efecto producido es distinto, por que trata de provocar un impacto visual.

c) Herramientas de resumen bidimensional –variables cuantitativas.

Las herramientas usadas para el resumen de información de dos variables cuantitativas fundamentalmente son los análisis de correlación y regresión en su contexto descriptivo.

El análisis de correlación permite apreciar el grado de asociación entre dos variables bajo el supuesto la relación es lineal y las variables sean cuantitativas.

Para cuantificar la relación entre dos variables cuantitativas se usa el coeficiente de correlación de Pearson que oscila entre –1 y 1, donde el valor de "0" representa la relación nula o independencia entre las variables, 1 relación perfecta y positiva, y –1 relación perfecta y negativa. Este coeficiente es una medida del grado de asociación entre dos variables pero no necesariamente un coeficiente alto conlleva que exista una alta relación entre las variable. El siguiente ejemplo muestra un análisis de correlación.

⇒ Ejemplo

La materia prima que se utiliza en la producción de una fibra sintética se almacena en un lugar que no tiene control de humedad, las medidas de la humedad relativa y del contenido de humedad de muestras de la materia prima ambos en porcentajes en doce días dieron los siguientes resultados:

Humedad Relativa	Contenido de Humedad
46	12
53	14
37	11
42	13
34	10
29	8
60	17
44	12
41	10
48	15
33	9
40	13

Se busca medir la correlación de las variables y además pronosticar el contenido de humedad cuando la relativa es del 38%.

⇒ CONCEPTO – DEFINICIÓN

ANÁLISIS DE CORRELACIÓN; EN LA INVESTIGACIÓN FRECUENTEMENTE NOS INTERESA RESUMIR FUERZAS DE LA ASOCIACIÓN ENTRE DOS VARIABLES QUE EN ESTE CASO SERÁN NUMÉRICAS. LA HERRAMIENTA ES EL COEFICIENTE DE CORRELACIÓN DE PEARSON " R ".

$$r = \frac{COV(X,Y)}{S_X S_Y} \qquad COV(X,Y) = \frac{1}{n-1}\left\{\sum xy - \frac{1}{n}\left(\sum x\right)\left(\sum y\right)\right\}$$

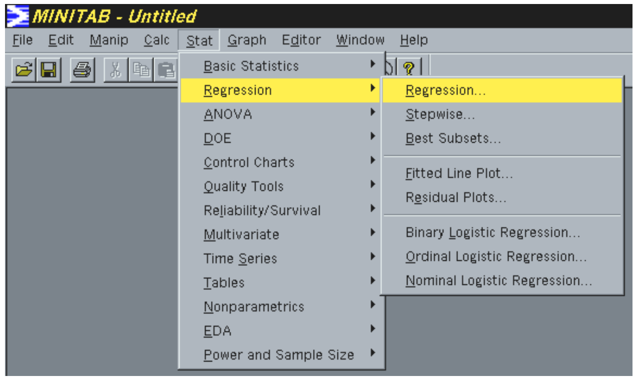
Regresión Lineal Simple; el modelo contempla una parte determinista en la que Y está determinada por completo por X. El valor de Y puede predecirse perfectamente si se conoce "a" y "b" sin embargo, en la investigación muy pocas relaciones son deterministas. De modo que el procedimiento de regresión agrega un término de error que debe tomarse en cuenta para la naturaleza probabilística o causal de la relación. La ecuación de regresión básica se convierte en:

$Y_i = a + b\chi_i$

La estimación de los parámetros de regresión, a y b , es relativamente sencilla.

$$a = \bar{y} - b\bar{x} \qquad b = \frac{\operatorname{cov}(x,y)}{Sx^2}$$

⇒ **Procedimiento MINITAB**



Los resultados son:

La ecuación de regresión es:

Contenido de humedad (Y) = 0.49 + 0.272 Humedad relativa (X)

Predictor	**Coef**	**SE Coef**
Constant	0.490	1.579
Humedad	0.272	0.03665

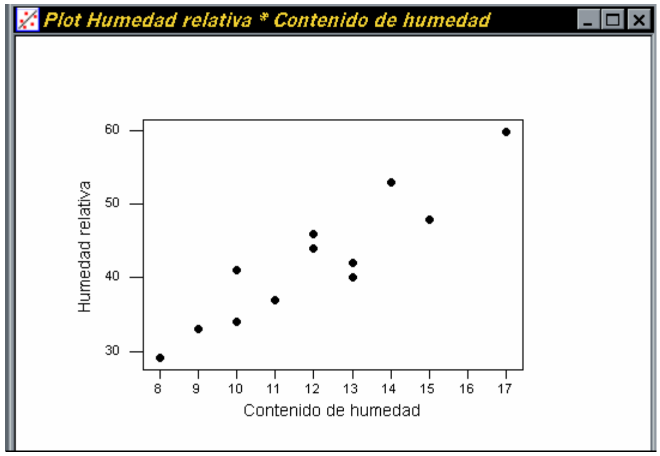
S = 1.065 R-Sq = 84.7% R-Sq(adj) = 83.1%

Interpretación

Como el valor de la correlación es positiva significa que nos encontramos en el primer caso descrito en la parte conceptual y los datos siguen una tendencia lineal ascendente.

La correlación entre las dos variables es de 0.92 ó $(0.847)^{1/2}$, es decir que existe un grado de asociación del 92% (los cuales se encuentran dentro de la recta de regresión proyectada).

La herramienta grafica que se utiliza en este caso es **Graph / Plot**, lo que muestra:



d) herramientas de resumen bidimensional –variables cualitativas

EL OBJETIVO DE ESTAS HERRAMIENTAS ES EL DE RESUMIR Y ANALIZAR LOS DATOS DE DOS VARIABLES CUALITATIVAS AL MISMO TIEMPO. SE USAN TABLAS A DOS DIMENSIONES COMO LA SIGUIENTE:

Y / X	Y_1	Y_2	Y_3	Y_i
X_1				
X_2		$\mathbf{N_{ij}}$		
X_3				
Xj				

n_{ij} (frecuencias)

⇒ **Ejemplo**

Un fabricante de televisores tienen los datos de una encuesta hecha a 10 hombres y 10 mujeres donde se les pregunto cual es su preferencia ante los televisores de pantalla pequeña (de menos de 20 pulgadas), pantalla mediana (de 20 a 27 pulgadas), pantalla grande (de más de 27 pulgadas), a fin de determinar los adecuados programas de producción para el mes siguiente.

Pantalla Pequeña: 1
Pantalla Mediana: 2
Pantalla Grande: 3

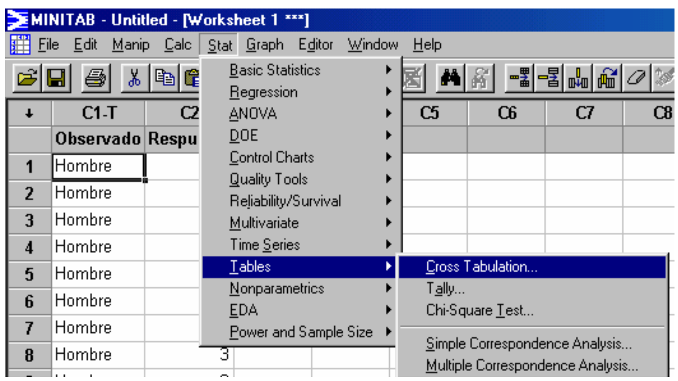
Hombres		**Mujeres**	
1	1	3	2
1	1	3	2
2	1	2	1
2	1	2	1
2	3	2	3
1	2	1	3

3	3	1	3
3	2	1	1
3	1	1	2
1	2	3	2

⇒ **Concepto – definición**

Cuando están implicadas dos variables de tipo cualitativas, las frecuencias observadas se organizan en una tabla de doble clasificación o tabla de Contingencia. Las dimensiones de estas tablas están definidas por m x n, donde m indica el numero de líneas y n el numero de columnas, y se describen los porcentajes de frecuencia para cada variable de observación con su variable respuesta.

⇒ **Procedimiento MINITAB**



Los resultados son:

Rows: Observad Columns: Respuest

	1	2	3	All
Hombre	9	6	5	20
	45,00	30,00	25,00	100,00
	56,25	46,15	45,45	50,00
	22,50	15,00	12,50	50,00
Mujer	7	7	6	20
	35,00	35,00	30,00	100,00
	43,75	53,85	54,55	50,00
	17,50	17,50	15,00	50,00
All	16	13	11	40
	40,00	32,50	27,50	100,00
	100,00	100,00	100,00	100,00
	40,00	32,50	27,50	100,00

Cell Contents --

Count
% of Row
% of Col
% of Tbl

Interpretación

Son 9 de 20 encuestados hombres que prefieren pantalla pequeña de televisor, es decir, que el 45% de los hombres encuestados prefieren pantallas pequeñas (Frec. Condicionada en Fila) y 56.25% (Frec. Condicionada en Columna) de los que prefieren pantallas pequeñas son hombres.

Para las mujeres son 6 de cada 20 que prefieren pantallas grandes, es decir, que el 30% de ellas los prefieren, y son 54.55% de los que prefieren pantallas grandes son mujeres.

Son 13 personas encuestas que prefieren pantallas medianas, es decir son el 32.50% personas encuestadas de un total de 40 encuestadas.

IV. DISTRIBUCIONES DE PROBABILIDAD.

Una distribución de probabilidad es una función que asigna a todo valor o conjunto de valores de una variable aleatoria su probabilidad correspondiente.

1. DISTRIBUCIONES DISCRETAS.

Una distribución de tipo discreta es aquella en la que se analizan, el comportamiento del numero de éxitos en un experimento repetitivo de sucesos independientes, para una probabilidad asociada al evento conocido.

a) Distribución Binomial

⇒ Ejemplo

Se ha lanzado una moneda tres veces:

a) Determinar la probabilidad de obtener dos caras.
b) La probabilidad de obtener 2 ó mas caras.

⇒ Concepto – definición

Si repetimos “n” veces, en condiciones idénticas de modo independiente, una experiencia aleatoria en la que el resultado de aparición del evento A ocurre con una probabilidad “p” o la no aparición del evento A con una probabilidad 1-p, el número X de apariciones del evento A en “n” experiencias aleatorias independientes ($0 \leq X \leq n$) sigue una distribución Binomial de parámetros n y p, ($X \Rightarrow$ Bin (n,p)).

Condiciones de uso de la distribución.

- El número de pruebas es Fijo.
- Los resultados del experimentó son dicotómicos: Éxito y Fracaso.
- La probabilidad de éxito en cada prueba es constante y se designa por p, donde la probabilidad de fracaso es q = 1 - p.
- Las pruebas son independientes del resultado anterior o posterior a una prueba, no influyen en su resultado.

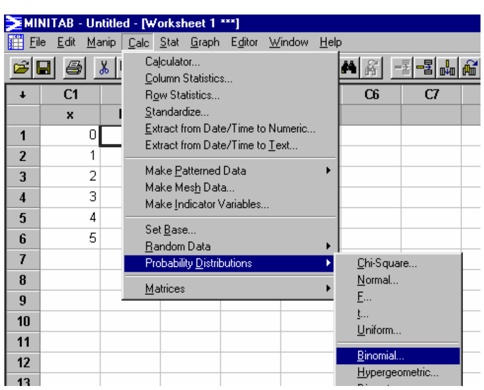
Teniendo en cuenta las características anteriores la probabilidad de un experimento que concuerdan con dichas características se obtiene aplicando la expresión:

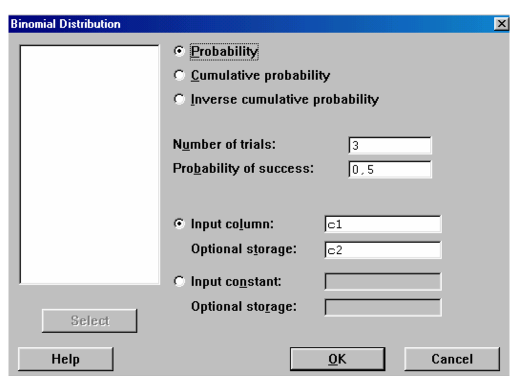
$$p(x) = \left(\frac{n}{x}\right) * p^{x} * q^{n-x}$$

n: Numero de pruebas del experimento.
X: Variable Aleatoria.
p: Probabilidad de aparición del evento.
q: Probabilidad de no aparición del evento (1-p).

⇒ **Procedimiento MINITAB**

- Se efectúa un numero de pruebas fija igual a 3.
- Los resultados del experimento se clasifican en:
    - Obtener Cara – Éxito.
    - Obtener Sello – Fracaso.
- La probabilidad de tener éxito para obtener cara en cada prueba es de ½ es decir en constante.
- Que una prueba salga cara no considera precisamente que en la prueba anterior también salga cara, las pruebas son independientes.





Interpretación

a) Existe un 37.5 % de probabilidad de obtener dos caras en los tres lanzamientos.

b) La probabilidad de obtener dos o más caras es:
P(x >= 2) = p(x=2) + p(x=3) = 0.375 + 0.125 = 0.5 = 50% .

b) Distribución Hipergeométrica

⇒ **Ejemplo**

En un curso de 9 alumnos, 5 son del Loyola y 4 son del San Agustín. Si estamos interesados en los alumnos del Loyola se obtiene una muestra de 3 alumnos. Determine:

a) Cual es la probabilidad de obtener exactamente un alumno del San Agustín.
b) Cual es la probabilidad de obtener 2 del Loyola a lo más.

⇒ **Concepto – definición**

Para obtener una formula análoga a la distribución Binomial que sea válida para el muestreo sin reemplazo (ensayos no independientes) se obtiene una variable aleatoria que sigue una distribución de tipo Hiper geométrica con parámetros (N, N1, n).

Condiciones de uso de la distribución:

- El número de pruebas es Fijo.
- Los resultados del experimentó son dicotómicos: Éxito y Fracaso.
- La probabilidad de éxito en cada prueba es variable.
- Las pruebas son dependientes, es decir un resultado es afectado por el anterior o el siguiente.

