

CONTENIDO Y LÍMITES DE LA ESTADÍSTICA

1.1 Antecedentes históricos de la Estadística.- Toda persona, tanto en el ejercicio de su profesión como en su actividad cotidiana está en contacto con el periódico, televisión, Internet, revistas y otros medios, los cuales le ofrecen información en forma de datos.

Todo aquello que se relaciona con la recolección, procesamiento, análisis e interpretación de datos cualitativos y cuantitativos pertenece al campo de la Estadística.

Consecuentemente, algún conocimiento de Estadística le será de utilidad a la población en general, pero en particular, el conocimiento estadístico será vital para quienes estén inmersos en el estudio de la ciencia y la ingeniería, en cualquiera de sus campos.

La Estadística surge porque en la mayoría de los procesos existe variabilidad. La variabilidad es el resultado de los cambios que ocurren en las condiciones en medio de las cuales se realizan dichos procesos.

La Estadística como tal es el resultado de la unión de dos disciplinas que evolucionaron independientemente hasta confluir en el siglo XIX: la primera es el "cálculo de probabilidades", que se origina aproximadamente en el siglo XVII como teoría matemática de los juegos de azar y, la segunda es como "ciencia del Estado", resultado de la necesidad de efectuar una descripción numérica de entidades políticas tales como ciudades, provincias, países, etc..

Sin embargo, mucho antes del siglo XVII, la gente registraba y utilizaba datos. El Antiguo Testamento contiene informes sobre levantamiento de censos. En la Edad Media, los gobernantes empezaron a registrar la propiedad de la tierra. En el año 762, Carlomagno solicitó la descripción detallada de la

cantidad de siervos que existían en cada feudo. Cerca al año 1086, Guillermo el Conquistador ordenó que se escribiera el *Domesday Book*, un registro de la propiedad, extensión y valor de las tierras de Inglaterra. Posteriormente, cerca al año 1500, tanto en Inglaterra como en Francia, se efectuaron censos de bautismos, defunciones y matrimonio, etc..

1.2 Significado de la Estadística.- Es posible comprender el sentido de la Estadística como:

“La Estadística es una *ciencia* que utiliza instrumentos de la Matemática, particularmente el cálculo de probabilidades, para estudiar fenómenos que dependen del azar (fenómenos aleatorios) a partir de una muestra, tal que las reglas de comportamiento obtenidas de la muestra se refieran a la población”

1.3 División de la Estadística.- Para un mejor estudio, la Estadística se divide en dos partes:

a) Estadística Descriptiva, cuya función es organizar, presentar y resumir los datos de una población o una muestra, a fin de describir y analizar sus características mediante valores numéricos y atributos.

Muchas de las técnicas de la Estadística Descriptiva se han empleado desde hace más de 200 años y se han originado en estudios y actividades de censos. La moderna tecnología de las computadoras, en particular las gráficas por computadora y los paquetes estadísticos, han ampliado en forma considerable el campo de la Estadística Descriptiva en los últimos años.

b) Estadística Inferencial, cuya función es de inducción o inferencia, es decir, su propósito es formular leyes generales sobre el comportamiento de los elementos de una población a partir el examen de una muestra extraída de ella.

La mayor parte de las técnicas de la Estadística Inferencial se han desarrollado en los últimos 80 años, en consecuencia, es una rama de las Estadística mucho más reciente que la Estadística Descriptiva.

1.4 Población y muestra.-

1.4.1 Población.- Es el total de un conjunto de elementos (animales, personas u objetos) que poseen determinadas propiedades o características comunes que los hace distinguibles respecto a elementos de otra población. Los elementos que caracterizan a una población se denominan *parámetros* y se representan mediante las letras del alfabeto griego.

Según la cantidad de elementos, existen dos tipos de población:

- a) Población finita.** Es aquella población que está conformada por un conjunto numerable de elementos. Generalmente se asume a aquella población con menos de 500.000 unidades.
- b) Población infinita.** Es aquella que está conformada por un conjunto no numerable de elementos o cuando la cantidad sobrepasa las 500.000 unidades.

1.4.2. Muestra.- Es un subconjunto de la población y se utiliza con el propósito de representar la población o universo y permitir los trabajos empíricos. La muestra es aplicable en los universos o poblaciones que no son manejables puesto que se reducen costos y tiempo y, las conclusiones no distan mucho de la realidad.

La muestra cumple la función de caracterizar los elementos de una población a partir de un número limitado de los elementos de la población. Los elementos que caracterizan a una muestra se denominan *estadígrafos* y se representan por las letras del alfabeto latino.

BIBLIOGRAFÍA:

- (1) HINES Walter y MONTGOMERY David (1996): "*Probabilidad y Estadística para Ingeniería y Administración*". McGraw-Hill, México.
- (2) KINNEAR Thomas y TAYLOR James (1994). "*Investigación de mercados*". McGraw-Hill, México.
- (3) MILLER Irwin, FREUND John y JOHNSON Richard (1994): "*Probabilidad y estadística para ingenieros*", México.
- (4) MOYA Rufino (1988): "*Estadística Descriptiva*". Perú.

=====

ÍNDICE

	Pág.
1.1. Antecedentes históricos de la Estadística.....	1

1.2.	Significado de la Estadística.....	2
1.3.	División de la Estadística.....	2
1.4.	Población y muestra.....	3
1.4.1.	Población.....	3
1.4.2.	Muestra.....	3

MÉTODOS DE RECOLECCIÓN DE DATOS

2.1 Introducción.- La recolección de datos es un proceso que permite reunir información de tal modo que de lugar a describir una población o muestra.

2.2 Fuente de datos.- La recolección de datos puede efectuarse a partir de dos tipos de fuentes:

- a) Fuentes primarias. En este caso, la información proviene de elementos, personas o situaciones que se observan en una población o muestra de modo tal que responda a las necesidades de la investigación; para ello se aplica diversos instrumentos tales como: la observación, entrevista, cuestionario, experimentación, lluvia de ideas, grupos focales, etc..
- b) Fuentes secundarias. En este caso, la información requerida se encuentra en informes, anuarios, folletos, Internet, etc., y son datos que se recolectan con propósitos diferentes de las necesidades específicas de la investigación.

2.3 La observación.- Es una forma de recolección de datos mediante la cual se utiliza cualquiera de los sentidos, en especial el de la vista, para efectuar un reconocimiento y registro del comportamiento de las personas, objetos o sucesos.

2.3.1 Clases de observación.- En este proceso de conocer y descubrir, la observación utiliza distintos procedimientos y adopta modalidades diferentes. Las distintas formas de observación que pueden presentarse son las siguientes:

- a) Estructurada o no estructurada. La observación *no estructurada* se denomina también simple, ordinaria o libre y consiste en recoger la información sin un guión preestablecido y sin la ayuda de medios o instrumentos

especiales. Se utiliza para tener el primer contacto con la realidad, percibir sus problemas, ordenarlos y formular hipótesis.

La observación *estructurada* es conocida también como observación sistemática por que sigue un plan determinado para alcanzar propósitos definidos y utiliza diversos instrumentos para captar los hechos con mayor objetividad tales como cuadros, anotaciones, escalas, etc..

b) Participante o no participante, según si el observador obtiene los datos participando de modo activo o no en la vida de la población o muestra sujeto de estudio.

c) Individual o en equipo, según si la observación se realiza por un individuo o un equipo, que se hace responsable de la validez, confiabilidad y precisión de los datos.

Las modalidades anteriores no son excluyentes y en el proceso de observación puede utilizarse una modalidad o modalidades combinadas, según el plan de observación y los propósitos de la investigación.

2.3.2 Ventajas.- Las principales ventajas de la observación son:

- Favorece la recolección de la información permitiendo que ésta sea independiente del deseo, capacidad y veracidad de los elementos que son motivos de observación.
- Permite recolectar información y efectuar el análisis de ésta sin la necesidad de intermediarios.

2.3.3 Desventajas.- Las principales desventajas de la observación son:

- La información obtenida puede ser distorsionada por los observadores o por el uso inadecuado de los instrumentos de la observación.
- Constituye un procedimiento muy costoso.
- No es conveniente cuando se estudian poblaciones muy numerosas.

2.4 La entrevista.- La entrevista es una técnica de recolección de información que consiste en que una persona llamada entrevistador, mediante la interacción verbal solicita información a otra persona llamada entrevistada.

Los instrumentos que ayudan para lograr una mayor eficacia con el uso de esta técnica son: la grabadora, esquemas, filmadoras, etc..

Para efectuar la entrevista puede utilizarse la conversación libre o el interrogatorio estructurado. En todos los casos el entrevistador debe conducir la entrevista empleando un guión o bosquejo de asuntos que oriente sus tareas.

2.4.1. Tipos de entrevista.- La entrevista puede ser clasificada en dos grandes grupos:

2.4.1.1 Entrevista no estructurada.- Consiste en dar libre iniciativa al entrevistado. En este caso se utilizan preguntas abiertas, que son respondidas por el entrevistado en su vocabulario y según su capacidad de comprensión.

2.4.1.2 Entrevista estructurada.- El entrevistador utiliza un formulario y somete a los entrevistados a un mismo orden de preguntas, planteadas con el mismo énfasis y en los mismos términos.

2.4.2 Ventajas.- Las ventajas más importantes de la entrevista son:

- Permite obtener información que es susceptible de manejo y tratamiento estadístico.
- Es posible obtener información sobre situaciones pasadas y actitudes futuras.

2.4.3 Desventajas.- Las desventajas más importantes de la entrevista son:

- La información ha obtener depende de la memoria y la buena fé del entrevistado, por lo que existe riesgo de distorsión en la información proporcionada.
- Con frecuencia se requiere personal capacitado y bastante tiempo para llevar a cabo las entrevistas, lo cual significa erogar grandes cantidades de recursos económicos.

2.5 El cuestionario.- Es un procedimiento que permite la recolección de información aplicando un formulario a una persona denominada encuestada, a objeto de proporcionar respuestas por escrito y sin la presencia del encuestador.

2.5.1 Ventajas.- Las ventajas del cuestionario son las siguientes:

- Puede aplicarse a una gran cantidad de personas, distribuidas en un área geográfica grande.
- El estudio se realiza con menores gastos de adiestramiento de personal y trabajo de campo, respecto a otras técnicas
- Las respuestas por escrito dan mayor libertad a la expresión y permite mantener el anonimato de las personas encuestadas.
- Se disminuye el riesgo de distorsión de la información que proviene de la presencia e influencia del encuestador.

2.5.2 Desventajas.- Las limitaciones del cuestionario son:

- Existe la posibilidad de altos índices de formularios sin respuestas o de cuestionarios incompletos.
- Existe imposibilidad de cooperar en la comprensión de preguntas y en la aplicación de las normas e instrucciones para el llenado del cuestionario.
- Existe exclusión de las personas que no saben leer o escribir.
- Existe recepción tardía de cuestionarios y pérdidas de cuestionarios por envíos equivocados.

2.6 La experimentación.- En la recolección de información requerida para propósitos de estudio se utilizan también métodos experimentales.

Un experimento se dice que es un modelo que está determinado por un conjunto de condiciones establecidas y por los resultados que se obtienen al efectuar el experimento en las condiciones establecidas.

Un experimento puede ser de dos clases: determinístico y aleatorio.

2.6.1. Experimento determinístico.- El experimento es determinístico si las condiciones que se establecen determinan el único modo en que aparecen los sucesos o resultados.

2.6.2. Experimento aleatorio.- El experimento es aleatorio si determinadas las condiciones de modo artificial o independientemente de la voluntad del experimentador, no es posible determinar los sucesos a obtener.

BIBLIOGRAFÍA:

- (1) HANKE Jhon y REITSCH Arthur (1996). *"Pronósticos en los negocios"*. Prentice Hall. México.
- (2) KINNEAR Thomas y TAYLOR James (1994). *"Investigación de mercados"*. McGraw-Hill, México.
- (3) MOYA Rufino (1988): *"Estadística Descriptiva"*. Perú.

=====

ÍNDICE

	Pág.
2.1. Introducción.....	1
2.2. Fuentes de datos.....	1
2.3. La observación.....	1
2.3.1. Clases de observación.....	1
2.3.2. Ventajas.....	2
2.3.3. Desventajas.....	2
2.4. La entrevista.....	3
2.4.1. Tipos de entrevista.....	3
2.4.1.1. Entrevista no estructurada.....	3
2.4.1.2. Entrevista estructurada.....	3
2.4.2. Ventajas.....	3
2.4.3. Desventajas.....	3
2.5. El cuestionario.....	4
2.5.1. Ventajas.....	4
2.5.2. Desventajas.....	4
2.6. La experimentación.....	4
2.6.1. Experimento determinístico.....	5
2.6.2. Experimento aleatorio.....	5

ORGANIZACIÓN Y PRESENTACIÓN DE DATOS

3.1 Introducción.- Una vez efectuada la recolección de los datos y con el objeto de proporcionar utilidad a la información obtenida, es necesario efectuar la organización de dichos datos para posteriormente presentarlos en forma tabular y en forma gráfica.

3.2 Caracteres de una población o muestra.- Se refieren a los rasgos distintivos que tiene los elementos que forman la población y la muestra. Estos rasgos característicos pueden ser de dos tipos: cualitativo y cuantitativo, por lo que esta distinción da lugar a la comprensión del atributo y de la variable.

3.2.1 Atributo.- Son los rangos característicos o propiedades cualitativas de un conjunto de elementos. Los atributos se designan con las primeras letras del alfabeto (A, B, C).

Los elementos distintos de un atributo se denominan modalidades y se diferencia con un subíndice en las letras utilizadas (a_i , b_j , c_k).

Existen dos clases de atributos:

3.2.1.1 Atributo nominal.- Se refiere al atributo cuyas modalidades son susceptibles de ordenamiento.

3.2.1.2 Atributo ordinal.- Se refiere a aquel atributo cuyas modalidades presentan jerarquía o capacidad de ordenamiento.

3.2.2 Variable.- Son los rasgos característicos o propiedades cuantificables de los elementos de una población o una muestra. Una variable se designa con las últimas letras del alfabeto (X, Y, Z).

Las magnitudes concretas de una variable se denominan valores y se diferencia con un subíndice en las letras utilizadas (x_i , y_j , z_k).

Existen dos clases de variables:

3.2.2.1 Variable discreta.- Es aquella variable que no permite que entre dos valores consecutivos pueda incluirse algún valor intermedio.

3.2.2.2 Variable continua.- Es aquella variable que permite que entre dos valores consecutivos pueda incluirse un valor intermedio.

3.3 Distribución de frecuencias.- Los datos obtenidos mediante cualquiera de las técnicas anteriormente descritas deben ser sometidos a un tratamiento estadístico, construyendo un cuadro en el cual se presentan las frecuencias de repetición de cada modalidad del atributo o valor de la variable.

Las partes esenciales de un cuadro de Distribución de frecuencias son:

- a) Número. Es el código de identificación del cuadro. Este número se escribe a continuación de la palabra "Cuadro".
- b) Título. Es la identificación que preside al cuadro y es colocado en la parte superior del mismo. Debe reunir dos condiciones:
 - Debe ser completo, es decir, debe especificar a qué se refieren los datos presentados en el cuadro, a qué periodo de tiempo y el lugar al cual se refiere la información.
 - Debe ser conciso, es decir, el título debe ser breve sin perder la claridad necesaria.
- c) Matriz de datos. Es una matriz que contiene la información y consta de un conjunto de casillas o celdas, dispuestas en columnas y filas.
- d) Notas explicativas. Estas notas contienen habitualmente la fuente de los datos y cualquier aclaración sobre el contenido del cuadro.

3.4 Distribución de frecuencias de un atributo.- Para el caso de atributos nominales u ordinales, una distribución de frecuencias se construye de la siguiente manera:

3.4.1 Construcción de un cuadro de Distribución de Frecuencias.- Para construir la distribución de frecuencias de un atributo se deben seguir las siguientes reglas:

- En la primera columna se nombran las distintas modalidades del atributo.
- En la segunda columna se calcula la *frecuencia absoluta* (n_i), es decir, el número de veces que se repite la modalidad del atributo.
- En la tercera columna se calcula la *frecuencia relativa* (h_i) que puede ser expresada en forma de proporción o porcentaje, mediante la ecuación (3.1).

$$h_i = \frac{n_i}{n} \quad (3.1)$$

En el cuadro (3.1) se muestra el modelo de una Distribución de Frecuencias de un Atributo.

CUADRO (3.1)
DISTRIBUCION DE FRECUENCIAS DEL ATRIBUTO A

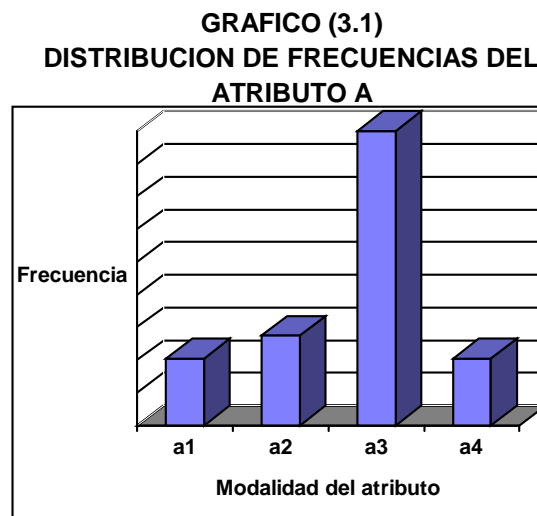
MODALIDADES DEL ATRIBUTO A	FRECUENCIA ABSOLUTA (n_i)	FRECUENCIA RELATIVA (h_i)
a_1	n_1	h_1
a_2	n_2	h_2
a_3	n_3	h_3
...
a_n	n_n	h_n
TOTAL		

Fuente:

3.4.2 Representación gráfica de una distribución de frecuencias de un atributo.- La representación gráfica de una distribución de frecuencias de un atributo, ya sea nominal u ordinal, se efectúa mediante un gráfico de barras, una gráfica de pastel o un gráfico rectangular, de tal manera que pueda apreciarse la información contenida en un golpe de vista.

3.4.2.1 Gráfico de Barras.- Consiste en presentar las frecuencias que corresponden a las diferentes modalidades de un atributo mediante barras, rectángulos o paralelepípedos, los cuales pueden dibujarse horizontal o verticalmente, siendo la longitud de las barras igual a la frecuencia absoluta o a la frecuencia relativa.

El gráfico de barras se representa en un sistema de ejes coordenados, en el cual las modalidades del atributo se representan en el eje de las abscisas y las frecuencias absolutas o relativas se representan en el eje de las ordenadas, tal como se muestra en el Gráfico (3.1).



3.4.2.2 Gráfica de pastel.- Esta gráfica es especialmente apropiada para ilustrar divisiones de una cantidad total, de tal forma que puedan efectuarse comparaciones de una serie de modalidades del atributo, comparada con el total.

En este caso, las modalidades del atributo corresponden a cada segmento del pastel y se representan mediante la ecuación (3.2).

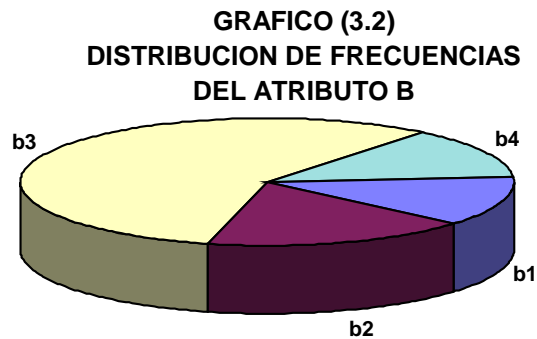
$$p_i = 360^\circ * h_i \quad (3.2)$$

En la que:

p_i = porción del pastel.

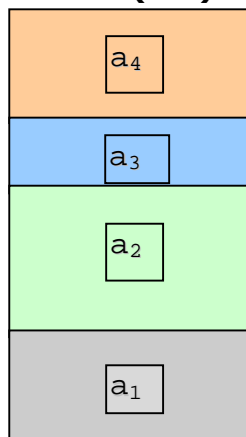
h_i = frecuencia relativa en proporción.

En el gráfico (3.2) se ilustra una gráfica de pastel típico.



3.4.2.3. Gráfica rectangular.- Consiste en representar las modalidades del atributo mediante segmentos rectangulares. Para tal efecto se construye un rectángulo de cualquier base y con altura que sea igual al total de elementos de la población o muestra. El segmento rectangular correspondiente a cada modalidad se determina tomando los valores de la frecuencia absoluta o relativa dimensionada en la altura del rectángulo, tal como se muestra en el Gráfico (3.3).

GRAFICO (3.3)



3.5 Distribución de frecuencias de una variable.- La información recolectada mediante la observación, la entrevista, el cuestionario o cualquier otra técnica, es sometida a un tratamiento estadístico teniendo en cuenta la cantidad de observaciones y la cantidad de valores diferentes de la variable, dando lugar a las llamadas distribuciones de frecuencias de variables ó funciones de frecuencia.

Existen dos tipos de distribuciones de frecuencia:

3.5.1 Distribución Tipo I.- En general, esta distribución se utiliza para presentar variables discretas. Es aquella distribución que se aplica cuando se han obtenido pocos valores diferentes de la variable. En este caso la información se dispone tomando en cuenta las siguientes normas:

- En la primera columna se escriben los valores de la variable.
- En la segunda columna se anota el número de veces que aparece cada valor de la variable. Este número se denomina *frecuencia absoluta* (n_i).
- En la tercera columna se registran las proporciones o porcentajes de aparición de cada valor diferente de la variable o lo que es lo mismo la *frecuencia relativa* (h_i).
- En la cuarta columna se registran los valores acumulados, en orden ascendente o descendente, de la frecuencia absoluta. El conjunto conforma la *frecuencia acumulada absoluta* (N_i).
- En la quinta columna se registran los valores acumulados de modo ascendente o descendente de frecuencia relativa, es decir, la *frecuencia acumulada relativa* (H_i).

El modelo de presentación para este tipo de distribuciones se muestra en el cuadro (3.2).

CUADRO (3.2)
DISTRIBUCION DE FRECUENCIAS DE LA VARIABLE X

VALORES DE LA VARIABLE X	FRECUENCIA ABSOLUTA (n_i)	FRECUENCIA RELATIVA (h_i)	FRECUENCIA ACUMULADA ABSOLUTA (N_i)	FRECUENCIA ACUMULADA RELATIVA (H_i)
x_1	n_1	h_1	N_1	H_1
x_2	n_2	h_2	N_2	H_2
x_3	n_3	h_3	N_3	H_3
...
x_n	n_n	h_n	N_n	H_n
TOTAL				

Fuente:

Existen dos formas tradicionales de presentar gráficamente los datos de una Distribución Tipo I. En ambos casos se utiliza un sistema de ejes coordenados, en el cual se utilizan las abscisas para presentar los valores de la variable y en las ordenadas las frecuencias observadas.