Las probabilidades de los eventos de un experimento que concuerdan con dichas características se resuelven aplicando la expresión:

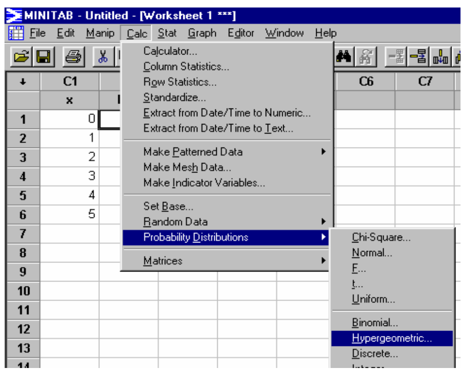
$$p(x) = \frac{\left(\frac{N1}{x}\right)\left(\frac{N-N1}{n-x}\right)}{\left(\frac{N}{n}\right)}$$

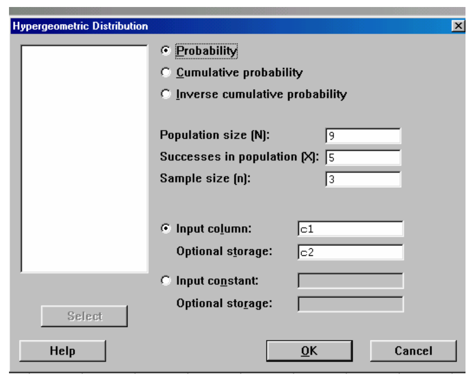
N: Total de elementos de la Prueba.
N1: Es el numero de elementos de una clase.

n: Numero de pruebas del experimento.
x: Variable Aleatoria.

⇒ **Procedimiento MINITAB**

- Se efectúa un numero de pruebas fija igual a 3.
- Los resultados del experimento se clasifican en:
  - o Obtener alumnos del Loyola – Éxito.
  - o Obtener alumnos del San Agustín – Fracaso.
- La probabilidad de ser Loyola es variable en cada prueba, por tanto si es diferente la probabilidad de ser del San Agustín.
- Las pruebas son dependientes, este resultado se ve afectado por la prueba aleatoria y afecta al siguiente. El procedimiento es:





Interpretación

El 48% de los casos se podrán elegir a los más 2 estudiantes del Loyola, es decir 52% (1-0.48) de probabilidad de escoger un estudiante del San Agustín.

La probabilidad de elegir a lo más 2 estudiantes del Loyola es:
P(x <= 2)= p(x=0)+p(x=1)+p(x=2)= 0.0476 + 0.3571 + 0.4761 = 0.88.

c) Distribución Poisson

⇒ **Ejemplo**

Maria recibe en promedio 8 llamadas telefónicas por día. Cual es la probabilidad de que un día en particular reciba menos de 8 llamadas.

⇒ **Concepto – definición**

Un experimento aleatorio sigue una Distribución de Poisson cuando reúne todas las características de una distribución Binomial en la que la probabilidad de éxito es Pequeña y el tamaño de la muestra es grande es decir: $(n \times p)\rangle 5$

La distribución de Poisson se utiliza con frecuencia en variables aleatorias distribuidas en el tiempo y espacio.

Condiciones de uso de la distribución Poisson:

- La variable aleatoria X es el número de ocurrencias de un suceso durante cierto intervalo.
- Las ocurrencias deben ser aleatorias.
- Las ocurrencias deben ser independientes unas y otras.
- Las ocurrencias deben estar uniformemente distribuidas dentro el intervalo empleado.

Las probabilidades de los eventos de un experimento que concuerdan con dichas características se resuelven aplicando la expresión:

$$p(x,\mu) = \frac{e^{-\mu}.\mu^{x}}{x!} \qquad \text{(ecuación 3.3)}$$

μ : Promedio asociada a la variable aleatoria x.
e: Base logarítmica = 2.71828.
x: Variable Aleatoria.

La distribución de poisson difiere de la distribución binomial en estos importantes aspectos:

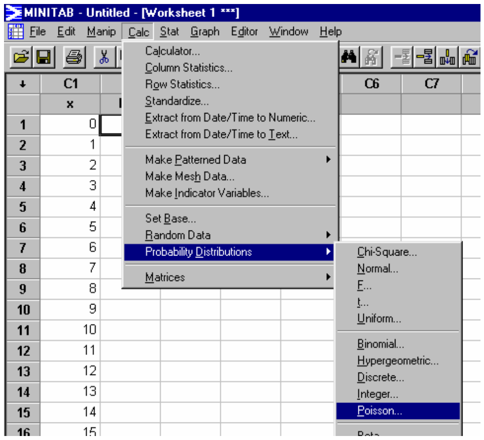
1. La distribución binomial es afectada por el tamaño de muestra *n* y la probabilidad *p*, mientras que la distribución de Poisson sólo es afectada por la media μ.
2. En una distribución binomial, los posibles valores de la variable aleatoria X son 0,1,2,....,n, pero una distribución de Poisson tiene como posibles valores de X 0,1,2,3,..... sin limite superior.

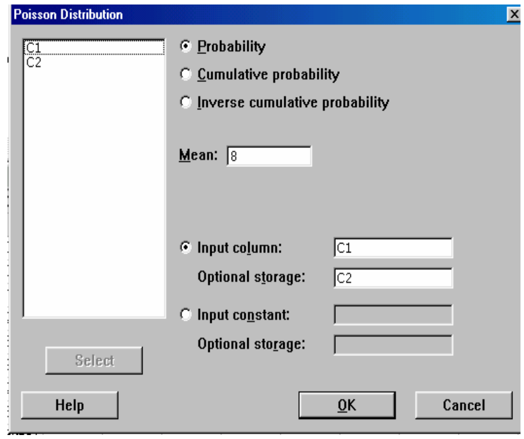
⇒ **Procedimiento MINITAB**

- Los resultados del experimento se clasifican en:
  - Obtener menos de 8 llamadas – Éxito.
  - Obtener mas de 8 llamadas – Fracaso.

- La probabilidad de recibir la llamada es constante en cada prueba.

- Las pruebas son independientes, este resultado no se ve afectado por la prueba aleatoria y no afecta al siguiente.





Interpretación

Existe una probabilidad acumulada del 45,3 % de los casos en que las llamadas serán menor a 8 por día.

2. DISTRIBUCIONES CONTINUAS.

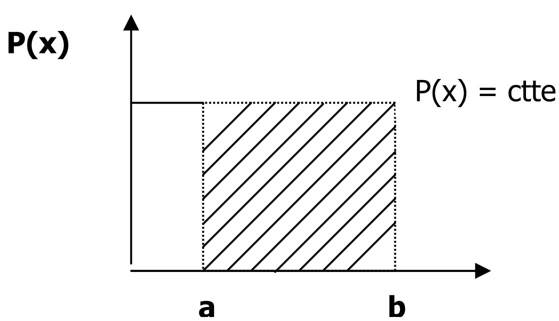
d) Distribución Uniforme

⇒ **Ejemplo**

Suponga que la concentración de cierto contaminante se encuentra distribuido en un intervalo de 4 a 20 partes por galón, si se consideran como toxico una concentración de 15 partes por galón o más. Cual es la probabilidad de que al tomar una muestra esta no sea tóxica.

⇒ **Concepto – definición**

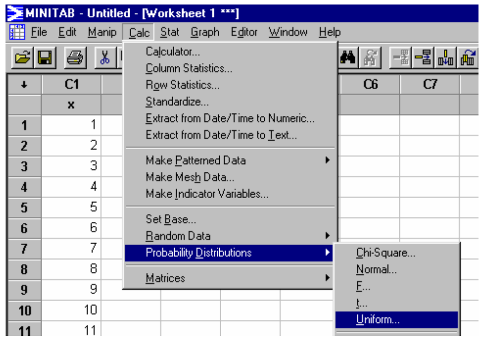
Una distribución uniforme es aquella en que todos los valores de la variable aleatoria son igualmente probables.



$$f(x) = \frac{1}{b-a}$$

⇒ PROCEDIMIENTO MINITAB.

- La probabilidad de recibir es constante en cada prueba.



Interpretación:

La probabilidad será igual a: (0.0625+0.0625+........+0.0625) ó 0.0625*11, que es igual a 68.75 % de probabilidad de no encontrar una muestra toxica.

e) Distribución Normal Estándar

⇒ **Ejercicio**

Los tiempos de reemplazo para los televisores tiene una distribución normal con una media de 8.2 años y una desviación estándar de 1.1 años. Determine la probabilidad de que un televisor seleccionado al azar tenga tiempo de reemplazo de menos de 7 años.

⇒ **Conceptos – Definición**

La densidad de probabilidad y la función de repartición para la Distribución Normal esta dada por la relación siguiente:

$$F(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\alpha}^{x} \exp\left[-\frac{1}{2}\left(\frac{t-m}{\sigma}\right)^2 .dt \right]$$

En estas expresiones, μ y σ son los dos parámetros que definen la distribución normal y representan el promedio y la desviación estándar de la distribución (X⇒ N (μ,σ^2)).

Si μ = 0 y σ = 1, la distribución se llama *Distribución Normal Reducida,* que se examinara más en detalle antes de presentar el caso en general. Considerando su importancia, la variable normal reducida, su función de densidad de probabilidad y su función de repartición son designadas por los símbolos particulares: z, (z) y ϕ(z). Se tiene en ese caso:

$$\phi(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) \qquad\qquad F(t) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right).dt$$

El análisis de la función de densidad de probabilidad muestra que la distribución es una campana simétrica respecto al eje de las ordenadas y posee dos puntos de inflexión situados en z = ± 1; la función es asintótica en el eje de las abscisas.

La función de repartición tiene una forma en S; el punto de inflexión, y de simetría, está situada en z = 0 y F(z) = 0.5; ella es asintótica al eje de las abscisas y a la recta paralela al eje de las abscisas y la ecuación ϕ(z) = 1.

A partir de la tabla que da los valores de las funciones de repartición se puede determinar las probabilidades siguientes:

P(-1< z <1) = 0.6826

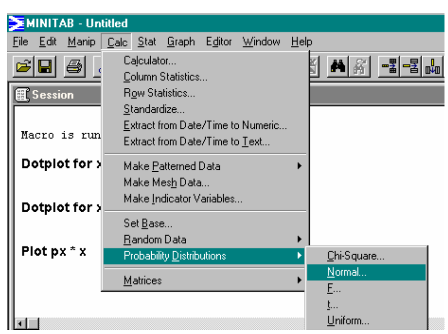
P(-1< z <1) = 0.9546

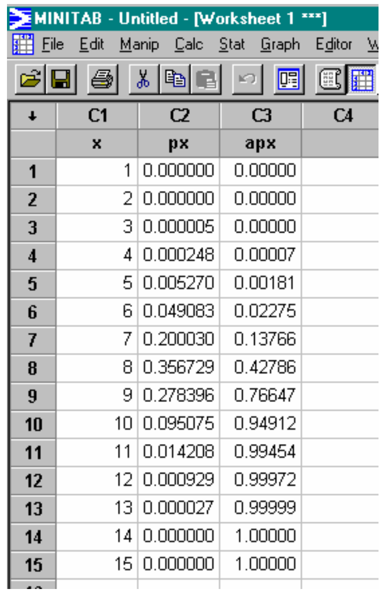
P(-1< z <1) = 0.9973

Como calcular probabilidades a partir de puntajes Z?

Distancia a lo largo de la escala horizontal de la Grafica.

⇒ Procedimiento Minitab.





↓	C1	C2	C3	C4
	x	px	apx	
1	1	0.000000	0.00000	
2	2	0.000000	0.00000	
3	3	0.000005	0.00000	
4	4	0.000248	0.00007	
5	5	0.005270	0.00181	
6	6	0.049083	0.02275	
7	7	0.200030	0.13766	
8	8	0.356729	0.42786	
9	9	0.278396	0.76647	
10	10	0.095075	0.94912	
11	11	0.014208	0.99454	
12	12	0.000929	0.99972	
13	13	0.000027	0.99999	
14	14	0.000000	1.00000	
15	15	0.000000	1.00000	

Interpretación:

a) La probabilidad de reemplazo menor a 7 años es del 13.76%.

b) La probabilidad de que sea reemplazado entre los 9 a 8 años es de 33.86% de los casos.

c) El tiempo de vida estimado para un televisor es de 10 años en el mejor de los casos.