- **Gráfico de barras.-** El gráfico de barras sirve para representar frecuencias absolutas y relativas mediante barras cuyas alturas corresponden a las frecuencias de cada valor de variable observada.
- **Gráfico acumulativo de frecuencias.-** En este caso para cada valor de la variable se levantan ordenadas del tamaño de la frecuencia acumulada absoluta o relativa respectiva. Los brazos continuos en forma de gradas conforman el gráfico acumulativo de frecuencias. Esa línea discontinua es una función empírica que se simboliza con la ecuación (3.3).

$$F(x) = \begin{cases} = 0 & x < x_1 \\ = N_i \text{ ó } H_i & x_1 \leq x \leq x_n \\ = n & x > x_n \end{cases} \quad (3.3)$$

3.5.2 Distribución de frecuencias de Tipo II.- Este tipo de tratamiento estadístico se utiliza preferentemente para variables continuas o también cuando se han efectuado muchas observaciones y se han obtenido muchos valores diferentes de la variable.

Para construir una distribución de frecuencias de Tipo II se siguen los siguientes pasos:

- Se determina el número de estratos o clases, el cual puede ser fijado arbitrariamente (según necesidades de estudio o investigación) o mediante la ecuación (3.4).

$$e = \sqrt{n} \quad (3.4)$$

En la que:

e = número de estratos

n = cantidad total de observaciones

- Se calcula el *recorrido* o *rango* de la variable, siendo ésta la diferencia del mayor valor observado y el menor valor observado, es decir:

$$r = X_{\max} - X_{\min} \quad (3.5)$$

- Se determina la *longitud del intervalo de clase*. Esta distancia puede ser constante o no. En el caso de tener intervalos de clase de longitud constante, ésta se calcula dividiendo el recorrido de la variable entre el número de estratos.

$$c_i = \frac{r}{e} \quad (3.6)$$

Según la ecuación (3.6), los intervalos de clase se encuentran en función de los estratos, de tal modo que permitan la ordenación de los datos en forma exhaustiva y tal que sean mutuamente excluyentes.

- En la primera columna, se colocan los intervalos de clase ($L_{i-1} - L_i$).

A este efecto es imprescindible que en cada estrato o clase se diferencie:

- * El límite inferior (L_{i-1}).
 - * El límite superior (L_i), valor hasta el cual se puedan incluir los valores a clasificar.
- En la segunda columna se calcula la *marca de clase* que es el valor central de cada intervalo de clase y se designa mediante x_i , calculándose con la ecuación (3.7).

$$x_i = \frac{(L_{i-1} + L_i)}{2} \quad (3.7)$$

En este caso la marca de clase sirve para transformar una Distribución Tipo II en una Distribución de Tipo I y de esta manera efectuar operaciones para la determinación de indicadores estadísticos.

- En la tercera columna se calcula la *frecuencia absoluta* (n_i).
- En la cuarta columna se registran los valores de la *frecuencia relativa* (h_i).
- En la quinta columna se obtiene la *frecuencia acumulada absoluta* (N_i).
- En la sexta columna se registran los valores de la *frecuencia acumulada relativa* (H_i).

En el cuadro (3.3), se muestra el formato de una Distribución de Frecuencias de Tipo II.

CUADRO (3.3)
DISTRIBUCIÓN DE FRECUENCIAS DE LA VARIABLE X

INTERVALOS DE CLASE ($L_{i-1} - L_i$)	MARCA DE CLASE (X_i)	FRECUENCIA ABSOLUTA (n_i)	FRECUENCIA RELATIVA (h_i)	FRECUENCIA ACUMULADA ABSOLUTA (N_i)	FRECUENCIA ACUMULADA RELATIVA (H_i)
$L_0 - L_1$	X_1	n_1	h_1	N_1	H_1
$L_1 - L_2$	X_2	n_2	h_2	N_2	H_2
$L_2 - L_3$	X_3	n_3	h_3	N_3	H_3
....
$L_{n-1} - L_n$	X_n	n_n	h_n	N_n	H_n
TOTAL					

Fuente:

La Distribución de Tipo II implica pérdida de información, debido a que los valores observados se tabulan en intervalos de clase, a diferencia, de la Distribución Tipo I donde la información se tabula tal como es, sin ninguna pérdida.

La Distribución Tipo II puede representarse gráficamente mediante:

- **Histograma.-** En un sistema de ejes coordenados, empleando rectángulos cuyas alturas representan los valores de las frecuencias, se utiliza el eje de las abscisas para representar los intervalos de clase ($L_{i-1} - L_i$) y en el eje de ordenadas se representan los valores de las frecuencias absolutas o relativas.

En el histograma de frecuencias, si los puntos medios de los rectángulos (marcas de clase) se unen mediante una línea se forma el polígono de frecuencias.

- **Polígono acumulativo de frecuencias u ojiva.-** En un sistema de ejes coordenados se utilizan las abscisas para representar los valores de los intervalos de clase y en el eje de ordenadas se representan las frecuencias acumuladas absolutas o relativas. Para obtener el polígono, se une los puntos límites de los intervalos de clase correspondientes a las ordenadas cuyo valor es de la frecuencia acumulada.

BIBLIOGRAFÍA:

(1) MOYA Rufino (1988): "*Estadística Descriptiva*". Perú.

=====

ÍNDICE

	Pág.
3.1. Introducción.....	1
3.2. Caracteres de una población o muestra.....	1
3.2.1. Atributo.....	1
3.2.1.1. Atributo nominal.....	1
3.2.1.2. Atributo ordinal.....	1
3.2.2. Variable.....	1
3.2.2.1. Variable discreta.....	2
3.2.2.2. Variable continua.....	2
3.3. Distribución de Frecuencias.....	2
3.4. Distribución de frecuencias de un atributo.....	3
3.4.1. Construcción de un cuadro de Distribución de Frecuencias.....	3
3.4.2. Representación gráfica de una distribución de frecuencias de un atributo.....	4
3.4.2.1. Gráfica de barras.....	4
3.4.2.2. Gráfica de pastel.....	4
3.4.2.3. Gráfica rectangular.....	5
3.5. Distribución de frecuencias de una variable.....	6
3.5.1. Distribución Tipo I.....	6
3.5.2. Distribución Tipo II.....	6

MEDIDAS DE POSICIÓN

4.1 Medidas descriptivas.- Una vez efectuada la organización y presentación de los datos mediante la Distribución de frecuencias y su representación gráfica, es necesario calcular ciertos valores que permitan resumir y transmitir los principales rasgos o características de la información recolectada.

Estos valores se cuantifican a través de ciertas medidas, entre las cuales se tiene: las medidas de posición, las medidas de dispersión, las medidas de asimetría, las medidas de curtosis, etc..

4.2 Medidas de posición.- En este caso, las medidas de posición son aquellas que describen la posición que ocupa la distribución de frecuencias respecto a un valor de la variable o modalidad del atributo.

Las medidas de posición más conocidas son: moda, mediana, media aritmética, media cuadrática, media armónica, media geométrica, cuartiles, deciles y percentiles.

4.3 La moda.- La moda es una medida de posición que corresponde al valor de la variable o modalidad del atributo determinada por la frecuencia (absoluta o relativa) con mayor valor. La moda se denota frecuentemente por M_o , $M_o(x)$, etc..

En el caso de que una distribución de frecuencia tuviera una sola moda se denomina unimodal, si tiene 2 modas se llama bimodal y si tiene más de tres modas se denomina multimodal.

4.3.1 Cálculo de la moda.- Para la obtención de la moda se consideran dos procedimientos:

4.3.1.1 Primer procedimiento.- Se aplica en el caso de estar presentes frente a una Distribución de Tipo I o una Distribución de frecuencias de un atributo, siendo la moda el valor de la variable o la modalidad del atributo correspondiente a la frecuencia con mayor valor.

4.3.1.2 Segundo procedimiento.- Se aplica cuando los datos se han agrupado en una Distribución Tipo II y se utiliza la ecuación (4.1).

$$Mo = L_{i-1} + \frac{d_1 c_i}{d_1 + d_2} \quad (4.1)$$

En la que:

L_{i-1} = Límite inferior correspondiente al intervalo con la frecuencia absoluta con mayor valor.

c_i = Longitud del intervalo de la clase modal.

$d_1 = n_i - n_{i-1}$

$d_2 = n_i - n_{i+1}$

n_i = frecuencia absoluta correspondiente a la clase modal.

4.3.2 Ventajas de la moda.- Las principales ventajas de la moda son:

- La moda se puede utilizar como una medida de posición para datos cualitativos como cuantitativos.
- La moda no está afectada por los valores extremos.

4.3.3 Desventajas de la moda.- Las principales desventajas de la moda son:

- En muchos casos, no existe un valor modal, puesto que el conjunto de datos no contiene valores que se repiten más de una vez.
- Cuando el conjunto de observaciones contiene dos, tres o más modas, éstas son difíciles de interpretar y comparar.

4.4 La mediana.- La mediana se define como el valor de la variable tal que si se ordenan los valores en forma creciente o decreciente, divide en dos partes iguales la distribución. La mediana frecuentemente se denota por: Me , $Me(x)$, μ , etc..

4.4.1 Cálculo de la mediana.- En el cálculo de la mediana es posible distinguir tres procedimientos.

4.4.1.1 Primer procedimiento.- Se utiliza en caso de que los valores de la variable no se encuentren agrupados. En este caso se aplica el concepto de la mediana para determinar su valor, considerando los siguientes pasos:

- Se ordenan los valores no agrupados en forma ascendente o descendente.
- Si el número de datos es impar el valor de la mediana corresponde al valor de la variable que ocupa la posición central.
- Si el número de datos es par el valor de la mediana es el promedio aritmético de los valores centrales de la variable.

4.4.1.2 Segundo procedimiento.- Se aplica para datos agrupados en distribuciones Tipo I y para su cálculo se siguen los siguientes pasos:

- Se determina la cantidad media del universo, es decir: $n/2$.
- Se ubica el valor $n/2$ entre dos valores consecutivos de la frecuencia acumulada absoluta, tal que el límite superior sea mayor o igual a $n/2$.

$$N_{i-1} < \frac{n}{2} \leq N_i$$

En tal caso se presentan dos situaciones:

- Si $N_i > n/2$, entonces el valor de la mediana es igual al valor de la correspondiente x_i , es decir:

$$Me = x_i \quad (4.2)$$

- Si $N_i = n/2$, entonces el valor de la mediana es el promedio de los valores centrales de la variable, es decir:

$$Me = \frac{X_i + X_{i+1}}{2} \quad (4.3)$$

4.4.1.3 Tercer procedimiento.- Se utiliza cuando se manejan distribuciones Tipo II y se siguen los pasos que a continuación se detallan:

- Se determina la cantidad media del universo, es decir: $n/2$.
- Se ubica el valor $n/2$ entre dos valores consecutivos de la frecuencia acumulada absoluta, tal que el límite superior sea mayor o igual a $n/2$.

$$N_{i-1} < \frac{n}{2} \leq N_i$$

En tal caso se pueden presentar dos casos:

- Si $N_i > n/2$ entonces el valor de la mediana se calcula con la ecuación (4.4).

$$Me = L_{i-1} + \frac{c_i \left(\frac{n}{2} - N_{i-1} \right)}{N_i - N_{i-1}} \quad (4.4)$$

- Si $N_i = n/2$ entonces el valor de la mediana corresponde al límite inferior del intervalo al cual pertenece el valor de N_i , es decir:

$$Me = L_{i-1} \quad (4.5)$$

4.4.2 Ventajas de la mediana.- Las ventajas de utilizar la mediana son:

- La mediana es fácil de entender y puede ser calculada en cualquier clase de datos.
- La mediana es afectada por el número de observaciones y no por la presencia de valores extremos.

4.3.3 Desventajas de la mediana.- Las desventajas de emplear la mediana son:

- Se deben organizar los datos antes de realizar cualquier tipo de cálculo. Esto consume mucho tiempo para un conjunto de datos con muchos elementos.

- La mediana no es adecuada para efectuar manipulaciones algebraicas posteriores.

4.5 La media aritmética.- Es una de las medidas de posición más conocidas y utilizadas en el campo de la Estadística y se define como el valor del centro de gravedad de la distribución. La media aritmética actúa como punto de equilibrio o balanceo del conjunto de valores, de modo que las observaciones que son menores se equilibran con las mayores. En forma simbólica se expresa como: μ , μ_x , $M(x)$, etc..

4.5.1 Cálculo de la media aritmética.- Dependiendo del tipo de Distribución en el que se presentan los datos, se tiene:

4.5.1.1 Primer procedimiento.- Se aplica en datos no agrupados y se determina como la suma de los valores de la variable dividida entre el número de ellos, en otras palabras, empleando la ecuación (4.6).

$$\mu = \sum \frac{x_i}{n} \quad (4.6)$$

4.5.1.2 Segundo procedimiento.- Se utiliza cuando se tiene una distribución Tipo I y se obtiene aplicando cualquiera de las siguientes ecuaciones:

$$\mu = \sum \frac{x_i n_i}{n} \quad (4.7)$$

$$\mu = \sum x_i h_i \quad (4.8)$$

4.5.1.3 Tercer procedimiento.- Se aplica en Distribuciones Tipo II y se obtiene aplicando la expresión utilizada para Distribuciones Tipo I, pero obteniendo previamente, las marcas de clase que vendrían a conformar los valores de x_i en la ecuación.

Cabe resaltar que la media aritmética que se obtiene utilizando las ecuaciones (4.7) y (4.8) es un valor aproximado al valor de la media aritmética que se obtendría si se utilizara la ecuación (4.6), puesto que al ser la marca de clase un valor representativo del intervalo analizado, no toma en cuenta los valores que se encuentran en dicho intervalo.

4.5.2 Propiedades de la media aritmética.- La media aritmética como medida de posición goza de las siguientes propiedades:

- La media aritmética de una constante es la constante misma.

$$M(k) = k \quad (4.9)$$

- La media aritmética de una variable multiplicada por una constante es igual a la constante por la media de la variable.

$$M(k) = kM(x) \quad (4.10)$$

- La media aritmética de una variable más/menos una constante es igual a la media aritmética del variable más/menos la constante.

$$M(k \pm x) = k \pm M(x) \quad (4.11)$$

- Si una variable x es particionable mediante “ r ” variables, es decir:

$$x = x_1 \pm x_2 \pm x_3 \pm \dots \pm x_r \quad \forall x \text{ var. independientes}$$

entonces se cumple:

$$\begin{aligned} M(x) &= M(x_1 \pm x_2 \pm x_3 \pm \dots \pm x_r) \\ M(x) &= M(x_1) \pm M(x_2) \pm M(x_3) \pm \dots \pm M(x_r) \end{aligned} \quad (4.12)$$

4.5.3 Ventajas de la media aritmética.- La media aritmética tiene las siguientes ventajas:

- Es un concepto conocido e intuitivo a la mayoría de las personas.
- Es una medida que es única, puesto que cada conjunto de datos tiene una y sólo una media aritmética.
- Para su cálculo se toma en cuenta todas las observaciones del conjunto de datos.

4.5.4. Desventajas de la media aritmética.- El uso de la media aritmética presenta las siguientes desventajas:

- Puede estar afectada por valores extremos que no son representativos del resto de las observaciones.
- El cálculo de la media aritmética es tedioso por que se utilizan todas las observaciones para efectuar los cálculos.
- No es posible calcular la media aritmética para un conjunto de datos que tiene intervalos de clase abiertos en los extremos.

4.6 Otras medidas de posición.

4.6.1 La media armónica.- La media armónica es una medida que se utiliza preferentemente cuando los datos tienen unidades que se encuentran divididas en unidades de tiempo, es decir, sustituye a la media aritmética como indicador de posición cuando los datos se refieren a determinadas tasas de utilización.

La media armónica se calcula con la ecuación (4.13).

$$M. H. = \frac{1}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} \quad (4.13)$$

4.6.2 La media geométrica.- La media geométrica es una medida que se emplea en casos en los que la variable se presenta en porcentaje, especialmente en tasas de crecimiento de tipo poblacional (crecimientos geométricos).

La media geométrica se calcula con la ecuación (4.14).

$$M. G. = \sqrt[n]{x_1 x_2 \dots x_n} \quad (4.14)$$

4.6.3 Cuantiles.- Los cuantiles permiten dividir a los datos en otras proporciones, siendo un concepto extendido de la mediana. Los cuantiles más usados en el análisis estadístico son: cuartiles, deciles y percentiles.

4.6.3.1 Cuartiles.- Los cuartiles son valores que dividen a un conjunto de datos ordenados en forma ascendente o descendente en cuatro partes iguales.

4.6.3.2 Deciles.- Los deciles son valores que dividen a un conjunto de datos ordenados en forma ascendente o descendente en diez partes iguales.

4.6.3.3 Percentiles.- Los percentiles son valores que dividen a un conjunto de datos ordenados en forma ascendente o descendente en cien partes iguales.

BIBLIOGRAFÍA:

(1) MOYA Rufino (1988): “*Estadística Descriptiva*”. Perú.

=====

ÍNDICE

	Pág.
4.1. Medidas descriptivas.....	1
4.2. Medidas de posición.....	1
4.3. La moda.....	1
4.3.1. Cálculo de la moda.....	1
4.3.1.1.Primer procedimiento.....	1
4.3.1.2.Segundo procedimiento.....	2
4.3.2. Ventajas de la moda.....	2
4.3.3. Desventajas de la moda.....	2
4.4. La mediana.....	3
4.4.1. Cálculo de la mediana.....	3
4.4.1.1.Primer procedimiento.....	3
4.4.1.2.Segundo procedimiento.....	3
4.4.1.3.Tercer procedimiento.....	4
4.4.2. Ventajas de la mediana.....	4
4.4.3. Desventajas de la mediana.....	4
4.5. La media aritmética.....	5
4.5.1. Cálculo de la media aritmética.....	5
4.5.1.1.Primer procedimiento.....	5
4.5.1.2.Segundo procedimiento.....	5
4.4.1.3 Tercer procedimiento.....	5
4.5.2. Propiedades de la media aritmética.....	6
4.5.3. Ventajas de la media aritmética.....	6
4.5.4. Desventajas de la media aritmética.....	7
4.6. Otras medidas de posición.....	7
4.6.1. La media armónica.....	7
4.6.2. La media geométrica.....	7
4.6.3. Cuantiles.....	7
4.6.3.1. Cuartiles.....	8
4.6.3.2. Deciles.....	8
4.6.3.3. Percentiles.....	8

MEDIDAS DE DISPERSIÓN, COMPARACIÓN Y ASIMETRÍA

5.1 Medidas de dispersión.- Las medidas de dispersión son aquellas que permiten determinar el grado en que los datos tienden a esparcirse en referencia a una medida de posición (generalmente la media aritmética).

Las principales medidas de dispersión son:

- El recorrido de la variable.
- La desviación media absoluta.
- La varianza.
- La desviación standard.

En términos generales, si cualquiera de las medidas anteriores adopta valores grandes, se dice que existe una dispersión elevada y, por el contrario, si el valor es pequeño, la dispersión también es pequeña.

5.1.1 El recorrido de la variable.- Es la medida más sencilla y de cálculo más fácil puesto que proporciona una indicación global de la dispersión.

El cálculo del recorrido de la variable se efectúa mediante la diferencia del valor máximo y el valor mínimo de los valores observados para la variable, es decir:

$$r = X_{\max} - X_{\min} \quad (5.1)$$

En ciertas ocasiones el recorrido ofrece una medida errónea de la dispersión, debido a que para su cálculo sólo se emplean los valores extremos y se ignora la naturaleza de la variación entre todas las demás observaciones, además está altamente influenciado por los valores extremos.

5.1.2. Desviación media absoluta.- Es la media aritmética en valor absoluto de las desviaciones de los valores observados de la variable respecto de la media aritmética. Se calcula mediante las ecuaciones (5.2) y (5.3).

- Para datos no agrupados:

$$DMA = \sum \frac{|x_i - \mu|}{n} \quad (5.2)$$

- Para datos agrupados.

$$DMA = \sum \frac{|x_i - \mu|n_i}{n} \quad (5.3)$$

La desviación media absoluta está expresada en las mismas unidades de la variable y se puede utilizar como medida de dispersión en todas aquellas distribuciones en las que la medida de tendencia central más significativa es la media aritmética.

La desviación media absoluta presenta las siguientes ventajas:

- En su cálculo se utiliza todas las observaciones.
- Es menos sensible a la presencia de valores extremos respecto al recorrido de la variable.

La desviación media absoluta presenta las siguientes desventajas:

- Su manejo algebraico es complicado.
- Su interpretación teórica es relativamente difícil.

5.1.3. Varianza.- La varianza se define como la media aritmética del cuadrado de las desviaciones de los valores de la variable respecto de la media aritmética.

La varianza se representa mediante: σ^2 , $V(x)$, $VAR(x)$, σ_x^2 , etc..

5.1.3.1 Cálculo de la varianza.- Dependiendo de si los datos se encuentran agrupados o no, la varianza se calcula mediante dos procedimientos:

5.1.3.1.1 Para datos no agrupados.- Cuando la cantidad de datos es pequeña, para su cálculo se aplica su misma definición.

$$\sigma^2 = \sum \frac{(x_i - \mu)^2}{n} \quad (5.4)$$

Si la cantidad de datos es relativamente numerosa es mejor aplicar la ecuación (5.5).

$$\sigma^2 = \sum \frac{x_i^2}{n} - \mu^2 \quad (5.5)$$

5.1.3.1.2 Para datos agrupados.- Este caso se presenta cuando los datos se agrupan en una Distribución Tipo I o cuando los datos de una Distribución Tipo II se reducen a una Distribución Tipo I. La varianza se calcula aplicando la definición de la siguiente manera:

$$\sigma^2 = \sum \frac{(x_i - \mu)^2 n_i}{n} \quad (5.6)$$

$$\sigma^2 = \sum (x_i - \mu)^2 h_i \quad (5.7)$$

El desarrollo de la ecuación (5.6) permite el cálculo de la varianza de un modo operativo más simple, dando origen a la ecuación (5.8).

$$\sigma^2 = \sum \frac{x_i^2 n_i}{n} - \mu^2 \quad (5.8)$$

5.1.3.2 Propiedades de la varianza.- La varianza goza de las siguientes propiedades:

- La varianza de un conjunto de datos siempre es un número no negativo.

$$V(x) \geq 0 \quad (5.9)$$

- La varianza de una constante es 0.

$$V(k) = 0 \quad (5.10)$$

- La varianza de una variable más/menos una constante es igual a la varianza de la variable.

$$V(x \pm k) = V(x) \quad (5.11)$$

- La varianza de la variable multiplicada por una constante es igual al cuadrado de la constante por la varianza de la variable.

$$V(kx) = k^2 V(x) \quad (5.12)$$

5.1.4 Desviación standard.- En vista de que las unidades de la varianza están elevadas al cuadrado, esto representa un inconveniente en cuanto a la interpretación de esta cantidad. En la práctica se utiliza otra medida basada en el valor de la varianza que sirve para dar una medida de la dispersión en la misma unidad en la que están los datos. Este indicador es la desviación standard o desviación típica.