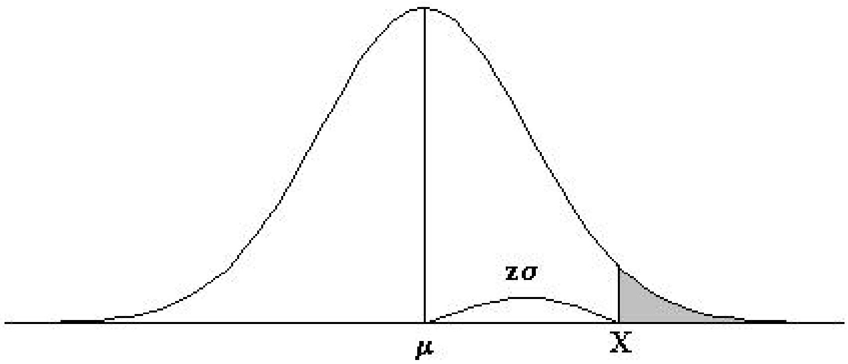
ANEXO B

TABLAS PROBABILÍSTICAS

APENDICE 1: TABLAS ESTADISTICAS

TABLA 1: DISTRIBUCIÓN NORMAL

Áreas bajo la curva normal



Ejemplo:

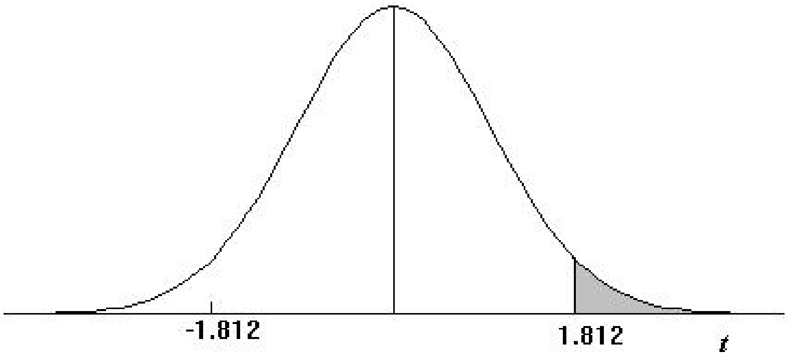
$$Z = \frac{X - \mu}{\sigma}$$

P [Z > 1] = 0.1587
P [Z > 1.96] = 0.0250

Desv. normal x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010

TABLA 2: DISTRIBUCIÓN t DE STUDENT

Puntos de porcentaje de la distribución t



Ejemplo

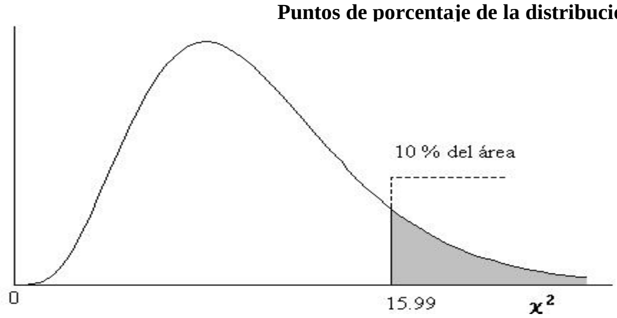
Para ϕ = 10 grados de libertad:

P[t > 1.812] = 0.05
P[t < -1.812] = 0.05

α / r	0,25	0,2	0,15	0,1	0,05	0,025	0,01	0,005	0,0005
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,656	636,578
2	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,600
3	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,768
24	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,689
28	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,660
30	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,679	0,848	1,045	1,296	1,671	2,000	2,390	2,660	3,460
120	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,290

TABLA 3: DISTRIBUCIÓN χ^2

Puntos de porcentaje de la distribución χ^2



Ejemplo:
Para ϕ = 10 grados de libertad

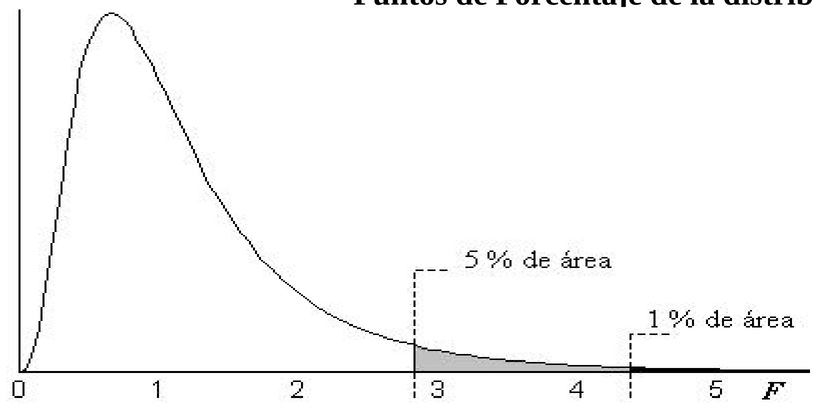
$P\ [\ \chi^2 > 15.99] = 0.10$

π / ϕ	0.995	0.99	0.975	0.95	0.9	0.75	0.5	0.25	0.1	0.05	0.025	0.01	0.005	π / ϕ
1	3.93E-05	1.57E-04	9.82E-04	3.93E-03	1.58E-02	0.102	0.455	1.323	2.71	3.84	5.02	6.63	7.88	1
2	1.00E-02	2.01E-02	5.06E-02	0.103	0.211	0.575	1.386	2.77	4.61	5.99	7.38	9.21	10.60	2
3	7.17E-02	0.115	0.216	0.352	0.584	1.213	2.37	4.11	6.25	7.81	9.35	11.34	12.84	3
4	0.207	0.297	0.484	0.711	1.064	1.923	3.36	5.39	7.78	9.49	11.14	13.28	14.86	4
5	0.412	0.554	0.831	1.145	1.610	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75	5
6	0.676	0.872	1.237	1.635	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55	6
7	0.989	1.239	1.690	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.3	7
8	1.344	1.647	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.1	22.0	8
9	1.735	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.7	23.6	9
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.5	23.2	25.2	10
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.9	24.7	26.8	11
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.0	23.3	26.2	28.3	12
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.4	24.7	27.7	29.8	13
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.1	23.7	26.1	29.1	31.3	14
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.3	25.0	27.5	30.6	32.8	15
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.5	26.3	28.8	32.0	34.3	16
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.5	24.8	27.6	30.2	33.4	35.7	17
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.6	26.0	28.9	31.5	34.8	37.2	18
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.7	27.2	30.1	32.9	36.2	38.6	19
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.8	28.4	31.4	34.2	37.6	40.0	20
21	8.03	8.90	10.28	11.59	13.24	16.34	20.3	24.9	29.6	32.7	35.5	38.9	41.4	21
22	8.64	9.54	10.98	12.34	14.04	17.24	21.3	26.0	30.8	33.9	36.8	40.3	42.8	22
23	9.26	10.20	11.69	13.09	14.85	18.14	22.3	27.1	32.0	35.2	38.1	41.6	44.2	23
24	9.89	10.86	12.40	13.85	15.66	19.04	23.3	28.2	33.2	36.4	39.4	43.0	45.6	24
25	10.52	11.52	13.12	14.61	16.47	19.94	24.3	29.3	34.4	37.7	40.6	44.3	46.9	25
26	11.16	12.20	13.84	15.38	17.29	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3	26
27	11.81	12.88	14.57	16.15	18.11	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6	27
28	12.46	13.56	15.31	16.93	18.94	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0	28
29	13.12	14.26	16.05	17.71	19.77	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3	29
30	13.79	14.95	16.79	18.49	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7	30
40	20.7	22.2	24.4	26.5	29.1	33.7	39.3	45.6	51.8	55.8	59.3	63.7	66.8	40
50	28.0	29.7	32.4	34.8	37.7	42.9	49.3	56.3	63.2	67.5	71.4	76.2	79.5	50
60	35.5	37.5	40.5	43.2	46.5	52.3	59.3	67.0	74.4	79.1	83.3	88.4	92.0	60
70	43.3	45.4	48.8	51.7	55.3	61.7	69.3	77.6	85.5	90.5	95.0	100.4	104.2	70
80	51.2	53.5	57.2	60.4	64.3	71.1	79.3	88.1	96.6	101.9	106.6	112.3	116.3	80
90	59.2	61.8	65.6	69.1	73.3	80.6	89.3	98.6	107.6	113.1	118.1	124.1	128.3	90
100	67.3	70.1	74.2	77.9	82.4	90.1	99.3	109.1	118.5	124.3	129.6	135.8	140.2	100
Z_α	-2.58	-2.33	-1.96	-1.64	-1.28	-0.674	0.000	0.674	1.282	1.645	1.96	2.33	2.58	Z_α

Para $\phi > 100$ tómese $\chi^2 = \frac{1}{2}\left(Z_\alpha + \sqrt{2\phi - 1}\right)^2$. Z_α es la desviación normal estandarizada correspondiente al nivel de significancia y se muestra en la parte superior de la tabla.

TABLA 4: DISTRIBUCIÓN F DE FISHER

Puntos de Porcentaje de la distribución F



Ejemplo:

Para n1 = 9, n2 = 12 grados de libertad:
P[F > 2.80] = 0.05
P [F > 4.39] = 0.01

n_2	5 % (normal) y 1 % (negritas) puntos para la distribución de F																								n_2
	n1 grados delibertad (para el mayor cuadrado medio)																								
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
1	161	199	216	225	230	234	237	239	241	242	243	244	245	246	248	249	250	251	252	253	253	254	254	254	1
	4052	**4999**	**5404**	**5624**	**5764**	**5859**	**5928**	**5981**	**6022**	**6056**	**6083**	**6107**	**6143**	**6170**	**6209**	**6234**	**6260**	**6286**	**6302**	**6324**	**6334**	**6350**	**6360**	**6366**	
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.40	19.41	19.42	19.43	19.45	19.45	19.46	19.47	19.48	19.48	19.49	19.49	19.49	19.50	2
	98.50	**99.00**	**99.16**	**99.25**	**99.30**	**99.33**	**99.36**	**99.38**	**99.39**	**99.40**	**99.41**	**99.42**	**99.43**	**99.44**	**99.45**	**99.46**	**99.47**	**99.48**	**99.48**	**99.48**	**99.49**	**99.49**	**99.50**	**99.50**	
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.71	8.69	8.66	8.64	8.62	8.59	8.58	8.56	8.55	8.54	8.53	8.53	3
	34.12	**30.82**	**29.46**	**28.71**	**28.24**	**27.91**	**27.67**	**27.49**	**27.34**	**27.23**	**27.13**	**27.05**	**26.92**	**26.83**	**26.69**	**26.60**	**26.50**	**26.41**	**26.35**	**26.28**	**26.24**	**26.18**	**26.15**	**26.13**	
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.87	5.84	5.80	5.77	5.75	5.72	5.70	5.68	5.66	5.65	5.64	5.63	4
	21.20	**18.00**	**16.69**	**15.98**	**15.52**	**15.21**	**14.98**	**14.80**	**14.66**	**14.55**	**14.45**	**14.37**	**14.25**	**14.15**	**14.02**	**13.93**	**13.84**	**13.75**	**13.69**	**13.61**	**13.58**	**13.52**	**13.49**	**13.46**	
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.64	4.60	4.56	4.53	4.50	4.46	4.44	4.42	4.41	4.39	4.37	4.37	5
	16.26	**13.27**	**12.06**	**11.39**	**10.97**	**10.67**	**10.46**	**10.29**	**10.16**	**10.05**	**9.96**	**9.89**	**9.77**	**9.68**	**9.55**	**9.47**	**9.38**	**9.29**	**9.24**	**9.17**	**9.13**	**9.08**	**9.04**	**9.02**	
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.96	3.92	3.87	3.84	3.81	3.77	3.75	3.73	3.71	3.69	3.68	3.67	6
	13.75	**10.92**	**9.78**	**9.15**	**8.75**	**8.47**	**8.26**	**8.10**	**7.98**	**7.87**	**7.79**	**7.72**	**7.60**	**7.52**	**7.40**	**7.31**	**7.23**	**7.14**	**7.09**	**7.02**	**6.99**	**6.93**	**6.90**	**6.88**	
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.53	3.49	3.44	3.41	3.38	3.34	3.32	3.29	3.27	3.25	3.24	3.23	7
	12.25	**9.55**	**8.45**	**7.85**	**7.46**	**7.19**	**6.99**	**6.84**	**6.72**	**6.62**	**6.54**	**6.47**	**6.36**	**6.28**	**6.16**	**6.07**	**5.99**	**5.91**	**5.86**	**5.79**	**5.75**	**5.70**	**5.67**	**5.65**	
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.24	3.20	3.15	3.12	3.08	3.04	3.02	2.99	2.97	2.95	2.94	2.93	8
	11.26	**8.65**	**7.59**	**7.01**	**6.63**	**6.37**	**6.18**	**6.03**	**5.91**	**5.81**	**5.73**	**5.67**	**5.56**	**5.48**	**5.36**	**5.28**	**5.20**	**5.12**	**5.07**	**5.00**	**4.96**	**4.91**	**4.88**	**4.86**	
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.03	2.99	2.94	2.90	2.86	2.83	2.80	2.77	2.76	2.73	2.72	2.71	9
	10.56	**8.02**	**6.99**	**6.42**	**6.06**	**5.80**	**5.61**	**5.47**	**5.35**	**5.26**	**5.18**	**5.11**	**5.01**	**4.92**	**4.81**	**4.73**	**4.65**	**4.57**	**4.52**	**4.45**	**4.41**	**4.36**	**4.33**	**4.31**	
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.86	2.83	2.77	2.74	2.70	2.66	2.64	2.60	2.59	2.56	2.55	2.54	10
	10.04	**7.56**	**6.55**	**5.99**	**5.64**	**5.39**	**5.20**	**5.06**	**4.94**	**4.85**	**4.77**	**4.71**	**4.60**	**4.52**	**4.41**	**4.33**	**4.25**	**4.17**	**4.12**	**4.05**	**4.01**	**3.96**	**3.93**	**3.91**	