La desviación standard es la raíz cuadrada positiva de la varianza y está expresada en las mismas unidades de la variable. Se representa por σ y se calcula mediante cualquiera de las siguientes ecuaciones:

- Para datos no agrupados:

$$\sigma = \sqrt{\sum \frac{(x_i - \mu)^2}{n}} \quad (5.13)$$

$$\sigma = \sqrt{\sum \frac{x_i^2}{n} - \mu^2} \quad (5.14)$$

- Para datos agrupados:

$$\sigma = \sqrt{\sum \frac{(x_i - \mu)^2 n_i}{n}} \quad (5.15)$$

$$\sigma = \sqrt{\sum \frac{x_i^2 n_i}{n} - \mu^2} \quad (5.16)$$

5.2 Medidas de comparación.- Para efectuar comparaciones entre dos distribuciones de frecuencias se buscan indicadores que no estén expresados en la escala de la medida de la variable y que manejen números abstractos.

Estas medidas son llamadas también medidas de dispersión relativa y las más utilizadas son:

- Coeficiente de variación.
- Variable standarizada.

5.2.1 Coeficiente de variación.- Es la relación que se establece entre la desviación standard y la media aritmética, es decir, expresa la desviación standard existente en la distribución por unidad de media aritmética.

El coeficiente de variación se calcula con la ecuación (5.17).

$$C.V. = \frac{\sigma}{\mu} \quad (5.17)$$

El coeficiente de variación es útil cuando se comparan la variabilidad de dos ó más conjuntos de datos que difieren de modo considerable en las magnitudes de las observaciones.

5.2.2 Variable estandarizada.- Permite conocer el número de desviaciones standard que una observación en particular ocupa por encima o por debajo de la media aritmética. Se calcula con la ecuación (5.18).

$$z_i = \frac{(x_i - \mu)}{\sigma} \quad (5.18)$$

Si el signo de "z" es positivo significa que la variable analizada (x_i) es mayor que la media aritmética y si es negativo es menor que la media aritmética.

5.3 Medidas de asimetría.- En forma general, el gráfico de una distribución de frecuencias se relaciona con un eje vertical o eje de simetría, el cual atraviesa el punto más alto de la distribución. En términos de este eje de simetría, la representación gráfica de una distribución se califica como simétrica o asimétrica.

El concepto de asimetría de una distribución indica la deformación horizontal de las distribuciones de frecuencia.

Una gráfica de distribución de frecuencias se dice que es **simétrica** cuando los valores de posición: media, mediana y moda, tienen el mismo valor, es decir, las dos partes divididas mediante el eje de simetría son iguales.

Cuando la gráfica de una distribución de frecuencias tiene "cola", "falda" o "rama" estirada a la derecha o izquierda, se dice que es **sesgada** o **asimétrica** a la derecha o a la izquierda, respectivamente. En este caso la mediana, la moda y la media tienen valores no coincidentes, es decir:

- Asimetría positiva o derecha: $Mo \leq Me \leq \mu$
- Asimetría negativa o izquierda: $\mu \leq Me \leq Mo$

Una medida que cuantifica la asimetría es el momento de tercer orden respecto a la media aritmética, que se obtiene mediante las ecuaciones (5.19) y (5.20).

- Para datos no agrupados:

$$m_3 = \sum \frac{(x_i - \mu)^3}{n} \quad (5.19)$$

- Para datos agrupados.

$$m_3 = \sum \frac{(x_i - \mu)^3 n_i}{n} \quad (5.20)$$

Si:

- $m_3 = 0$, la distribución es simétrica.
- $m_3 < 0$, la distribución es asimétrica a la izquierda.
- $m_3 > 0$, la distribución es asimétrica a la derecha.

BIBLIOGRAFÍA:

(1) MOYA Rufino (1988): "*Estadística Descriptiva*". Perú.

=====

ÍNDICE

	Pág.
5.1. Medidas de dispersión.....	1
5.1.1. El recorrido de la variable.....	1
5.1.2. Desviación media absoluta.....	1
5.1.3. Varianza.....	2
5.1.3.1.Cálculo de la varianza.....	2
5.1.3.1.1.Para datos no agrupados.....	2
5.1.3.1.2.Para datos agrupados.....	3
5.1.3.2.Propiedades de la varianza.....	3
5.1.4. Desviación standard.....	4
5.2. Medidas de comparación.....	4
5.2.1. Coeficiente de variación.....	5
5.2.2. Variable standarizada.....	5
5.3. Medidas de asimetría.....	5

**UNIVERSIDAD MAYOR DE SAN SIMÓN
FACULTAD DE CIENCIAS Y TECNOLOGÍA
DEPARTAMENTO DE MATEMÁTICAS**

ESTADÍSTICA I

CAPÍTULO V

**"MEDIDAS DE DISPERSIÓN, COMPARACIÓN
Y ASIMETRÍA"**

SEMESTRE: I/2004

Docente: Ing. Roberto Manchego C.

Cochabamba, Febrero de 2004

CÁLCULO DE PROBABILIDADES

6.1 Antecedentes.- La probabilidad es un concepto utilizado con bastante frecuencia en el cotidiano desenvolvimiento de nuestras actividades puesto que cada uno debe tomar decisiones y en la mayor parte de los casos existe incertidumbre respecto a los resultados posibles.

Con más razón, en el desarrollo de las actividades profesionales, al momento de tomar alguna decisión administrativa, técnica, económica, etc., el valor de la probabilidad proporciona una herramienta que sirve de apoyo a la decisión más adecuada.

6.2 Espacio muestral.- Es el conjunto de resultados o sucesos posibles e imaginables del experimento aleatorio.

Un espacio muestral puede expresarse mediante las modalidades que permiten expresar un conjunto, es decir, por extensión y comprensión.

- Por extensión, implica describir y enumerar los distintos sucesos.
- Por comprensión, implica incluir los sucesos en un concepto.

6.2.1 Sucesos de un espacio muestral.- El elemento que compone el espacio muestral se denominan suceso y puede clasificarse en simple y compuesto.

- Un suceso es **simple** cuando no es posible su descomposición en otros sucesos. Por ejemplo, en el lanzamiento de un dado, la aparición de cualquiera de los números en la cara superior, es un suceso simple.
- Un suceso es **compuesto** cuando es posible descomponer el suceso en uno o más sucesos simples. Por ejemplo, al lanzar dos dados, un suceso compuesto ocurre al estar interesados en la suma de los números que salen en las caras superiores.

6.2.2 Evento.- El evento constituye un subconjunto del espacio muestral. Está conformado por los resultados o sucesos que son de interés para el cálculo de probabilidades.

Ya que el espacio muestral y el evento son conjuntos, es posible aplicar las operaciones de conjuntos, las leyes y propiedades relacionadas con estos. Los eventos y espacios muestrales y, en particular las relaciones entre eventos son a menudo representados por medio de Diagramas de Venn.

6.3 Definición de probabilidad.- En general, la probabilidad es la medida de la incertidumbre. La probabilidad es la posibilidad de que suceda un evento particular. Esta medida se expresa mediante un número que se asocia al evento o conjunto de sucesos aleatorios que expresan una situación de incertidumbre.

6.3.1 Notación y terminología.- Las principales formas de anotar y sus correspondientes términos son:

- $p(E)$ = probabilidad de que ocurra el evento E .
- $p(E_1 \text{ o } E_2) = p(E_1 \cup E_2)$ = probabilidad de que ocurra el evento E_1 o el evento E_2
- $p(E_1 / E_2)$ = probabilidad de que ocurra el evento E_1 dado que ocurrió el evento E_2
- $p(E_1 E_2) = p(E_1 \cap E_2)$ = probabilidad de que ocurra el evento E_1 y el evento E_2

Se hace notar que la notación y terminología anterior se hace extensible a más de 2 eventos.

6.3.2 Axiomas de probabilidad.- Sea un experimento cualquiera, para el que se define: S el espacio muestral correspondiente, E un evento cualesquiera y $p(E)$ la probabilidad de que ocurra E , entonces $p(E)$ debe cumplir los siguientes axiomas:

1º La probabilidad es un número no negativo y menor que la unidad, es decir:

$$0 \leq p(E) \leq 1$$

2º La probabilidad de un evento cierto es igual a la unidad, es decir:

$$p(E) = 1 \text{ si } E \text{ es cierto}$$

3° Si E es un evento cualquiera y \bar{E} es un evento contrario, la probabilidad del evento contrario es igual a la unidad menos la probabilidad del evento E , es decir:

$$p(E) = 1 - p(\bar{E}) \quad (6.1)$$

6.3.3. Tipos de cálculo de probabilidades.- El concepto de probabilidad ha sido desarrollado de distintas maneras, lo que ha dado lugar a dos formas o definiciones, los cuales son: la definición subjetiva y la definición objetiva.

6.3.3.1 Definición subjetiva.- Esta definición se basa en el grado de creencia que tiene la persona que calcula la probabilidad, respecto a la ocurrencia de un determinado evento.

El grado de creencia es un número que difiere según la evidencia que posee la persona que efectúa la asignación.

6.3.3.2 Definición objetiva.- En este caso, la situación difiere de la anterior en vista si dos ó más personas deciden calcular la probabilidad de un evento particular, partiendo de los mismos datos y utilizando las mismas normas de cálculo deben llegar a los mismos resultados.

En lo que resta del tema se detalla el cálculo objetivo de probabilidad con todas sus reglas y teoremas.

6.4 Técnicas de conteo.- En ocasiones puede ser difícil o al menos tedioso, determinar el número de elementos de un espacio muestral por medio de la enumeración directa. Para manejar sistemáticamente este problema existen las siguientes técnicas de conteo:

a) **Enumeración explícita.-** Es la técnica más básica de conteo, que consiste en enumerar uno por uno, todos los elementos del espacio muestral.

- b) **Cuadro o matriz de doble entrada.**- Se utiliza cuando se desea conocer la cantidad y los elementos de un suceso compuesto de 2 sucesos simples.
- c) **Diagrama del árbol.**- Se emplea cuando se tiene un experimento aleatorio con sucesos compuestos de 2 ó más sucesos simples.
- d) **Permutación.**- Una permutación es un arreglo de objetos distintos, en la que una permutación difiere de otra si el orden del arreglo o el contenido difieren. Si se desea seleccionar permutaciones de r objetos a partir de n objetos distintos, el número de resultados está dado por la ecuación (6.2).

$$P_r^n = \frac{n!}{(n-r)!} \quad (6.2)$$

- e) **Combinación.**- Una combinación es un arreglo de objetos distintos, en la que una combinación difiere de otra, sólo si difiere el contenido del arreglo. Si se desea determinar el número de combinaciones cuando en n objetos distintos deben seleccionarse r a la vez, se utiliza la ecuación (6.3).

$$C_r^n = \frac{n!}{r!(n-r)!} \quad (6.3)$$

6.5 Definición clásica.- Esta definición se aplica para el caso de eventos con sucesos simples o cuando se tiene un evento con sucesos compuestos, cada uno de ellos conformado de sucesos simples que tienen la misma posibilidad de ocurrir.