n_2	5 % (normal) y 1 % (negritas) puntos para la distribución de F																								n_2
	n1 grados delibertad (para el mayor cuadrado medio)																								
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.74	2.70	2.65	2.61	2.57	2.53	2.51	2.47	2.46	2.43	2.42	2.40	11
	9.65	**7.21**	**6.22**	**5.67**	**5.32**	**5.07**	**4.89**	**4.74**	**4.63**	**4.54**	**4.46**	**4.40**	**4.29**	**4.21**	**4.10**	**4.02**	**3.94**	**3.86**	**3.81**	**3.74**	**3.71**	**3.66**	**3.62**	**3.60**	
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.64	2.60	2.54	2.51	2.47	2.43	2.40	2.37	2.35	2.32	2.31	2.30	12
	9.33	**6.93**	**5.95**	**5.41**	**5.06**	**4.82**	**4.64**	**4.50**	**4.39**	**4.30**	**4.22**	**4.16**	**4.05**	**3.97**	**3.86**	**3.78**	**3.70**	**3.62**	**3.57**	**3.50**	**3.47**	**3.41**	**3.38**	**3.36**	
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.55	2.51	2.46	2.42	2.38	2.34	2.31	2.28	2.26	2.23	2.22	2.21	13
	9.07	**6.70**	**5.74**	**5.21**	**4.86**	**4.62**	**4.44**	**4.30**	**4.19**	**4.10**	**4.02**	**3.96**	**3.86**	**3.78**	**3.66**	**3.59**	**3.51**	**3.43**	**3.38**	**3.31**	**3.27**	**3.22**	**3.19**	**3.17**	
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.48	2.44	2.39	2.35	2.31	2.27	2.24	2.21	2.19	2.16	2.14	2.13	14
	8.86	**6.51**	**5.56**	**5.04**	**4.69**	**4.46**	**4.28**	**4.14**	**4.03**	**3.94**	**3.86**	**3.80**	**3.70**	**3.62**	**3.51**	**3.43**	**3.35**	**3.27**	**3.22**	**3.15**	**3.11**	**3.06**	**3.03**	**3.00**	
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.42	2.38	2.33	2.29	2.25	2.20	2.18	2.14	2.12	2.10	2.08	2.07	15
	8.68	**6.36**	**5.42**	**4.89**	**4.56**	**4.32**	**4.14**	**4.00**	**3.89**	**3.80**	**3.73**	**3.67**	**3.56**	**3.49**	**3.37**	**3.29**	**3.21**	**3.13**	**3.08**	**3.01**	**2.98**	**2.92**	**2.89**	**2.87**	
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.37	2.33	2.28	2.24	2.19	2.15	2.12	2.09	2.07	2.04	2.02	2.01	16
	8.53	**6.23**	**5.29**	**4.77**	**4.44**	**4.20**	**4.03**	**3.89**	**3.78**	**3.69**	**3.62**	**3.55**	**3.45**	**3.37**	**3.26**	**3.18**	**3.10**	**3.02**	**2.97**	**2.90**	**2.86**	**2.81**	**2.78**	**2.75**	
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.33	2.29	2.23	2.19	2.15	2.10	2.08	2.04	2.02	1.99	1.97	1.96	17
	8.40	**6.11**	**5.19**	**4.67**	**4.34**	**4.10**	**3.93**	**3.79**	**3.68**	**3.59**	**3.52**	**3.46**	**3.35**	**3.27**	**3.16**	**3.08**	**3.00**	**2.92**	**2.87**	**2.80**	**2.76**	**2.71**	**2.68**	**2.65**	
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.29	2.25	2.19	2.15	2.11	2.06	2.04	2.00	1.98	1.95	1.93	1.92	18
	8.29	**6.01**	**5.09**	**4.58**	**4.25**	**4.01**	**3.84**	**3.71**	**3.60**	**3.51**	**3.43**	**3.37**	**3.27**	**3.19**	**3.08**	**3.00**	**2.92**	**2.84**	**2.78**	**2.71**	**2.68**	**2.62**	**2.59**	**2.57**	
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31	2.26	2.21	2.16	2.11	2.07	2.03	2.00	1.96	1.94	1.91	1.89	1.88	19
	8.18	**5.93**	**5.01**	**4.50**	**4.17**	**3.94**	**3.77**	**3.63**	**3.52**	**3.43**	**3.36**	**3.30**	**3.19**	**3.12**	**3.00**	**2.92**	**2.84**	**2.76**	**2.71**	**2.64**	**2.60**	**2.55**	**2.51**	**2.49**	
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.22	2.18	2.12	2.08	2.04	1.99	1.97	1.93	1.91	1.88	1.86	1.84	20
	8.10	**5.85**	**4.94**	**4.43**	**4.10**	**3.87**	**3.70**	**3.56**	**3.46**	**3.37**	**3.29**	**3.23**	**3.13**	**3.05**	**2.94**	**2.86**	**2.78**	**2.69**	**2.64**	**2.57**	**2.54**	**2.48**	**2.44**	**2.42**	
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.20	2.16	2.10	2.05	2.01	1.96	1.94	1.90	1.88	1.84	1.83	1.81	21
	8.02	**5.78**	**4.87**	**4.37**	**4.04**	**3.81**	**3.64**	**3.51**	**3.40**	**3.31**	**3.24**	**3.17**	**3.07**	**2.99**	**2.88**	**2.80**	**2.72**	**2.64**	**2.58**	**2.51**	**2.48**	**2.42**	**2.38**	**2.36**	
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23	2.17	2.13	2.07	2.03	1.98	1.94	1.91	1.87	1.85	1.82	1.80	1.78	22
	7.95	**5.72**	**4.82**	**4.31**	**3.99**	**3.76**	**3.59**	**3.45**	**3.35**	**3.26**	**3.18**	**3.12**	**3.02**	**2.94**	**2.83**	**2.75**	**2.67**	**2.58**	**2.53**	**2.46**	**2.42**	**2.36**	**2.33**	**2.31**	
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.24	2.20	2.15	2.11	2.05	2.01	1.96	1.91	1.88	1.84	1.82	1.79	1.77	1.76	23
	7.88	**5.66**	**4.76**	**4.26**	**3.94**	**3.71**	**3.54**	**3.41**	**3.30**	**3.21**	**3.14**	**3.07**	**2.97**	**2.89**	**2.78**	**2.70**	**2.62**	**2.54**	**2.48**	**2.41**	**2.37**	**2.32**	**2.28**	**2.26**	
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.18	2.13	2.09	2.03	1.98	1.94	1.89	1.86	1.82	1.80	1.77	1.75	1.73	24
	7.82	**5.61**	**4.72**	**4.22**	**3.90**	**3.67**	**3.50**	**3.36**	**3.26**	**3.17**	**3.09**	**3.03**	**2.93**	**2.85**	**2.74**	**2.66**	**2.58**	**2.49**	**2.44**	**2.37**	**2.33**	**2.27**	**2.24**	**2.21**	
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16	2.11	2.07	2.01	1.96	1.92	1.87	1.84	1.80	1.78	1.75	1.73	1.71	25
	7.77	**5.57**	**4.68**	**4.18**	**3.85**	**3.63**	**3.46**	**3.32**	**3.22**	**3.13**	**3.06**	**2.99**	**2.89**	**2.81**	**2.70**	**2.62**	**2.54**	**2.45**	**2.40**	**2.33**	**2.29**	**2.23**	**2.19**	**2.17**	
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.09	2.05	1.99	1.95	1.90	1.85	1.82	1.78	1.76	1.73	1.71	1.69	26
	7.72	**5.53**	**4.64**	**4.14**	**3.82**	**3.59**	**3.42**	**3.29**	**3.18**	**3.09**	**3.02**	**2.96**	**2.86**	**2.78**	**2.66**	**2.58**	**2.50**	**2.42**	**2.36**	**2.29**	**2.25**	**2.19**	**2.16**	**2.13**	
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.17	2.13	2.08	2.04	1.97	1.93	1.88	1.84	1.81	1.76	1.74	1.71	1.69	1.67	27
	7.68	**5.49**	**4.60**	**4.11**	**3.78**	**3.56**	**3.39**	**3.26**	**3.15**	**3.06**	**2.99**	**2.93**	**2.82**	**2.75**	**2.63**	**2.55**	**2.47**	**2.38**	**2.33**	**2.26**	**2.22**	**2.16**	**2.12**	**2.10**	
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12	2.06	2.02	1.96	1.91	1.87	1.82	1.79	1.75	1.73	1.69	1.67	1.65	28
	7.64	**5.45**	**4.57**	**4.07**	**3.75**	**3.53**	**3.36**	**3.23**	**3.12**	**3.03**	**2.96**	**2.90**	**2.79**	**2.72**	**2.60**	**2.52**	**2.44**	**2.35**	**2.30**	**2.23**	**2.19**	**2.13**	**2.09**	**2.06**	
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.05	2.01	1.94	1.90	1.85	1.81	1.77	1.73	1.71	1.67	1.65	1.64	29
	7.60	**5.42**	**4.54**	**4.04**	**3.73**	**3.50**	**3.33**	**3.20**	**3.09**	**3.00**	**2.93**	**2.87**	**2.77**	**2.69**	**2.57**	**2.49**	**2.41**	**2.33**	**2.27**	**2.20**	**2.16**	**2.10**	**2.06**	**2.03**	
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.04	1.99	1.93	1.89	1.84	1.79	1.76	1.72	1.70	1.66	1.64	1.62	30
	7.56	**5.39**	**4.51**	**4.02**	**3.70**	**3.47**	**3.30**	**3.17**	**3.07**	**2.98**	**2.91**	**2.84**	**2.74**	**2.66**	**2.55**	**2.47**	**2.39**	**2.30**	**2.25**	**2.17**	**2.13**	**2.07**	**2.03**	**2.01**	
32	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14	2.10	2.07	2.01	1.97	1.91	1.86	1.82	1.77	1.74	1.69	1.67	1.63	1.61	1.59	32
	7.50	**5.34**	**4.46**	**3.97**	**3.65**	**3.43**	**3.26**	**3.13**	**3.02**	**2.93**	**2.86**	**2.80**	**2.70**	**2.62**	**2.50**	**2.42**	**2.34**	**2.25**	**2.20**	**2.12**	**2.08**	**2.02**	**1.98**	**1.96**	