Se define de la siguiente manera: "La probabilidad de un evento cualquiera es el número que se obtiene al dividir el número de casos favorables al evento entre el número de casos posibles", es decir:

$$p(E) = \frac{\text{número de casos favorables al evento (E)}}{\text{número de casos posibles}} \quad (6.4)$$

Ahora bien, asociada a esta situación existe el caso de utilizar las frecuencias de una distribución de frecuencias, por lo que la probabilidad de un evento es igual a la frecuencia relativa observada de la aparición de dicho evento, siempre y cuando el número de experimentos sea un valor muy grande, es decir:

$$h(E) = \frac{n(E)}{n} \quad (6.5)$$

$$p(E) \approx h(E) \quad \forall n \rightarrow \infty \quad (6.6)$$

La definición frecuencial de probabilidad supone:

- Gran cantidad de ensayos.
- Acepta el principio de regularidad estadística que consiste en que al repetir un fenómeno aleatorio de manera indefinida se puede observar que existe creciente estabilidad en la aparición de un suceso.

6.6 Regla de la adición.- La regla de la adición se utiliza cuando se desea determinar la probabilidad de que ocurra un evento u otro. La probabilidad de que ocurra el evento E_1 ó el evento E_2 , puede representarse en el lenguaje de la teoría de conjuntos como la unión del conjunto E_1 con el conjunto E_2 , y designarse mediante $p(E_1 \cup E_2)$.

En forma general, la probabilidad de que ocurra E_1 o E_2 es la suma de la probabilidad de que ocurra E_1 más la probabilidad de que ocurra E_2 menos la probabilidad conjunta de que ocurra E_1 y E_2 , es decir:

$$p(E_1 \text{ o } E_2) = p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 E_2) \quad (6.7)$$

La probabilidad $p(E_1 E_2)$, como se dijo, se denomina probabilidad conjunta de la ocurrencia de dos eventos y es la intersección de E_1 y E_2 designándose también como $p(E_1 \cap E_2)$.

En el caso de que E_1 y E_2 sean eventos mutuamente excluyentes:

$$p(E_1 \cap E_2) = 0 \quad (6.8)$$

por tanto, la ecuación (6.7) se convierte en:

$$p(E_1 \cup E_2) = p(E_1 \cap E_2) + p(E_1) + p(E_2) \quad (6.9)$$

6.7 Probabilidad condicional.- Frecuentemente, existe interés en obtener la probabilidad de un evento, el cual está condicionado a la ocurrencia de otro evento del espacio muestral. Es evidente, también que ello ocurre cuando ambos eventos son dependientes entre sí, es decir, cuando la ocurrencia o no ocurrencia de uno afecta la probabilidad de ocurrencia del otro evento.

En estas situaciones se utiliza el concepto de probabilidad condicional para designar la probabilidad de ocurrencia de un evento relacionado. Por tanto, si E_1 y E_2 son eventos cualesquiera que se encuentran en un espacio muestral S , la probabilidad condicional de E_2 , con respecto a E_1 es:

$$p\left(\frac{E_2}{E_1}\right) = \frac{p(E_1 \cap E_2)}{p(E_1)} \quad p(E_1) > 0 \quad (6.10)$$

es decir, es el cociente de la probabilidad conjunta de E_1 y E_2 con respecto a la probabilidad de E_1 .

6.8 Regla del producto.- La regla del producto se refiere a la determinación de la probabilidad de la ocurrencia de los eventos de E_1 y E_2 , es decir, $p(E_1 \cap E_2)$, siempre y cuando existan diferentes espacios muestrales y ocurra simultaneidad de sucesos y, los sucesos sean compuesto de sucesos simples con diferentes probabilidades.

6.8.1 Independencia estadística.- En el caso de que la ocurrencia de E_1 no tenga ningún efecto sobre la ocurrencia del evento de E_2 , en el sentido de que la probabilidad condicional $p(E_2 / E_1)$ es igual a la probabilidad $p(E_2)$ aún a pesar de que haya ocurrido el evento E_1 , se origina un concepto muy importante que se conoce como independencia estadística, la cual se define de la siguiente manera:

Sean E_1 y E_2 dos eventos cualesquiera de un espacio muestral S , de tal forma que es posible calcular:

$$p\left(\frac{E_2}{E_1}\right) = \frac{p(E_1 \cap E_2)}{p(E_1)} \quad (6.11)$$

se dice que el evento E_1 es estadísticamente independiente del evento E_2 si:

$$p\left(\frac{E_2}{E_1}\right) = p(E_2) \quad (6.12)$$

lo cual implica que:

$$p(E_1 \cap E_2) = p(E_1) * p(E_2) \quad (6.13)$$

Con cierta frecuencia se confunde la aparición de sucesos mutuamente excluyentes y no excluyentes, por un lado, y los conceptos de independencia y dependencia por el otro. La mutua exclusividad indica que dos sucesos no pueden ocurrir al mismo tiempo, en tanto que la independencia señala que la probabilidad de ocurrencia de un evento no es afectada por la ocurrencia del otro.

6.8.2 Dependencia estadística.- Al considerar la probabilidad condicional de algún evento E_2 , dada la ocurrencia de otro evento E_1 , siempre se ha implicado que las probabilidades de E_1 y E_2 , son dependientes entre sí. En otras palabras, la información con respecto a la ocurrencia de E_2 afectará la probabilidad de E_1 .

Para este caso la regla del producto se deduce de la definición de probabilidad condicional:

$$p\left(\frac{E_2}{E_1}\right) = \frac{p(E_1 \cap E_2)}{p(E_1)} \quad (6.14)$$

Despejando de la ecuación (6.14) se obtiene:

$$p(E_1 \cap E_2) = p(E_1) * p\left(\frac{E_2}{E_1}\right) \quad (6.15)$$

ecuación que representa la llamada regla del producto, es decir, la probabilidad de que ocurran los eventos E_1 y E_2 es igual a la probabilidad de la ocurrencia de uno de ellos multiplicado por la probabilidad condicional de que ocurra el segundo, dado que ocurrió el primero.

6.9 Probabilidad total.- En caso de tener eventos secuenciales, es posible calcular la probabilidad de que ocurra un evento, combinando la regla de la suma y la regla del producto, es decir:

Sean $B_1, B_2, B_3, \dots, B_n$, n eventos mutuamente excluyentes del espacio muestral S y A un evento arbitrario en S , entonces para $r = 1, 2, \dots, n$:

$$p(A) = \sum_{i=1}^r p(B_i) * p\left(\frac{A}{B_i}\right) \quad (6.16)$$

y se lee "probabilidad total de que ocurra A ".

6.10 Teorema de Bayes.- En su forma algebraica más simple, el Teorema de Bayes se refiere al cálculo de la probabilidad de un evento específico B_i , dado que ocurrió el evento A , es decir:

$$p\left(\frac{B_i}{A}\right) = \frac{p(B_i) * p\left(\frac{A}{B_i}\right)}{p(A)} \quad \forall i = 1, 2, 3, \dots, r \quad (6.17)$$

A primera vista, no es más que una aplicación de la probabilidad condicional. Sin embargo, la importancia especial del Teorema de Bayes consiste en que se aplica en el contexto de eventos secuenciales y además, proporciona la base para determinar la probabilidad condicional de un evento que ha ocurrido en la primera

posición secuencial, dado que se ha observado un evento específico en la segunda posición secuencial.

El Teorema de Bayes ofrece también el fundamento para obtener la llamada probabilidad "condicional hacia atrás" o "*a priori*", puesto que puede determinarse la probabilidad de que se haya observado un evento determinado en una primera etapa, dada la observación de otro evento en una segunda etapa.

BIBLIOGRAFÍA:

- (1) **LEVIN** Richard (1996): "*Estadística para Administración y Economía*". México
 (2) **MOYA** Rufino (1988): "*Probabilidad e Inferencia Estadística*". Perú.

=====

ÍNDICE

	<u>Pág.</u>
6.1. Antecedentes.....	1
6.2. Espacio muestral.....	1
6.2.1. Sucesos de un espacio muestral.....	1
6.2.2. Evento.....	2
6.3. Definición de probabilidad.....	2
6.3.1. Notación y terminología.....	2
6.3.2. Axiomas de probabilidad.....	2
6.3.3. Tipos de cálculo de probabilidad.....	3
6.3.3.1. Definición subjetiva.....	3
6.3.3.2. Definición objetiva.....	3
6.4. Técnicas de conteo.....	3
6.5. Definición clásica.....	4
6.6. Regla de la adición.....	5
6.7. Probabilidad condicional.....	6
6.8. Regla del producto.....	6
6.8.1. Independencia estadística.....	6
6.8.2. Dependencia estadística.....	7
6.9. Probabilidad total.....	8
6.10. Teorema de Bayes.....	8

DISTRIBUCIONES DE PROBABILIDAD

7.1 Variable aleatoria.- Una variable aleatoria “ x ” es aquella función de valor que permite asignar puntos y elementos del espacio muestral en la serie correspondiente a los números reales.

Se dice que “ x ” es “aleatoria” por que involucra la probabilidad de los resultados del espacio muestral.

7.1.1 Variable aleatoria discreta.- Una variable aleatoria “ x ” se denomina discreta si existe correspondencia con la serie de los números naturales.

Una variable aleatoria discreta se identifica cuando entre dos valores consecutivos de la variable no puede existir un valor intermedio.

7.1.2 Variable aleatoria continua.- Se dice que “ x ” es una variable aleatoria continua cuando ésta toma un continuo de valores o cuando entre dos valores consecutivos de la variable es posible la aparición de un valor intermedio.

7.2 Distribución de probabilidad de una variable aleatoria.- Una distribución de probabilidad de variable aleatoria 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, según sea el caso.

7.2.1 Distribución de probabilidad de una variable aleatoria discreta.- Es aquel conjunto formado por todos los valores numéricos posibles de una variable aleatoria con sus probabilidades correspondientes, tal como se muestra en el Cuadro (7.1).

Cuadro (7.1)
DISTRIBUCIÓN DE PROBABILIDAD DE LA VARIABLE X

x_i	$p(x_i)$
x_1	$p(x_1)$
x_2	$p(x_2)$
x_3	$p(x_3)$
...	...
x_n	$p(x_n)$

Fuente:

La distribución de probabilidad de una variable aleatoria discreta se denomina también *función de cuantía* y puede expresarse mediante una ecuación matemática o un listado.

Si X es una variable aleatoria discreta y $p(x_i)$ la probabilidad de que x_i tome un valor en particular, todos los valores de $p(x_i)$ deben satisfacer las siguientes condiciones:

a) Cualquier valor de $p(x_i)$, debe ser un número real positivo, es decir:

$$p(x_i) \geq 0 \quad \forall x_i \in N_x$$

b) La suma de todos los valores $p(x_i)$ es 1, es decir:

$$\sum p(x_i) = 1 \quad \forall x_i \in N_x$$

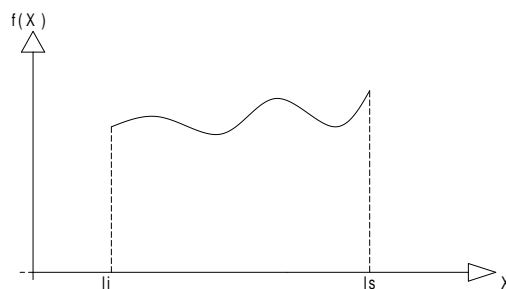
Cuando la distribución de probabilidad de variable aleatoria discreta implica la acumulación de probabilidades hasta un valor determinado de la variable, esa acumulación se denomina **Función de Distribución Acumulada de X**, es decir, representa la probabilidad de que x sea menor o igual a un valor específico de X y se designa mediante la ecuación (7.1).

$$F(x_0) = p(x \leq x_0) = \sum_{i=1}^{x_0} p(x_i) \quad \forall x_i \in N_x \quad (7.1)$$

La función de cuantía y la función de Distribución Acumulada pueden representarse gráficamente mediante:

- Un diagrama de barras, para lo cual en el eje de abscisas se colocan los valores de la variable y en las ordenadas se representan el valor de las probabilidades.
- Un diagrama acumulativo de probabilidades, para lo cual en el eje de las abscisas se colocan los valores de la variable y en las ordenadas se colocan los valores correspondientes a la Función de Distribución Acumulada.