n_2	5 % (normal) y 1 % (negritas) puntos para la distribución de F																								n_2
	n1 grados delibertad (para el mayor cuadrado medio)																								
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
34	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17	2.12	2.08	2.05	1.99	1.95	1.89	1.84	1.80	1.75	1.71	1.67	1.65	1.61	1.59	1.57	34
	7.44	**5.29**	**4.42**	**3.93**	**3.61**	**3.39**	**3.22**	**3.09**	**2.98**	**2.89**	**2.82**	**2.76**	**2.66**	**2.58**	**2.46**	**2.38**	**2.30**	**2.21**	**2.16**	**2.08**	**2.04**	**1.98**	**1.94**	**1.91**	
36	4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.15	2.11	2.07	2.03	1.98	1.93	1.87	1.82	1.78	1.73	1.69	1.65	1.62	1.59	1.56	1.55	36
	7.40	**5.25**	**4.38**	**3.89**	**3.57**	**3.35**	**3.18**	**3.05**	**2.95**	**2.86**	**2.79**	**2.72**	**2.62**	**2.54**	**2.43**	**2.35**	**2.26**	**2.18**	**2.12**	**2.04**	**2.00**	**1.94**	**1.90**	**1.87**	
38	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02	1.96	1.92	1.85	1.81	1.76	1.71	1.68	1.63	1.61	1.57	1.54	1.53	38
	7.35	**5.21**	**4.34**	**3.86**	**3.54**	**3.32**	**3.15**	**3.02**	**2.92**	**2.83**	**2.75**	**2.69**	**2.59**	**2.51**	**2.40**	**2.32**	**2.23**	**2.14**	**2.09**	**2.01**	**1.97**	**1.90**	**1.86**	**1.84**	
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.95	1.90	1.84	1.79	1.74	1.69	1.66	1.61	1.59	1.55	1.53	1.51	40
	7.31	**5.18**	**4.31**	**3.83**	**3.51**	**3.29**	**3.12**	**2.99**	**2.89**	**2.80**	**2.73**	**2.66**	**2.56**	**2.48**	**2.37**	**2.29**	**2.20**	**2.11**	**2.06**	**1.98**	**1.94**	**1.87**	**1.83**	**1.81**	
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.03	1.99	1.94	1.89	1.83	1.78	1.73	1.68	1.65	1.60	1.57	1.53	1.51	1.49	42
	7.28	**5.15**	**4.29**	**3.80**	**3.49**	**3.27**	**3.10**	**2.97**	**2.86**	**2.78**	**2.70**	**2.64**	**2.54**	**2.46**	**2.34**	**2.26**	**2.18**	**2.09**	**2.03**	**1.95**	**1.91**	**1.85**	**1.80**	**1.78**	
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98	1.92	1.88	1.81	1.77	1.72	1.67	1.63	1.59	1.56	1.52	1.49	1.48	44
	7.25	**5.12**	**4.26**	**3.78**	**3.47**	**3.24**	**3.08**	**2.95**	**2.84**	**2.75**	**2.68**	**2.62**	**2.52**	**2.44**	**2.32**	**2.24**	**2.15**	**2.07**	**2.01**	**1.93**	**1.89**	**1.82**	**1.78**	**1.75**	
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.15	2.09	2.04	2.00	1.97	1.91	1.87	1.80	1.76	1.71	1.65	1.62	1.57	1.55	1.51	1.48	1.46	46
	7.22	**5.10**	**4.24**	**3.76**	**3.44**	**3.22**	**3.06**	**2.93**	**2.82**	**2.73**	**2.66**	**2.60**	**2.50**	**2.42**	**2.30**	**2.22**	**2.13**	**2.04**	**1.99**	**1.91**	**1.86**	**1.80**	**1.76**	**1.73**	
48	4.04	3.19	2.80	2.57	2.41	2.29	2.21	2.14	2.08	2.03	1.99	1.96	1.90	1.86	1.79	1.75	1.70	1.64	1.61	1.56	1.54	1.49	1.47	1.45	48
	7.19	**5.08**	**4.22**	**3.74**	**3.43**	**3.20**	**3.04**	**2.91**	**2.80**	**2.71**	**2.64**	**2.58**	**2.48**	**2.40**	**2.28**	**2.20**	**2.12**	**2.02**	**1.97**	**1.89**	**1.84**	**1.78**	**1.73**	**1.70**	
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.99	1.95	1.89	1.85	1.78	1.74	1.69	1.63	1.60	1.55	1.52	1.48	1.46	1.44	50
	7.17	**5.06**	**4.20**	**3.72**	**3.41**	**3.19**	**3.02**	**2.89**	**2.78**	**2.70**	**2.63**	**2.56**	**2.46**	**2.38**	**2.27**	**2.18**	**2.10**	**2.01**	**1.95**	**1.87**	**1.82**	**1.76**	**1.71**	**1.68**	
55	4.02	3.16	2.77	2.54	2.38	2.27	2.18	2.11	2.06	2.01	1.97	1.93	1.88	1.83	1.76	1.72	1.67	1.61	1.58	1.53	1.50	1.46	1.43	1.41	55
	7.12	**5.01**	**4.16**	**3.68**	**3.37**	**3.15**	**2.98**	**2.85**	**2.75**	**2.66**	**2.59**	**2.53**	**2.42**	**2.34**	**2.23**	**2.15**	**2.06**	**1.97**	**1.91**	**1.83**	**1.78**	**1.71**	**1.67**	**1.64**	
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.86	1.82	1.75	1.70	1.65	1.59	1.56	1.51	1.48	1.44	1.41	1.39	60
	7.08	**4.98**	**4.13**	**3.65**	**3.34**	**3.12**	**2.95**	**2.82**	**2.72**	**2.63**	**2.56**	**2.50**	**2.39**	**2.31**	**2.20**	**2.12**	**2.03**	**1.94**	**1.88**	**1.79**	**1.75**	**1.68**	**1.63**	**1.60**	
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.03	1.98	1.94	1.90	1.85	1.80	1.73	1.69	1.63	1.58	1.54	1.49	1.46	1.42	1.39	1.37	65
	7.04	**4.95**	**4.10**	**3.62**	**3.31**	**3.09**	**2.93**	**2.80**	**2.69**	**2.61**	**2.53**	**2.47**	**2.37**	**2.29**	**2.17**	**2.09**	**2.00**	**1.91**	**1.85**	**1.77**	**1.72**	**1.65**	**1.60**	**1.57**	
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97	1.93	1.89	1.84	1.79	1.72	1.67	1.62	1.57	1.53	1.48	1.45	1.40	1.37	1.35	70
	7.01	**4.92**	**4.07**	**3.60**	**3.29**	**3.07**	**2.91**	**2.78**	**2.67**	**2.59**	**2.51**	**2.45**	**2.35**	**2.27**	**2.15**	**2.07**	**1.98**	**1.89**	**1.83**	**1.74**	**1.70**	**1.62**	**1.57**	**1.54**	
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.91	1.88	1.82	1.77	1.70	1.65	1.60	1.54	1.51	1.45	1.43	1.38	1.35	1.33	80
	6.96	**4.88**	**4.04**	**3.56**	**3.26**	**3.04**	**2.87**	**2.74**	**2.64**	**2.55**	**2.48**	**2.42**	**2.31**	**2.23**	**2.12**	**2.03**	**1.94**	**1.85**	**1.79**	**1.70**	**1.65**	**1.58**	**1.53**	**1.50**	
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.89	1.85	1.79	1.75	1.68	1.63	1.57	1.52	1.48	1.42	1.39	1.34	1.31	1.28	100
	6.90	**4.82**	**3.98**	**3.51**	**3.21**	**2.99**	**2.82**	**2.69**	**2.59**	**2.50**	**2.43**	**2.37**	**2.27**	**2.19**	**2.07**	**1.98**	**1.89**	**1.80**	**1.74**	**1.65**	**1.60**	**1.52**	**1.47**	**1.43**	
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.96	1.91	1.87	1.83	1.77	1.73	1.66	1.60	1.55	1.49	1.45	1.40	1.36	1.31	1.27	1.25	125
	6.84	**4.78**	**3.94**	**3.47**	**3.17**	**2.95**	**2.79**	**2.66**	**2.55**	**2.47**	**2.39**	**2.33**	**2.23**	**2.15**	**2.03**	**1.94**	**1.85**	**1.76**	**1.69**	**1.60**	**1.55**	**1.47**	**1.41**	**1.37**	
150	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.82	1.76	1.71	1.64	1.59	1.54	1.48	1.44	1.38	1.34	1.29	1.25	1.22	150
	6.81	**4.75**	**3.91**	**3.45**	**3.14**	**2.92**	**2.76**	**2.63**	**2.53**	**2.44**	**2.37**	**2.31**	**2.20**	**2.12**	**2.00**	**1.92**	**1.83**	**1.73**	**1.66**	**1.57**	**1.52**	**1.43**	**1.38**	**1.33**	
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.84	1.80	1.74	1.69	1.62	1.57	1.52	1.46	1.41	1.35	1.32	1.26	1.22	1.19	200
	6.76	**4.71**	**3.88**	**3.41**	**3.11**	**2.89**	**2.73**	**2.60**	**2.50**	**2.41**	**2.34**	**2.27**	**2.17**	**2.09**	**1.97**	**1.89**	**1.79**	**1.69**	**1.63**	**1.53**	**1.48**	**1.39**	**1.33**	**1.28**	
400	3.86	3.02	2.63	2.39	2.24	2.12	2.03	1.96	1.90	1.85	1.81	1.78	1.72	1.67	1.60	1.54	1.49	1.42	1.38	1.32	1.28	1.22	1.17	1.13	400
	6.70	**4.66**	**3.83**	**3.37**	**3.06**	**2.85**	**2.68**	**2.56**	**2.45**	**2.37**	**2.29**	**2.23**	**2.13**	**2.05**	**1.92**	**1.84**	**1.75**	**1.64**	**1.58**	**1.48**	**1.42**	**1.32**	**1.25**	**1.19**	
1000	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.80	1.76	1.70	1.65	1.58	1.53	1.47	1.41	1.36	1.30	1.26	1.19	1.13	1.08	1000
	6.66	**4.63**	**3.80**	**3.34**	**3.04**	**2.82**	**2.66**	**2.53**	**2.43**	**2.34**	**2.27**	**2.20**	**2.10**	**2.02**	**1.90**	**1.81**	**1.72**	**1.61**	**1.54**	**1.44**	**1.38**	**1.28**	**1.19**	**1.12**	
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75	1.69	1.64	1.57	1.52	1.46	1.39	1.35	1.28	1.24	1.17	1.11	1.00	∞
	6.63	**4.61**	**3.78**	**3.32**	**3.02**	**2.80**	**2.64**	**2.51**	**2.41**	**2.32**	**2.25**	**2.18**	**2.08**	**2.00**	**1.88**	**1.79**	**1.70**	**1.59**	**1.52**	**1.42**	**1.36**	**1.25**	**1.15**	**1.00**	

TABLA 5: PROBABILIDADES BINOMIALES

						p					
n	*k*	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
1	0	0.9500	0.9000	0.8500	0.8000	0.7500	0.7000	0.6500	0.6000	0.5500	0.5000
1	1	0.0500	0.1000	0.1500	0.2000	0.2500	0.3000	0.3500	0.4000	0.4500	0.5000
2	0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
2	1	0.0950	0.1800	0.2550	0.3200	0.3750	0.4200	0.4550	0.4800	0.4950	0.5000
2	2	0.0025	0.0100	0.0225	0.0400	0.0625	0.0900	0.1225	0.1600	0.2025	0.2500
3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
3	1	0.1354	0.2430	0.3251	0.3840	0.4219	0.4410	0.4436	0.4320	0.4084	0.3750
3	2	0.0071	0.0270	0.0574	0.0960	0.1406	0.1890	0.2389	0.2880	0.3341	0.3750
3	3	0.0001	0.0010	0.0034	0.0080	0.0156	0.0270	0.0429	0.0640	0.0911	0.1250
4	0	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
4	1	0.1715	0.2916	0.3685	0.4096	0.4219	0.4116	0.3845	0.3456	0.2995	0.2500
4	2	0.0135	0.0486	0.0975	0.1536	0.2109	0.2646	0.3105	0.3456	0.3675	0.3750
4	3	0.0005	0.0036	0.0115	0.0256	0.0469	0.0756	0.1115	0.1536	0.2005	0.2500
4	4	0.0000	0.0001	0.0005	0.0016	0.0039	0.0081	0.0150	0.0256	0.0410	0.0625
5	0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0313
5	1	0.2036	0.3281	0.3915	0.4096	0.3955	0.3602	0.3124	0.2592	0.2059	0.1563
5	2	0.0214	0.0729	0.1382	0.2048	0.2637	0.3087	0.3364	0.3456	0.3369	0.3125
5	3	0.0011	0.0081	0.0244	0.0512	0.0879	0.1323	0.1811	0.2304	0.2757	0.3125
5	4	0.0000	0.0005	0.0022	0.0064	0.0146	0.0284	0.0488	0.0768	0.1128	0.1563
5	5	0.0000	0.0000	0.0001	0.0003	0.0010	0.0024	0.0053	0.0102	0.0185	0.0313
6	0	0.7351	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0467	0.0277	0.0156
6	1	0.2321	0.3543	0.3993	0.3932	0.3560	0.3025	0.2437	0.1866	0.1359	0.0938
6	2	0.0305	0.0984	0.1762	0.2458	0.2966	0.3241	0.3280	0.3110	0.2780	0.2344
6	3	0.0021	0.0146	0.0415	0.0819	0.1318	0.1852	0.2355	0.2765	0.3032	0.3125
6	4	0.0001	0.0012	0.0055	0.0154	0.0330	0.0595	0.0951	0.1382	0.1861	0.2344
6	5	0.0000	0.0001	0.0004	0.0015	0.0044	0.0102	0.0205	0.0369	0.0609	0.0938
6	6	0.0000	0.0000	0.0000	0.0001	0.0002	0.0007	0.0018	0.0041	0.0083	0.0156
7	0	0.6983	0.4783	0.3206	0.2097	0.1335	0.0824	0.0490	0.0280	0.0152	0.0078
7	1	0.2573	0.3720	0.3960	0.3670	0.3115	0.2471	0.1848	0.1306	0.0872	0.0547
7	2	0.0406	0.1240	0.2097	0.2753	0.3115	0.3177	0.2985	0.2613	0.2140	0.1641
7	3	0.0036	0.0230	0.0617	0.1147	0.1730	0.2269	0.2679	0.2903	0.2918	0.2734
7	4	0.0002	0.0026	0.0109	0.0287	0.0577	0.0972	0.1442	0.1935	0.2388	0.2734
7	5	0.0000	0.0002	0.0012	0.0043	0.0115	0.0250	0.0466	0.0774	0.1172	0.1641
7	6	0.0000	0.0000	0.0001	0.0004	0.0013	0.0036	0.0084	0.0172	0.0320	0.0547
7	7	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0006	0.0016	0.0037	0.0078
8	0	0.6634	0.4305	0.2725	0.1678	0.1001	0.0576	0.0319	0.0168	0.0084	0.0039
8	1	0.2793	0.3826	0.3847	0.3355	0.2670	0.1977	0.1373	0.0896	0.0548	0.0313
8	2	0.0515	0.1488	0.2376	0.2936	0.3115	0.2965	0.2587	0.2090	0.1569	0.1094
8	3	0.0054	0.0331	0.0839	0.1468	0.2076	0.2541	0.2786	0.2787	0.2568	0.2188
8	4	0.0004	0.0046	0.0185	0.0459	0.0865	0.1361	0.1875	0.2322	0.2627	0.2734
8	5	0.0000	0.0004	0.0026	0.0092	0.0231	0.0467	0.0808	0.1239	0.1719	0.2188
8	6	0.0000	0.0000	0.0002	0.0011	0.0038	0.0100	0.0217	0.0413	0.0703	0.1094
8	7	0.0000	0.0000	0.0000	0.0001	0.0004	0.0012	0.0033	0.0079	0.0164	0.0313
8	8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0007	0.0017	0.0039

TABLA 5 (CONTINUACIÓN)

						p					
n	*k*	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
9	0	0.6302	0.3874	0.2316	0.1342	0.0751	0.0404	0.0207	0.0101	0.0046	0.0020
9	1	0.2985	0.3874	0.3679	0.3020	0.2253	0.1556	0.1004	0.0605	0.0339	0.0176
9	2	0.0629	0.1722	0.2597	0.3020	0.3003	0.2668	0.2162	0.1612	0.1110	0.0703
9	3	0.0077	0.0446	0.1069	0.1762	0.2336	0.2668	0.2716	0.2508	0.2119	0.1641
9	4	0.0006	0.0074	0.0283	0.0661	0.1168	0.1715	0.2194	0.2508	0.2600	0.2461
9	5	0.0000	0.0008	0.0050	0.0165	0.0389	0.0735	0.1181	0.1672	0.2128	0.2461
9	6	0.0000	0.0001	0.0006	0.0028	0.0087	0.0210	0.0424	0.0743	0.1160	0.1641
9	7	0.0000	0.0000	0.0000	0.0003	0.0012	0.0039	0.0098	0.0212	0.0407	0.0703
9	8	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0013	0.0035	0.0083	0.0176
9	9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0008	0.0020
10	0	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010
10	1	0.3151	0.3874	0.3474	0.2684	0.1877	0.1211	0.0725	0.0403	0.0207	0.0098
10	2	0.0746	0.1937	0.2759	0.3020	0.2816	0.2335	0.1757	0.1209	0.0763	0.0439
10	3	0.0105	0.0574	0.1298	0.2013	0.2503	0.2668	0.2522	0.2150	0.1665	0.1172
10	4	0.0010	0.0112	0.0401	0.0881	0.1460	0.2001	0.2377	0.2508	0.2384	0.2051
10	5	0.0001	0.0015	0.0085	0.0264	0.0584	0.1029	0.1536	0.2007	0.2340	0.2461
10	6	0.0000	0.0001	0.0012	0.0055	0.0162	0.0368	0.0689	0.1115	0.1596	0.2051
10	7	0.0000	0.0000	0.0001	0.0008	0.0031	0.0090	0.0212	0.0425	0.0746	0.1172
10	8	0.0000	0.0000	0.0000	0.0001	0.0004	0.0014	0.0043	0.0106	0.0229	0.0439
10	9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0016	0.0042	0.0098
10	10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0010
11	0	0.5688	0.3138	0.1673	0.0859	0.0422	0.0198	0.0088	0.0036	0.0014	0.0005
11	1	0.3293	0.3835	0.3248	0.2362	0.1549	0.0932	0.0518	0.0266	0.0125	0.0054
11	2	0.0867	0.2131	0.2866	0.2953	0.2581	0.1998	0.1395	0.0887	0.0513	0.0269
11	3	0.0137	0.0710	0.1517	0.2215	0.2581	0.2568	0.2254	0.1774	0.1259	0.0806
11	4	0.0014	0.0158	0.0536	0.1107	0.1721	0.2201	0.2428	0.2365	0.2060	0.1611
11	5	0.0001	0.0025	0.0132	0.0388	0.0803	0.1321	0.1830	0.2207	0.2360	0.2256
11	6	0.0000	0.0003	0.0023	0.0097	0.0268	0.0566	0.0985	0.1471	0.1931	0.2256
11	7	0.0000	0.0000	0.0003	0.0017	0.0064	0.0173	0.0379	0.0701	0.1128	0.1611
11	8	0.0000	0.0000	0.0000	0.0002	0.0011	0.0037	0.0102	0.0234	0.0462	0.0806
11	9	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0018	0.0052	0.0126	0.0269
11	10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0007	0.0021	0.0054
11	11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0005
12	0	0.5404	0.2824	0.1422	0.0687	0.0317	0.0138	0.0057	0.0022	0.0008	0.0002
12	1	0.3413	0.3766	0.3012	0.2062	0.1267	0.0712	0.0368	0.0174	0.0075	0.0029
12	2	0.0988	0.2301	0.2924	0.2835	0.2323	0.1678	0.1088	0.0639	0.0339	0.0161
12	3	0.0173	0.0852	0.1720	0.2362	0.2581	0.2397	0.1954	0.1419	0.0923	0.0537
12	4	0.0021	0.0213	0.0683	0.1329	0.1936	0.2311	0.2367	0.2128	0.1700	0.1208
12	5	0.0002	0.0038	0.0193	0.0532	0.1032	0.1585	0.2039	0.2270	0.2225	0.1934
12	6	0.0000	0.0005	0.0040	0.0155	0.0401	0.0792	0.1281	0.1766	0.2124	0.2256
12	7	0.0000	0.0000	0.0006	0.0033	0.0115	0.0291	0.0591	0.1009	0.1489	0.1934
12	8	0.0000	0.0000	0.0001	0.0005	0.0024	0.0078	0.0199	0.0420	0.0762	0.1208
12	9	0.0000	0.0000	0.0000	0.0001	0.0004	0.0015	0.0048	0.0125	0.0277	0.0537
12	10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0008	0.0025	0.0068	0.0161
12	11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0010	0.0029
12	12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002

TABLA 5 (CONTINUACIÓN)

						p					
n	*k*	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
13	0	0.5133	0.2542	0.1209	0.0550	0.0238	0.0097	0.0037	0.0013	0.0004	0.0001
13	1	0.3512	0.3672	0.2774	0.1787	0.1029	0.0540	0.0259	0.0113	0.0045	0.0016
13	2	0.1109	0.2448	0.2937	0.2680	0.2059	0.1388	0.0836	0.0453	0.0220	0.0095
13	3	0.0214	0.0997	0.1900	0.2457	0.2517	0.2181	0.1651	0.1107	0.0660	0.0349
13	4	0.0028	0.0277	0.0838	0.1535	0.2097	0.2337	0.2222	0.1845	0.1350	0.0873
13	5	0.0003	0.0055	0.0266	0.0691	0.1258	0.1803	0.2154	0.2214	0.1989	0.1571
13	6	0.0000	0.0008	0.0063	0.0230	0.0559	0.1030	0.1546	0.1968	0.2169	0.2095
13	7	0.0000	0.0001	0.0011	0.0058	0.0186	0.0442	0.0833	0.1312	0.1775	0.2095
13	8	0.0000	0.0000	0.0001	0.0011	0.0047	0.0142	0.0336	0.0656	0.1089	0.1571
13	9	0.0000	0.0000	0.0000	0.0001	0.0009	0.0034	0.0101	0.0243	0.0495	0.0873
13	10	0.0000	0.0000	0.0000	0.0000	0.0001	0.0006	0.0022	0.0065	0.0162	0.0349
13	11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0012	0.0036	0.0095
13	12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0016
13	13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
14	0	0.4877	0.2288	0.1028	0.0440	0.0178	0.0068	0.0024	0.0008	0.0002	0.0001
14	1	0.3593	0.3559	0.2539	0.1539	0.0832	0.0407	0.0181	0.0073	0.0027	0.0009
14	2	0.1229	0.2570	0.2912	0.2501	0.1802	0.1134	0.0634	0.0317	0.0141	0.0056
14	3	0.0259	0.1142	0.2056	0.2501	0.2402	0.1943	0.1366	0.0845	0.0462	0.0222
14	4	0.0037	0.0349	0.0998	0.1720	0.2202	0.2290	0.2022	0.1549	0.1040	0.0611
14	5	0.0004	0.0078	0.0352	0.0860	0.1468	0.1963	0.2178	0.2066	0.1701	0.1222
14	6	0.0000	0.0013	0.0093	0.0322	0.0734	0.1262	0.1759	0.2066	0.2088	0.1833
14	7	0.0000	0.0002	0.0019	0.0092	0.0280	0.0618	0.1082	0.1574	0.1952	0.2095
14	8	0.0000	0.0000	0.0003	0.0020	0.0082	0.0232	0.0510	0.0918	0.1398	0.1833
14	9	0.0000	0.0000	0.0000	0.0003	0.0018	0.0066	0.0183	0.0408	0.0762	0.1222
14	10	0.0000	0.0000	0.0000	0.0000	0.0003	0.0014	0.0049	0.0136	0.0312	0.0611
14	11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0010	0.0033	0.0093	0.0222
14	12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0019	0.0056
14	13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0009
14	14	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
15	0	0.4633	0.2059	0.0874	0.0352	0.0134	0.0047	0.0016	0.0005	0.0001	0.0000
15	1	0.3658	0.3432	0.2312	0.1319	0.0668	0.0305	0.0126	0.0047	0.0016	0.0005
15	2	0.1348	0.2669	0.2856	0.2309	0.1559	0.0916	0.0476	0.0219	0.0090	0.0032
15	3	0.0307	0.1285	0.2184	0.2501	0.2252	0.1700	0.1110	0.0634	0.0318	0.0139
15	4	0.0049	0.0428	0.1156	0.1876	0.2252	0.2186	0.1792	0.1268	0.0780	0.0417
15	5	0.0006	0.0105	0.0449	0.1032	0.1651	0.2061	0.2123	0.1859	0.1404	0.0916
15	6	0.0000	0.0019	0.0132	0.0430	0.0917	0.1472	0.1906	0.2066	0.1914	0.1527
15	7	0.0000	0.0003	0.0030	0.0138	0.0393	0.0811	0.1319	0.1771	0.2013	0.1964
15	8	0.0000	0.0000	0.0005	0.0035	0.0131	0.0348	0.0710	0.1181	0.1647	0.1964
15	9	0.0000	0.0000	0.0001	0.0007	0.0034	0.0116	0.0298	0.0612	0.1048	0.1527
15	10	0.0000	0.0000	0.0000	0.0001	0.0007	0.0030	0.0096	0.0245	0.0515	0.0916
15	11	0.0000	0.0000	0.0000	0.0000	0.0001	0.0006	0.0024	0.0074	0.0191	0.0417
15	12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0016	0.0052	0.0139
15	13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0010	0.0032
15	14	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005
15	15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

						p					
n	*k*	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
16	0	0.4401	0.1853	0.0743	0.0281	0.0100	0.0033	0.0010	0.0003	0.0001	0.0000
16	1	0.3706	0.3294	0.2097	0.1126	0.0535	0.0228	0.0087	0.0030	0.0009	0.0002
16	2	0.1463	0.2745	0.2775	0.2111	0.1336	0.0732	0.0353	0.0150	0.0056	0.0018
16	3	0.0359	0.1423	0.2285	0.2463	0.2079	0.1465	0.0888	0.0468	0.0215	0.0085
16	4	0.0061	0.0514	0.1311	0.2001	0.2252	0.2040	0.1553	0.1014	0.0572	0.0278
16	5	0.0008	0.0137	0.0555	0.1201	0.1802	0.2099	0.2008	0.1623	0.1123	0.0667
16	6	0.0001	0.0028	0.0180	0.0550	0.1101	0.1649	0.1982	0.1983	0.1684	0.1222
16	7	0.0000	0.0004	0.0045	0.0197	0.0524	0.1010	0.1524	0.1889	0.1969	0.1746
16	8	0.0000	0.0001	0.0009	0.0055	0.0197	0.0487	0.0923	0.1417	0.1812	0.1964
16	9	0.0000	0.0000	0.0001	0.0012	0.0058	0.0185	0.0442	0.0840	0.1318	0.1746
16	10	0.0000	0.0000	0.0000	0.0002	0.0014	0.0056	0.0167	0.0392	0.0755	0.1222
16	11	0.0000	0.0000	0.0000	0.0000	0.0002	0.0013	0.0049	0.0142	0.0337	0.0667
16	12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0011	0.0040	0.0115	0.0278
16	13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0008	0.0029	0.0085
16	14	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0018
16	15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002
16	16	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
17	0	0.4181	0.1668	0.0631	0.0225	0.0075	0.0023	0.0007	0.0002	0.0000	0.0000
17	1	0.3741	0.3150	0.1893	0.0957	0.0426	0.0169	0.0060	0.0019	0.0005	0.0001
17	2	0.1575	0.2800	0.2673	0.1914	0.1136	0.0581	0.0260	0.0102	0.0035	0.0010
17	3	0.0415	0.1556	0.2359	0.2393	0.1893	0.1245	0.0701	0.0341	0.0144	0.0052
17	4	0.0076	0.0605	0.1457	0.2093	0.2209	0.1868	0.1320	0.0796	0.0411	0.0182
17	5	0.0010	0.0175	0.0668	0.1361	0.1914	0.2081	0.1849	0.1379	0.0875	0.0472
17	6	0.0001	0.0039	0.0236	0.0680	0.1276	0.1784	0.1991	0.1839	0.1432	0.0944
17	7	0.0000	0.0007	0.0065	0.0267	0.0668	0.1201	0.1685	0.1927	0.1841	0.1484
17	8	0.0000	0.0001	0.0014	0.0084	0.0279	0.0644	0.1134	0.1606	0.1883	0.1855
17	9	0.0000	0.0000	0.0003	0.0021	0.0093	0.0276	0.0611	0.1070	0.1540	0.1855
17	10	0.0000	0.0000	0.0000	0.0004	0.0025	0.0095	0.0263	0.0571	0.1008	0.1484
17	11	0.0000	0.0000	0.0000	0.0001	0.0005	0.0026	0.0090	0.0242	0.0525	0.0944
17	12	0.0000	0.0000	0.0000	0.0000	0.0001	0.0006	0.0024	0.0081	0.0215	0.0472
17	13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0021	0.0068	0.0182
17	14	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0016	0.0052
17	15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0010
17	16	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
17	17	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
18	0	0.3972	0.1501	0.0536	0.0180	0.0056	0.0016	0.0004	0.0001	0.0000	0.0000
18	1	0.3763	0.3002	0.1704	0.0811	0.0338	0.0126	0.0042	0.0012	0.0003	0.0001
18	2	0.1683	0.2835	0.2556	0.1723	0.0958	0.0458	0.0190	0.0069	0.0022	0.0006
18	3	0.0473	0.1680	0.2406	0.2297	0.1704	0.1046	0.0547	0.0246	0.0095	0.0031
18	4	0.0093	0.0700	0.1592	0.2153	0.2130	0.1681	0.1104	0.0614	0.0291	0.0117
18	5	0.0014	0.0218	0.0787	0.1507	0.1988	0.2017	0.1664	0.1146	0.0666	0.0327
18	6	0.0002	0.0052	0.0301	0.0816	0.1436	0.1873	0.1941	0.1655	0.1181	0.0708
18	7	0.0000	0.0010	0.0091	0.0350	0.0820	0.1376	0.1792	0.1892	0.1657	0.1214
18	8	0.0000	0.0002	0.0022	0.0120	0.0376	0.0811	0.1327	0.1734	0.1864	0.1669
18	9	0.0000	0.0000	0.0004	0.0033	0.0139	0.0386	0.0794	0.1284	0.1694	0.1855
18	10	0.0000	0.0000	0.0001	0.0008	0.0042	0.0149	0.0385	0.0771	0.1248	0.1669
18	11	0.0000	0.0000	0.0000	0.0001	0.0010	0.0046	0.0151	0.0374	0.0742	0.1214
18	12	0.0000	0.0000	0.0000	0.0000	0.0002	0.0012	0.0047	0.0145	0.0354	0.0708
18	13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0012	0.0045	0.0134	0.0327
18	14	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0011	0.0039	0.0117
18	15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0009	0.0031
18	16	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0006
18	17	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
18	18	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