7.2.2 Distribución de probabilidad de una variable aleatoria continua.- Es aquella en la que los valores numéricos posibles de una variable aleatoria se determinan a través de una función matemática y se ilustra en forma gráfica por medio de una curva de probabilidad. La distribución de probabilidad de una variable aleatoria continua se denomina **función de densidad** y se representan generalmente como **$f(x)$** .



Si $f(x)$ existe, debe cumplir dos condiciones:

a) La función $f(x)$ debe proporcionar valores mayores o iguales que 0, es decir:

$$f(x) \geq 0 \quad \forall \text{ l.i} < x < \text{l.s}$$

b) La integral de la función debe ser igual a 1.

$$\int_{\text{l.i}}^{\text{l.s}} f(x) dx = 1$$

En la que:

- l.i = límite inferior
- l.s = límite superior

Es importante indicar que $f(x)$ no representa ninguna probabilidad como tal y que solamente cuando la función se integra entre dos puntos proporciona un valor de probabilidad, es decir:

$$p(a < x < b) = p(a \leq x \leq b) = \int_a^b f(x) dx \quad \forall x \in R_x \quad (7.2)$$

Gráficamente, la probabilidad queda expresada por el área bajo la curva entre los puntos **a** y **b**.

La función de distribución acumulada de **X** para el caso de una variable aleatoria continua se designa por **F(x_o)** y permite determinar la probabilidad acumulada correspondiente a un nivel determinado de variable y se define como:

$$F(x_o) = p(X \leq x_o) = \int_{\text{l.i}}^{x_o} f(x) dx \quad \forall x \in R_x \quad (7.3)$$

La función de distribución acumulada, para que sea tal, debe cumplir dos condiciones:

- $F(x_o = \text{l.i}) = 0$
- $F(x_o = \text{l.s}) = 1$

Cuando la función de distribución acumulada es derivable en el recorrido de la variable, su derivada da origen a la función de densidad, es decir:

$$\frac{dF(x_o)}{dx} = f(x) \quad (7.4)$$

7.3 Medidas de posición y dispersión de una distribución de probabilidad de variable aleatoria.-

7.3.1 Moda.- En caso de que la variable aleatoria fuese discreta, la moda se define como aquel valor de la variable que tiene mayor probabilidad de ocurrencia. Para el caso continuo, la moda de la distribución se calcula aplicando las condiciones que corresponden a la determinación del valor máximo correspondiente, es decir, aplicando la ecuación (7.5).

$$f'(x) = \frac{df(x)}{dx} \quad (7.5)$$

Se verifica si existe moda calculando la segunda derivada, para lo cual debe cumplirse:

$$f''(x) < 0$$

Ahora bien en caso de no poder aplicarse el cálculo diferencial, es posible utilizar el análisis de puntos extremos o análisis de contorno.

7.3.2 Mediana.- La mediana para una función de densidad $f(x)$ se calcula aplicando la definición, es decir, es el valor de la variable que divide la distribución en dos partes iguales. La expresión que permite el cálculo de la mediana se muestra en la ecuación (7.6).

$$\int_{li}^{Me} f(x)dx = \int_{Me}^{ls} f(x)dx = 0.5 \quad (7.6)$$

7.3.3 Valor esperado.- El valor esperado de una variable aleatoria X es el valor promedio después de un número grande de experimentos, llamado también valor a la larga o esperanza matemática; se define según las ecuaciones (7.7) y (7.8).

$$E(x) = \sum x_i p(x_i) \quad \forall x \quad \text{v.a. discreta} \quad (7.7)$$

$$E(x) = \int_{li}^{ls} x f(x) dx \quad \forall x \quad \text{v.a. continua} \quad (7.8)$$

7.3.4 Varianza.- La varianza de una variable aleatoria X se define según la ecuación (7.9).

$$V(x) = \sigma^2 = E(x - \mu)^2 \quad (7.9)$$

Para el caso de tener una variable discreta y otra continua, se tiene las ecuaciones (7.10) y (7.11).

$$\sigma^2 = \sum (x_i - \mu)^2 p(x_i) \quad \forall x \quad \text{v.a. discreta} \quad (7.10)$$

$$\sigma^2 = \int_{li}^{ls} (x - \mu)^2 f(x) dx \quad \forall x \quad \text{v.a. continua} \quad (7.11)$$

Otras formas alternativas de cálculo son las que se muestran en las ecuaciones (7.12) y (7.13).

$$\sigma^2 = \sum x_i^2 p(x_i) - \mu^2 \quad \forall x \quad \text{v.a. discreta} \quad (7.12)$$

$$\sigma^2 = \int_{li}^{ls} x^2 f(x) dx - \mu^2 \quad \forall x \quad \text{v.a. continua} \quad (7.13)$$

BIBLIOGRAFÍA:

(1) **MOYA** Rufino y **SARAVIA** Gregorio (1988): “*Probabilidad e Inferencia Estadística*”. Perú.

=====

INDICE

	Pág.
7.1 Variable aleatoria.....	1
7.1.1 Variable aleatoria discreta.....	1
7.1.2 Variable aleatoria continua.....	1
7.2 Distribución de probabilidad de una variable aleatoria.....	1
7.2.1 Distribución de probabilidad de una variable aleatoria discreta.....	1
7.2.2 Distribución de probabilidad de una variable aleatoria continua.....	3
7.3 Medidas de posición y dispersión de una distribución de probabilidad de variable aleatoria.....	5
7.3.1 Moda.....	5
7.3.2 Mediana.....	5
7.3.3 Valor esperado.....	6
7.3.4 Varianza.....	6

DISTRIBUCIONES TEÓRICAS DE PROBABILIDAD

8.1. Distribuciones de probabilidad de variable aleatoria discreta.- Son aquellas en las que la variable aleatoria adopta valores enteros. Las principales son:

8.1.1. Distribución Bernoulli.- Se dice que un experimento sigue una distribución Bernoulli, si cumple las siguientes condiciones:

- a) Se realiza una única prueba.
- b) Sólo existen dos posibles resultados en cada ensayo, llamados comúnmente: “éxito” y “fracaso”.

La distribución de probabilidad que describe el comportamiento de la variable aleatoria x es:

$$\begin{aligned} p(x) &= p^x q^{1-x} & x = 0, 1 \\ p(x) &= 0 & \text{en otro caso} \end{aligned} \quad (8.1)$$

En la que:

p = probabilidad de éxito ($x = 1$, éxito)

q = probabilidad de fracaso ($x = 0$, fracaso)

El valor esperado y la varianza de la distribución son:

$$\begin{aligned} E(x) &= p & (8.2) \\ V(x) &= pq & (8.3) \end{aligned}$$

Esta función no es muy utilizada, siendo que en realidad se emplea para desarrollar el modelo de la Distribución Binomial.

8.1.2. Distribución Binomial.- Cuando un experimento que sigue una distribución Bernoulli se repite n veces, se dice que se comporta como una distribución Binomial.

La distribución Binomial debe cumplir las siguientes condiciones:

- a) El número de pruebas es fijo e igual a n .
- b) Sólo existen dos posibles resultados en cada ensayo (generalmente denominados “éxito” y “fracaso”).

- c) Las pruebas son independientes unas de otras, lo cual implica que las probabilidades de éxito y fracaso son constantes o que el experimento es con devolución.

De forma general, para deducir la ecuación Binomial, se determina la probabilidad en n ensayos, que ocurran x éxitos consecutivos seguidos de $(n-x)$ fracasos consecutivos. Puesto que los n ensayos son independientes, se tiene:

$$\underbrace{p * p * p * p \dots \dots \dots}_{x \text{ términos}} * \underbrace{q * q * q * \dots \dots \dots}_{(n-x) \text{ términos}} = p^x q^{(n-x)} \quad (8.4)$$

La probabilidad de obtener exactamente x éxitos y $(n-x)$ fracasos en cualquier orden es el producto de $p^x q^{(n-x)}$ por el número de ordenaciones distintas, es decir, el número de combinaciones de n objetos tomados de x a la vez, o lo que es lo mismo:

$$p(x) = \left\{ \begin{array}{ll} = \binom{n}{x} p^x q^{n-x} & \forall x = 0, 1, 2, \dots, n \\ = 0 & \text{en otros casos} \end{array} \right\} \quad (8.5)$$

En la que:

x = número de éxitos
 n = número de pruebas
 p = probabilidad de éxito.
 q = probabilidad de fracaso.

El valor esperado y la varianza son, respectivamente:

$$E(x) = np \quad (8.6)$$

$$V(x) = npq \quad (8.7)$$

Sus áreas de aplicación incluyen inspecciones de calidad, ventas, mercadotecnia, medicina, investigaciones de opinión y otras.

8.1.3. Distribución Poisson.- La distribución Poisson se presenta frecuentemente en dos situaciones:

- a) Cuando un evento o el cambio del estado de un sistema ocurre aleatoriamente sobre el tiempo o espacio (distancias, áreas, volúmenes, pesos, etc.), procesos que se denominan *Procesos Poisson*, por ejemplo:
- Número de automóviles por hora que llegan a un autobanco.
 - Número de repuestos producidos por minuto.
 - Número de bacterias por cm^2 en un cultivo.
- b) Esta distribución constituye una aproximación a la Distribución Binomial cuando $n \rightarrow \infty$ y $p \rightarrow 0$, generalmente se asume cuando: $p \leq 0.10$ y $n \geq 50$, es decir, cuando se cumplen las anteriores condiciones en una distribución Binomial, es conveniente utilizar una Distribución Poisson.

Se dice que una variable aleatoria x sigue una distribución Poisson si tiene una distribución de probabilidad con la ecuación (8.9).

$$p(x) = \begin{cases} = \frac{e^{-\lambda} \lambda^x}{x!} & x = 0, 1, 2, 3, \dots, n \\ = 0 & \text{en otros casos} \end{cases} \quad (8.8)$$

En la que:

$$\lambda = np \quad (8.9)$$

En este caso, el valor esperado y varianza de la variable aleatoria son iguales:

$$E(x) = np \quad (8.10)$$

$$V(x) = np \quad (8.11)$$

La distribución Poisson es el principal modelo de probabilidad empleado para analizar líneas de espera.

8.1.4. Distribución Hipergeométrica.- Se dice que una variable aleatoria x sigue una distribución Hipergeométrica, si cumple las siguientes condiciones:

- a) El número de pruebas es fijo e igual a n .
- b) Los resultados de cada prueba se clasifican en éxito o fracaso.

- c) Las probabilidades de éxito varían en cada prueba puesto que el muestreo es sin reemplazo, lo cual implica que los sucesos son dependientes unos de otros.

De forma general, si en una población de N elementos se extrae sin reemplazo una muestra de tamaño n y se sabe que existen k elementos con cierta característica en la población N . Considerando x el número de elementos de dicha clase en la muestra n y $(n-x)$ el número de elementos que no son de esa clase. El número total de formas de obtener x elementos de k es $\binom{k}{x}$ y las formas de obtener $(n-x)$ elementos de un total $(N-k)$ es igual a $\binom{N-k}{n-x}$. Luego, el número total de formas en que pueden presentarse ambos casos es $\binom{N-k}{n-x} \binom{k}{x}$.