TABLA 5 (CONTINUACIÓN)

						p					
n	*k*	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
19	0	0.3774	0.1351	0.0456	0.0144	0.0042	0.0011	0.0003	0.0001	0.0000	0.0000
19	1	0.3774	0.2852	0.1529	0.0685	0.0268	0.0093	0.0029	0.0008	0.0002	0.0000
19	2	0.1787	0.2852	0.2428	0.1540	0.0803	0.0358	0.0138	0.0046	0.0013	0.0003
19	3	0.0533	0.1796	0.2428	0.2182	0.1517	0.0869	0.0422	0.0175	0.0062	0.0018
19	4	0.0112	0.0798	0.1714	0.2182	0.2023	0.1491	0.0909	0.0467	0.0203	0.0074
19	5	0.0018	0.0266	0.0907	0.1636	0.2023	0.1916	0.1468	0.0933	0.0497	0.0222
19	6	0.0002	0.0069	0.0374	0.0955	0.1574	0.1916	0.1844	0.1451	0.0949	0.0518
19	7	0.0000	0.0014	0.0122	0.0443	0.0974	0.1525	0.1844	0.1797	0.1443	0.0961
19	8	0.0000	0.0002	0.0032	0.0166	0.0487	0.0981	0.1489	0.1797	0.1771	0.1442
19	9	0.0000	0.0000	0.0007	0.0051	0.0198	0.0514	0.0980	0.1464	0.1771	0.1762
19	10	0.0000	0.0000	0.0001	0.0013	0.0066	0.0220	0.0528	0.0976	0.1449	0.1762
19	11	0.0000	0.0000	0.0000	0.0003	0.0018	0.0077	0.0233	0.0532	0.0970	0.1442
19	12	0.0000	0.0000	0.0000	0.0000	0.0004	0.0022	0.0083	0.0237	0.0529	0.0961
19	13	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0024	0.0085	0.0233	0.0518
19	14	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0006	0.0024	0.0082	0.0222
19	15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0022	0.0074
19	16	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0018
19	17	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003
19	18	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
19	19	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
20	0	0.3585	0.1216	0.0388	0.0115	0.0032	0.0008	0.0002	0.0000	0.0000	0.0000
20	1	0.3774	0.2702	0.1368	0.0576	0.0211	0.0068	0.0020	0.0005	0.0001	0.0000
20	2	0.1887	0.2852	0.2293	0.1369	0.0669	0.0278	0.0100	0.0031	0.0008	0.0002
20	3	0.0596	0.1901	0.2428	0.2054	0.1339	0.0716	0.0323	0.0123	0.0040	0.0011
20	4	0.0133	0.0898	0.1821	0.2182	0.1897	0.1304	0.0738	0.0350	0.0139	0.0046
20	5	0.0022	0.0319	0.1028	0.1746	0.2023	0.1789	0.1272	0.0746	0.0365	0.0148
20	6	0.0003	0.0089	0.0454	0.1091	0.1686	0.1916	0.1712	0.1244	0.0746	0.0370
20	7	0.0000	0.0020	0.0160	0.0545	0.1124	0.1643	0.1844	0.1659	0.1221	0.0739
20	8	0.0000	0.0004	0.0046	0.0222	0.0609	0.1144	0.1614	0.1797	0.1623	0.1201
20	9	0.0000	0.0001	0.0011	0.0074	0.0271	0.0654	0.1158	0.1597	0.1771	0.1602
20	10	0.0000	0.0000	0.0002	0.0020	0.0099	0.0308	0.0686	0.1171	0.1593	0.1762
20	11	0.0000	0.0000	0.0000	0.0005	0.0030	0.0120	0.0336	0.0710	0.1185	0.1602
20	12	0.0000	0.0000	0.0000	0.0001	0.0008	0.0039	0.0136	0.0355	0.0727	0.1201
20	13	0.0000	0.0000	0.0000	0.0000	0.0002	0.0010	0.0045	0.0146	0.0366	0.0739
20	14	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0012	0.0049	0.0150	0.0370
20	15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003	0.0013	0.0049	0.0148
20	16	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003	0.0013	0.0046
20	17	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0011
20	18	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002
20	19	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
20	20	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

TABLA 6: PROBABILIDADES DE POISSON

					λ					
k	0.005	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.9950	0.9900	0.9802	0.9704	0.9608	0.9512	0.9418	0.9324	0.9231	0.9139
1	0.0050	0.0099	0.0196	0.0291	0.0384	0.0476	0.0565	0.0653	0.0738	0.0823
2	0.0000	0.0000	0.0002	0.0004	0.0008	0.0012	0.0017	0.0023	0.0030	0.0037
3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001

					λ					
k	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066	0.3679
1	0.0905	0.1637	0.2222	0.2681	0.3033	0.3293	0.3476	0.3595	0.3659	0.3679
2	0.0045	0.0164	0.0333	0.0536	0.0758	0.0988	0.1217	0.1438	0.1647	0.1839
3	0.0002	0.0011	0.0033	0.0072	0.0126	0.0198	0.0284	0.0383	0.0494	0.0613
4	0.0000	0.0001	0.0003	0.0007	0.0016	0.0030	0.0050	0.0077	0.0111	0.0153
5	0.0000	0.0000	0.0000	0.0001	0.0002	0.0004	0.0007	0.0012	0.0020	0.0031
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0003	0.0005
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001

					λ					
k	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0	0.3329	0.3012	0.2725	0.2466	0.2231	0.2019	0.1827	0.1653	0.1496	0.1353
1	0.3662	0.3614	0.3543	0.3452	0.3347	0.3230	0.3106	0.2975	0.2842	0.2707
2	0.2014	0.2169	0.2303	0.2417	0.2510	0.2584	0.2640	0.2678	0.2700	0.2707
3	0.0738	0.0867	0.0998	0.1128	0.1255	0.1378	0.1496	0.1607	0.1710	0.1804
4	0.0203	0.0260	0.0324	0.0395	0.0471	0.0551	0.0636	0.0723	0.0812	0.0902
5	0.0045	0.0062	0.0084	0.0111	0.0141	0.0176	0.0216	0.0260	0.0309	0.0361
6	0.0008	0.0012	0.0018	0.0026	0.0035	0.0047	0.0061	0.0078	0.0098	0.0120
7	0.0001	0.0002	0.0003	0.0005	0.0008	0.0011	0.0015	0.0020	0.0027	0.0034
8	0.0000	0.0000	0.0001	0.0001	0.0001	0.0002	0.0003	0.0005	0.0006	0.0009
9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0002

					λ					
k	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
0	0.1225	0.1108	0.1003	0.0907	0.0821	0.0743	0.0672	0.0608	0.0550	0.0498
1	0.2572	0.2438	0.2306	0.2177	0.2052	0.1931	0.1815	0.1703	0.1596	0.1494
2	0.2700	0.2681	0.2652	0.2613	0.2565	0.2510	0.2450	0.2384	0.2314	0.2240
3	0.1890	0.1966	0.2033	0.2090	0.2138	0.2176	0.2205	0.2225	0.2237	0.2240
4	0.0992	0.1082	0.1169	0.1254	0.1336	0.1414	0.1488	0.1557	0.1622	0.1680
5	0.0417	0.0476	0.0538	0.0602	0.0668	0.0735	0.0804	0.0872	0.0940	0.1008
6	0.0146	0.0174	0.0206	0.0241	0.0278	0.0319	0.0362	0.0407	0.0455	0.0504
7	0.0044	0.0055	0.0068	0.0083	0.0099	0.0118	0.0139	0.0163	0.0188	0.0216
8	0.0011	0.0015	0.0019	0.0025	0.0031	0.0038	0.0047	0.0057	0.0068	0.0081
9	0.0003	0.0004	0.0005	0.0007	0.0009	0.0011	0.0014	0.0018	0.0022	0.0027
10	0.0001	0.0001	0.0001	0.0002	0.0002	0.0003	0.0004	0.0005	0.0006	0.0008
11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0002	0.0002
12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001

						λ				
k	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
0	0.0450	0.0408	0.0369	0.0334	0.0302	0.0273	0.0247	0.0224	0.0202	0.0183
1	0.1397	0.1304	0.1217	0.1135	0.1057	0.0984	0.0915	0.0850	0.0789	0.0733
2	0.2165	0.2087	0.2008	0.1929	0.1850	0.1771	0.1692	0.1615	0.1539	0.1465
3	0.2237	0.2226	0.2209	0.2186	0.2158	0.2125	0.2087	0.2046	0.2001	0.1954
4	0.1733	0.1781	0.1823	0.1858	0.1888	0.1912	0.1931	0.1944	0.1951	0.1954
5	0.1075	0.1140	0.1203	0.1264	0.1322	0.1377	0.1429	0.1477	0.1522	0.1563
6	0.0555	0.0608	0.0662	0.0716	0.0771	0.0826	0.0881	0.0936	0.0989	0.1042
7	0.0246	0.0278	0.0312	0.0348	0.0385	0.0425	0.0466	0.0508	0.0551	0.0595
8	0.0095	0.0111	0.0129	0.0148	0.0169	0.0191	0.0215	0.0241	0.0269	0.0298
9	0.0033	0.0040	0.0047	0.0056	0.0066	0.0076	0.0089	0.0102	0.0116	0.0132
10	0.0010	0.0013	0.0016	0.0019	0.0023	0.0028	0.0033	0.0039	0.0045	0.0053
11	0.0003	0.0004	0.0005	0.0006	0.0007	0.0009	0.0011	0.0013	0.0016	0.0019
12	0.0001	0.0001	0.0001	0.0002	0.0002	0.0003	0.0003	0.0004	0.0005	0.0006
13	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002
14	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001

						λ				
k	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5.0
0	0.0166	0.0150	0.0136	0.0123	0.0111	0.0101	0.0091	0.0082	0.0074	0.0067
1	0.0679	0.0630	0.0583	0.0540	0.0500	0.0462	0.0427	0.0395	0.0365	0.0337
2	0.1393	0.1323	0.1254	0.1188	0.1125	0.1063	0.1005	0.0948	0.0894	0.0842
3	0.1904	0.1852	0.1798	0.1743	0.1687	0.1631	0.1574	0.1517	0.1460	0.1404
4	0.1951	0.1944	0.1933	0.1917	0.1898	0.1875	0.1849	0.1820	0.1789	0.1755
5	0.1600	0.1633	0.1662	0.1687	0.1708	0.1725	0.1738	0.1747	0.1753	0.1755
6	0.1093	0.1143	0.1191	0.1237	0.1281	0.1323	0.1362	0.1398	0.1432	0.1462
7	0.0640	0.0686	0.0732	0.0778	0.0824	0.0869	0.0914	0.0959	0.1002	0.1044
8	0.0328	0.0360	0.0393	0.0428	0.0463	0.0500	0.0537	0.0575	0.0614	0.0653
9	0.0150	0.0168	0.0188	0.0209	0.0232	0.0255	0.0281	0.0307	0.0334	0.0363
10	0.0061	0.0071	0.0081	0.0092	0.0104	0.0118	0.0132	0.0147	0.0164	0.0181
11	0.0023	0.0027	0.0032	0.0037	0.0043	0.0049	0.0056	0.0064	0.0073	0.0082
12	0.0008	0.0009	0.0011	0.0013	0.0016	0.0019	0.0022	0.0026	0.0030	0.0034
13	0.0002	0.0003	0.0004	0.0005	0.0006	0.0007	0.0008	0.0009	0.0011	0.0013
14	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002	0.0003	0.0003	0.0004	0.0005
15	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002

						λ				
k	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6.0
0	0.0061	0.0055	0.0050	0.0045	0.0041	0.0037	0.0033	0.0030	0.0027	0.0025
1	0.0311	0.0287	0.0265	0.0244	0.0225	0.0207	0.0191	0.0176	0.0162	0.0149
2	0.0793	0.0746	0.0701	0.0659	0.0618	0.0580	0.0544	0.0509	0.0477	0.0446
3	0.1348	0.1293	0.1239	0.1185	0.1133	0.1082	0.1033	0.0985	0.0938	0.0892
4	0.1719	0.1681	0.1641	0.1600	0.1558	0.1515	0.1472	0.1428	0.1383	0.1339
5	0.1753	0.1748	0.1740	0.1728	0.1714	0.1697	0.1678	0.1656	0.1632	0.1606
6	0.1490	0.1515	0.1537	0.1555	0.1571	0.1584	0.1594	0.1601	0.1605	0.1606
7	0.1086	0.1125	0.1163	0.1200	0.1234	0.1267	0.1298	0.1326	0.1353	0.1377
8	0.0692	0.0731	0.0771	0.0810	0.0849	0.0887	0.0925	0.0962	0.0998	0.1033
9	0.0392	0.0423	0.0454	0.0486	0.0519	0.0552	0.0586	0.0620	0.0654	0.0688
10	0.0200	0.0220	0.0241	0.0262	0.0285	0.0309	0.0334	0.0359	0.0386	0.0413
11	0.0093	0.0104	0.0116	0.0129	0.0143	0.0157	0.0173	0.0190	0.0207	0.0225
12	0.0039	0.0045	0.0051	0.0058	0.0065	0.0073	0.0082	0.0092	0.0102	0.0113
13	0.0015	0.0018	0.0021	0.0024	0.0028	0.0032	0.0036	0.0041	0.0046	0.0052
14	0.0006	0.0007	0.0008	0.0009	0.0011	0.0013	0.0015	0.0017	0.0019	0.0022
15	0.0002	0.0002	0.0003	0.0003	0.0004	0.0005	0.0006	0.0007	0.0008	0.0009
16	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002	0.0002	0.0003	0.0003
17	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001

TABLA 6 (CONTINUACION)

	λ									
k	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	7.0
0	0.0022	0.0020	0.0018	0.0017	0.0015	0.0014	0.0012	0.0011	0.0010	0.0009
1	0.0137	0.0126	0.0116	0.0106	0.0098	0.0090	0.0082	0.0076	0.0070	0.0064
2	0.0417	0.0390	0.0364	0.0340	0.0318	0.0296	0.0276	0.0258	0.0240	0.0223
3	0.0848	0.0806	0.0765	0.0726	0.0688	0.0652	0.0617	0.0584	0.0552	0.0521
4	0.1294	0.1249	0.1205	0.1162	0.1118	0.1076	0.1034	0.0992	0.0952	0.0912
5	0.1579	0.1549	0.1519	0.1487	0.1454	0.1420	0.1385	0.1349	0.1314	0.1277
6	0.1605	0.1601	0.1595	0.1586	0.1575	0.1562	0.1546	0.1529	0.1511	0.1490
7	0.1399	0.1418	0.1435	0.1450	0.1462	0.1472	0.1480	0.1486	0.1489	0.1490
8	0.1066	0.1099	0.1130	0.1160	0.1188	0.1215	0.1240	0.1263	0.1284	0.1304
9	0.0723	0.0757	0.0791	0.0825	0.0858	0.0891	0.0923	0.0954	0.0985	0.1014
10	0.0441	0.0469	0.0498	0.0528	0.0558	0.0588	0.0618	0.0649	0.0679	0.0710
11	0.0244	0.0265	0.0285	0.0307	0.0330	0.0353	0.0377	0.0401	0.0426	0.0452
12	0.0124	0.0137	0.0150	0.0164	0.0179	0.0194	0.0210	0.0227	0.0245	0.0263
13	0.0058	0.0065	0.0073	0.0081	0.0089	0.0099	0.0108	0.0119	0.0130	0.0142
14	0.0025	0.0029	0.0033	0.0037	0.0041	0.0046	0.0052	0.0058	0.0064	0.0071
15	0.0010	0.0012	0.0014	0.0016	0.0018	0.0020	0.0023	0.0026	0.0029	0.0033
16	0.0004	0.0005	0.0005	0.0006	0.0007	0.0008	0.0010	0.0011	0.0013	0.0014
17	0.0001	0.0002	0.0002	0.0002	0.0003	0.0003	0.0004	0.0004	0.0005	0.0006
18	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002	0.0002
19	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001

	λ									
k	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	8.0
0	0.0008	0.0007	0.0007	0.0006	0.0006	0.0005	0.0005	0.0004	0.0004	0.0003
1	0.0059	0.0054	0.0049	0.0045	0.0041	0.0038	0.0035	0.0032	0.0029	0.0027
2	0.0208	0.0194	0.0180	0.0167	0.0156	0.0145	0.0134	0.0125	0.0116	0.0107
3	0.0492	0.0464	0.0438	0.0413	0.0389	0.0366	0.0345	0.0324	0.0305	0.0286
4	0.0874	0.0836	0.0799	0.0764	0.0729	0.0696	0.0663	0.0632	0.0602	0.0573
5	0.1241	0.1204	0.1167	0.1130	0.1094	0.1057	0.1021	0.0986	0.0951	0.0916
6	0.1468	0.1445	0.1420	0.1394	0.1367	0.1339	0.1311	0.1282	0.1252	0.1221
7	0.1489	0.1486	0.1481	0.1474	0.1465	0.1454	0.1442	0.1428	0.1413	0.1396
8	0.1321	0.1337	0.1351	0.1363	0.1373	0.1381	0.1388	0.1392	0.1395	0.1396
9	0.1042	0.1070	0.1096	0.1121	0.1144	0.1167	0.1187	0.1207	0.1224	0.1241
10	0.0740	0.0770	0.0800	0.0829	0.0858	0.0887	0.0914	0.0941	0.0967	0.0993
11	0.0478	0.0504	0.0531	0.0558	0.0585	0.0613	0.0640	0.0667	0.0695	0.0722
12	0.0283	0.0303	0.0323	0.0344	0.0366	0.0388	0.0411	0.0434	0.0457	0.0481
13	0.0154	0.0168	0.0181	0.0196	0.0211	0.0227	0.0243	0.0260	0.0278	0.0296
14	0.0078	0.0086	0.0095	0.0104	0.0113	0.0123	0.0134	0.0145	0.0157	0.0169
15	0.0037	0.0041	0.0046	0.0051	0.0057	0.0062	0.0069	0.0075	0.0083	0.0090
16	0.0016	0.0019	0.0021	0.0024	0.0026	0.0030	0.0033	0.0037	0.0041	0.0045
17	0.0007	0.0008	0.0009	0.0010	0.0012	0.0013	0.0015	0.0017	0.0019	0.0021
18	0.0003	0.0003	0.0004	0.0004	0.0005	0.0006	0.0006	0.0007	0.0008	0.0009
19	0.0001	0.0001	0.0001	0.0002	0.0002	0.0002	0.0003	0.0003	0.0003	0.0004
20	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002
21	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001

TABLA 7: TABLA DE NÚMEROS AL AZAR

Primera serie

	1 - 4		5 - 8		9 - 12		13 - 16		17 - 20		21 - 24		25 - 28		29 - 32		33 - 36		37 - 40	
1	20	77	81	43	63	92	68	61	70	79	88	81	05	47	63	07	13	10	46	19
2	09	42	70	58	14	76	27	02	29	73	87	66	47	73	31	59	02	96	41	07
3	93	04	55	07	83	92	26	76	50	57	05	97	12	85	01	30	82	45	52	08
4	95	99	93	67	54	96	35	98	84	64	80	88	29	39	07	00	97	95	59	24
5	40	82	40	08	05	80	60	50	33	93	68	58	83	62	06	09	20	56	91	36
6	78	11	44	01	19	42	06	02	32	19	99	23	94	02	29	27	29	38	17	82
7	56	41	30	34	77	26	83	55	26	08	69	53	66	16	19	43	77	69	70	77
8	76	42	48	52	69	11	70	01	23	21	99	22	30	75	42	61	99	20	91	90
9	41	14	93	39	41	11	56	76	60	04	24	75	18	06	14	42	91	25	31	92
10	17	42	51	18	60	28	10	87	61	25	88	92	04	30	90	80	32	26	91	22
11	96	66	80	87	48	97	22	47	84	24	58	51	41	10	54	26	93	19	90	20
12	24	81	91	42	70	40	96	75	48	30	48	66	21	54	20	98	12	00	86	61
13	78	65	68	07	07	95	15	50	67	10	01	62	36	75	93	76	40	54	97	68
14	29	27	78	63	25	00	14	51	15	18	18	14	03	96	63	08	85	49	16	14
15	34	16	38	45	71	04	00	72	44	03	63	46	49	56	50	76	57	32	84	43
16	82	76	24	97	43	39	05	39	93	69	61	80	25	47	90	15	70	06	74	13
17	18	93	50	05	65	07	39	37	51	99	78	42	52	78	82	86	81	17	69	09
18	46	84	90	64	55	19	12	20	32	11	56	30	00	54	75	95	54	22	80	38
19	59	52	94	41	54	33	08	80	51	39	35	64	22	90	59	82	79	76	23	22
20	38	12	76	09	53	32	80	07	19	34	18	55	60	86	33	22	36	15	79	85
21	14	72	18	71	55	19	09	25	27	36	10	35	60	87	96	55	74	86	08	54
22	44	29	94	19	34	91	62	94	56	81	35	00	79	15	62	92	66	16	67	29
23	50	10	67	79	43	27	66	85	52	00	97	65	07	58	31	74	90	09	24	75
24	11	19	88	34	80	11	94	03	56	28	53	52	86	83	51	38	97	02	50	20
25	12	16	81	62	90	38	45	23	13	08	18	57	67	45	15	75	86	07	77	57

Segunda serie

	1 - 4		5 - 8		9 - 12		13 - 16		17 - 20		21 - 24		25 - 28		29 - 32		33 - 36		37 - 40	
1	83	79	21	68	54	51	23	50	78	17	73	51	94	71	91	31	29	97	13	16
2	27	98	04	97	52	48	19	33	82	16	00	34	30	67	58	00	80	80	92	26
3	75	57	42	55	59	91	72	75	66	75	39	70	55	24	09	19	70	22	42	10
4	60	73	07	27	71	94	68	70	48	81	40	91	16	24	45	24	54	03	18	03
5	27	42	15	67	94	44	48	62	37	53	83	15	90	90	60	19	78	62	44	70
6	57	27	77	75	38	18	24	32	91	53	78	91	33	38	31	95	85	11	33	87
7	22	79	42	39	32	65	83	60	74	63	56	77	47	21	88	36	43	10	19	41
8	61	94	32	42	83	81	92	40	99	00	05	66	33	61	32	10	43	15	49	25
9	24	28	53	76	13	61	73	22	50	51	75	08	10	90	71	58	22	42	46	83
10	06	75	46	13	77	68	97	05	56	73	34	86	42	22	37	75	94	87	57	72
11	83	81	98	41	67	38	97	30	12	31	87	76	81	07	32	88	42	29	94	58
12	33	66	09	91	21	26	52	57	47	14	27	75	07	84	50	96	95	12	25	01
13	63	14	10	59	10	68	27	91	00	17	36	79	01	79	65	43	13	98	52	21
14	84	43	66	38	65	72	14	55	93	78	24	57	38	10	54	53	92	41	82	56
15	36	75	92	36	76	77	89	27	06	57	37	70	36	09	99	90	66	44	91	89
16	17	23	20	39	81	03	49	79	68	20	94	45	95	92	63	06	55	29	20	82
17	57	09	57	60	40	64	00	77	31	05	83	44	96	62	56	42	42	68	46	42
18	33	49	47	96	50	21	75	68	28	12	46	25	72	64	50	72	75	16	67	00
19	80	52	82	00	02	12	67	61	43	23	14	53	10	66	16	29	06	60	23	20
20	45	29	44	64	44	70	22	10	70	26	43	49	28	51	69	52	85	95	98	58
21	16	04	54	67	60	53	51	60	82	45	56	26	08	89	42	25	93	76	69	73
22	81	09	97	91	19	60	91	96	53	66	45	33	75	09	67	42	59	02	18	97
23	20	61	83	96	58	12	64	99	78	24	94	36	99	60	07	70	27	40	97	15
24	43	17	00	05	86	39	95	58	12	35	84	31	97	75	50	52	06	88	27	79
25	27	84	00	90	41	44	05	43	36	93	01	37	91	93	24	78	00	12	04	12

TABLA7 (CONTINUACION)

Tercera serie

	01 - 04		05 - 08		09 - 12		13 - 16		17 - 20		21 - 24		25 - 28		29 - 32		33 - 36		37 - 40	
1	35	87	68	11	29	27	78	34	74	92	86	13	33	22	34	75	59	44	61	63
2	16	58	21	36	34	35	92	56	72	85	03	91	31	33	26	62	08	30	95	09
3	44	20	51	20	95	87	87	37	37	05	06	43	61	75	51	99	52	03	23	29
4	36	04	26	10	46	13	47	92	34	02	53	54	97	57	24	18	37	62	98	35
5	92	15	17	81	81	67	17	00	76	64	77	34	20	18	88	01	88	31	18	40
6	29	12	08	74	26	94	48	03	85	60	23	69	65	52	53	18	26	72	47	62
7	19	37	78	71	07	12	90	53	13	34	19	49	81	36	85	96	19	81	68	65
8	14	92	75	64	31	51	00	20	86	47	07	04	40	26	57	62	34	95	57	17
9	02	19	07	98	20	98	21	10	69	15	39	55	80	23	80	26	63	33	81	26
10	31	18	25	46	66	31	15	27	09	80	13	44	72	58	42	93	71	72	69	17
11	51	96	70	26	99	94	81	55	72	29	57	37	29	36	99	55	35	99	22	79
12	11	78	29	90	66	21	63	60	39	87	77	34	55	26	56	94	74	43	71	04
13	93	10	34	79	41	18	07	79	30	84	77	27	07	40	54	24	90	96	43	63
14	19	32	42	52	58	31	29	38	39	31	00	68	71	25	28	05	27	55	56	18
15	26	84	50	36	62	73	35	95	56	00	76	61	16	40	08	80	71	87	20	48
16	05	50	36	40	59	22	90	33	76	82	58	66	30	65	68	23	38	98	39	92
17	45	14	18	03	19	34	32	93	93	37	97	11	37	04	64	62	60	53	86	33
18	71	44	13	65	14	74	67	80	11	11	01	90	47	58	36	86	13	68	85	64
19	47	46	52	39	25	41	96	96	98	22	11	51	90	91	82	16	96	43	11	05
20	33	95	58	94	59	65	09	99	47	45	70	06	60	66	75	22	43	93	79	62
21	28	74	81	01	28	31	47	96	55	75	26	61	48	51	10	20	50	43	83	28
22	33	68	25	37	96	38	70	64	36	19	60	48	71	41	05	53	77	56	81	78
23	85	38	32	52	85	14	64	87	82	11	85	11	19	29	80	35	09	99	54	23
24	04	29	07	21	28	33	85	69	93	51	57	30	02	14	58	12	84	34	30	71
25	47	20	21	03	47	88	47	25	51	49	70	21	07	71	07	03	08	69	69	35

Cuarta serie

	01 - 04		05 - 08		09 - 12		13 - 16		17 - 20		21 - 24		25 - 28		29 - 32		33 - 36		37 - 40	
1	72	43	49	72	61	99	36	48	80	20	10	31	66	31	84	83	87	39	93	25
2	34	71	89	58	13	15	56	47	41	24	57	70	96	69	59	17	28	25	55	05
3	94	75	59	95	98	99	88	03	83	67	75	75	92	73	56	60	91	80	59	53
4	28	97	88	17	22	80	49	80	81	22	07	83	41	63	15	27	25	72	00	01
5	75	89	87	84	83	13	56	59	46	28	26	11	62	28	81	61	42	86	35	50
6	77	44	16	92	59	93	41	27	08	86	09	45	47	43	35	19	66	71	81	08
7	46	06	66	76	61	89	59	75	61	68	68	50	47	98	47	21	84	05	52	95
8	53	46	28	21	43	74	40	07	15	60	52	15	39	94	67	29	90	83	08	61
9	55	23	44	40	08	34	48	33	12	12	28	12	69	81	14	58	25	01	00	13
10	38	44	74	56	06	24	85	13	22	04	07	70	00	18	43	99	03	53	77	98
11	08	12	50	56	81	86	61	59	77	00	11	71	00	47	29	62	68	87	25	30
12	34	27	77	14	64	22	20	77	22	41	50	92	51	67	70	54	14	26	54	47
13	19	98	90	19	27	12	80	34	87	97	73	09	98	80	66	74	77	59	11	54
14	39	05	72	32	69	87	95	50	82	76	50	79	82	18	72	77	88	60	92	11
15	38	96	83	88	05	76	23	20	09	33	08	02	10	74	93	22	09	75	83	23
16	72	94	21	37	57	90	48	48	43	96	66	75	33	08	09	65	62	00	94	38
17	86	61	52	23	80	46	97	81	15	83	88	88	71	98	73	73	43	63	93	74
18	14	78	87	24	89	77	62	94	19	26	16	08	78	08	97	24	19	03	47	46
19	24	17	92	19	81	85	71	91	02	41	45	08	44	99	72	36	84	23	59	51
20	92	52	08	30	47	44	31	36	80	12	73	65	98	06	15	69	59	18	38	58
21	94	47	16	55	89	88	64	29	02	59	48	90	06	90	57	14	65	55	75	97
22	47	21	92	90	36	20	85	34	96	73	11	69	71	36	65	16	49	89	41	78
23	76	89	16	29	20	24	00	47	33	89	74	11	59	62	84	53	03	14	74	77
24	90	93	86	13	61	32	15	62	99	30	63	47	88	87	86	12	59	30	69	44
25	44	93	41	83	33	43	31	74	85	47	44	45	91	36	40	48	56	84	10	09