Por otra parte, el número de maneras distintas en que pueden seleccionarse una muestra de n elementos de un total de N es $\binom{N}{n}$ y cada muestra con una probabilidad de selección igual a $\frac{1}{\binom{N}{n}}$. De esta forma la probabilidad de seleccionar x elementos que poseen cierta característica es:

$$p(x) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}} \quad (8.12)$$

Por tanto, la distribución de probabilidad Hipergeométrica es:

$$p(x) = \left\{ \begin{array}{ll} \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}} & x = 0, 1, 2, 3, \dots, n \\ = 0 & \text{en otros casos} \end{array} \right\} \quad (8.13)$$

En la que:

x = variable aleatoria asociada a una de las dos categorías de estudio

N = número total de objetos de la población.

k = número de objetos en la población, de la categoría asociada a la variable aleatoria.

n = tamaño de la muestra.

La esperanza y la varianza de la distribución son:

$$E(x) = \frac{nk}{N} \quad (8.14)$$

$$V(x) = \frac{nk(N-k)(N-n)}{N^2(N-1)} \quad (8.15)$$

La Distribución Hipergeométrica se utiliza con mucha frecuencia en el control de calidad estadístico.

8.2. Distribuciones de probabilidad de variable aleatoria continua.-

8.2.1. Distribución Normal.- La distribución Normal o distribución de Gauss es una de las más importantes y de mayor uso de todas las distribuciones continuas de probabilidad. Es fundamental en la aplicación de la inferencia estadística, ya que las distribuciones de muchas estadísticas muestrales tienden a la distribución Normal conforme crece el tamaño de la muestra.

Se dice que una variable aleatoria x está normalmente distribuida si su función de densidad está dada por:

$$f(x) = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sqrt{2\pi\sigma^2}} \quad -\infty < x < +\infty \quad (8.16)$$

En la que:

μ = media aritmética
 σ^2 = varianza

$-\infty < \mu < +\infty$
 $\sigma^2 > 0$

La gráfica de la distribución Normal es una curva simétrica con forma de campana, que se extiende sin límites tanto en la dirección positiva como en la negativa.

La distribución Normal queda completamente definida una vez que se especifica la media y la varianza, es decir, estos dos parámetros determinan la posición y la forma de la distribución Normal.

La probabilidad de que una variable aleatoria normalmente distribuida sea menor o igual a un valor específico, está dada por la función de distribución acumulada:

$$p(x \leq a) = \int_{-\infty}^a \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sqrt{2\pi\sigma^2}} dx \quad (8.16)$$

La función **f(x)** no es una función simple y su integración no puede realizarse en forma sencilla. Si se tabulara la función de densidad de la distribución Normal, la tabla que se elaboraría sería para un par de valores de μ y σ^2 , es decir, se tendría que elaborar tablas para cada par de valores de μ y σ^2 , tarea virtualmente imposible.

Por otra parte, esta distribución tiene la propiedad de que cualquier transformación lineal de una variable aleatoria normalmente distribuida sigue teniendo la misma distribución, por ejemplo, si **x** tiene una distribución Normal con media μ y varianza σ^2 y **z** = **ax + b**, entonces **z** tendrá una distribución Normal con media **aμ + b** y varianza **a²σ²**.

Por tanto, para reducir el problema anterior y utilizando dicha propiedad, es necesario hacer una transformación que permita presentar los resultados en una sola tabla, dicha transformación es:

$$z = \frac{x - \mu}{\sigma} \quad (8.17)$$

Luego, se origina la llamada Distribución Normal standard:

$$f(z) = \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi\sigma^2}} \quad (8.18)$$

con:

$$E(z) = 0$$

$$V(z) = 1$$

De manera que:

$$p(x \leq a) = p\left(z \leq \frac{a - \mu}{\sigma}\right) = \int_{-\infty}^{\frac{a - \mu}{\sigma}} \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} dz \quad (8.19)$$

Con métodos de cálculo integral, todavía sigue siendo difícil integrar la función de densidad acumulada de la distribución Normal standarizada, sin embargo, por medio del análisis numérico se han obtenido tablas para dicha función.

BIBLIOGRAFÍA:

- (1) **LEVIN** Richard (1996): “*Estadística para Administración y Economía*”. México
- (2) **MOYA** Rufino (1988): “*Probabilidad e Inferencia Estadística*”. Perú.

=====

ÍNDICE

	Pag.
8.1 Distribuciones de probabilidad de variable aleatoria discreta.....	1
8.1.1 Distribución Bernoulli.....	1
8.1.2 Distribución Binomial.....	1
8.1.3 Distribución Poisson.....	2
8.1.4 Distribución Hipergeométrica.....	4
8.2 Distribuciones de probabilidad de variable aleatoria continua.....	5
8.2.1 Distribución Normal.....	5

IX DISTRIBUCIONES BIDIMENSIONALES

9.1 Introducción.- Todo universo ó población es posible caracterizarlo recurriendo a dos ó mas criterios de clasificación, es decir, una población puede ser estudiada, considerando dos ó mas caracteres cualitativos o cuantitativos.

Cuando una población es estudiada por 2 atributos, 2 variables ó un atributo y una variable, se dice que la distribución que permite manejar observaciones y valores es de caracter bidimensional.

9.2 Tipos de Distribución Bidimensional.- Según el número de observaciones y valores o modalidades diferentes, las distribuciones de frecuencia pueden ser de las siguientes clases:

9.2.1 Distribución Bidimensional Tipo I.- Son las distribuciones que corresponden a una situación en que se han tomado pocas observaciones o modalidades diferentes. En este caso no se requiere tratamiento estadístico y la información obtenida se registra en 2 columnas, organizando sus valores en forma ascendente o descendente.

La Distribución Tipo I tiene la siguiente conformación:

DISTRIBUCION TIPO I

X_i	Y_j
x_1	y_1
x_2	y_2
x_3	y_3
..	...
x_r	y_s

9.2.2 Distribución Bidimensional Tipo II.- Se utiliza cuando existen muchas observaciones del par (x_i, y_j) y son pocos los valores distintos de la variable. La información se presenta en un cuadro de doble entrada, tal que los valores de x_i se presentan en las filas y los valores de y_j en las columnas y en el cuerpo se registran las frecuencias absolutas o relativas.

Su disposición tabular para el caso de frecuencias absolutas es de la siguiente forma:

DISTRIBUCION TIPO II

	y_1	y_2	y_3	...	y_s
x_1	n_{11}	n_{12}	n_{13}	...	n_{1s}
x_2	n_{21}	n_{22}	n_{23}	...	n_{2s}
x_3	n_{31}	n_{32}	n_{33}	...	n_{3s}
...
x_r	n_{r1}	n_{r2}	n_{r3}	...	n_{rs}

Donde n_{ij} representan el número de veces que se repite el par (x_i, y_j) .

Por otra parte es posible definir las distribuciones marginales en las cuales solo interesa la información para una de las variables, prescindiendo de la otra.

Las distribuciones marginales se muestran en los siguientes cuadros:

- Para la variable x :

DISTRIBUCION MARGINAL DE X

x_i	$n(x_i)$
x_1	$n(x_1) =$
x_2	$n(x_2) =$
x_3	$n(x_3) =$
....	
x_r	$n(x_r) =$

- Para la variable Y :

DISTRIBUCION MARGINAL DE Y

Y_i	$n(Y_i)$
Y_1	$n(Y_1) =$
Y_2	$n(Y_2) =$
Y_3	$n(Y_3) =$
....	
Y_S	$n(Y_S) =$

9.2.3 Distribución Bidimensional Tipo III.- Se realiza cuando se han efectuado muchas observaciones y obtenido muchos valores distintos de la variable o cuando las observaciones pueden clasificarse como una variable continua.

Para elaborar una Distribución tipo II se procede de la siguiente manera:

- Se define el recorrido de cada una de las variables del par $(x_i; y_j)$ tal que el recorrrido de x e y es:

$$r_x = X_{\max} - X_{\min} \quad (9.1)$$

$$r_y = Y_{\max} - Y_{\min} \quad (9.2)$$

- Se establece el número de clases, estratos o categorías, según las necesidades de la investigación.
- Se determina la amplitud del intervalo de clase, mediante la división del recorrido de la variable por el correspondiente número de clases o estratos.

$$a_x = r_x / N^\circ \text{ de estratos de } x \quad (9.3)$$

$$a_y = r_y / N^\circ \text{ de estratos de } y \quad (9.4)$$

Con los elementos anteriores se construye la Distribución Bidimensional de Tipo III, tal como se muestra en el siguiente cuadro:

DISTRIBUCION TIPO III

	$Y_0 - Y_1$	$Y_1 - Y_2$	$Y_2 - Y_3$...	$Y_{s-1} - Y_s$
$X_0 - X_1$	n_{11}	n_{12}	n_{13}	...	n_{1s}
$X_1 - X_2$	n_{21}	n_{22}	n_{23}	...	n_{2s}
$X_2 - X_3$	n_{31}	n_{32}	n_{33}	...	n_{3s}
...
$X_{r-1} - X_r$	n_{r1}	n_{r2}	n_{r3}	...	n_{rs}

Donde n_{ij} representan el número de veces que se repite el par $(x_{i-1} - x_i, Y_{j-1} - Y_j)$.

También para este caso es posible construir las Distribuciones Marginales de X y Y.

Para obtener los estadígrafos de posición y dispersión necesarios para caracterizar la distribución en forma de indicadores se transforma la Distribución Tipo III en una Distribución Tipo II. Además, con las distribuciones marginales obtenidas de la distribución bidimensional, aplicando las formas de cálculo correspondientes pueden obtenerse los indicadores bidimensionales conocidos y resolver cualquier problema referido a las variable considerada.

9.3 Covarianza.- Es un estadígrafo que permite estudiar la dependencia estadística de las variables contenidas en una distribución bidimensional.

La covarianza se define como la media aritmética del producto de las desviaciones de los valores de cada variable respecto de su media aritmética y se designa con el símbolo σ_{xy} .

La covarianza se calcula mediante las siguientes ecuaciones:

- Para datos no agrupados:

$$\sigma_{xy} = \sum (x_i - \mu_x)(y_i - \mu_y)/n \quad (9.5)$$

$$\sigma_{xy} = \sum x_i y_i / n - (\sum x_i / n)(\sum y_i / n) \quad (9.6)$$

- Para datos agrupados:

$$\sigma_{xy} = \sum (x_i - \mu_x)(y_i - \mu_y)n_i / n \quad (9.7)$$

$$\sigma_{xy} = \sum x_i y_i n_i / n - (\sum x_i n_i / n)(\sum y_i n_i / n) \quad (9.8)$$

9.4 Funciones de probabilidad bidimensional.- Dado un experimento aleatorio E, si el espacio muestral resultante se considera X y Y, como variables aleatorias y las funciones de probabilidad:

$P(x,y)$, para el caso discreto y $f(x,y)$ para el caso continuo es posible desarrollar todo lo referente al cálculo de probabilidades de la siguiente manera:

9.4.1 Funciones de cuantía bidimensional.

9.4.1.1 Condiciones para que una función sea de cuantía.- Al igual que para el caso unidimensional, para que una función sea de cuantía debe cumplir dos condiciones:

a) Las probabilidades deben ser mayores o iguales a cero, es decir.

$$p(x,y) = 0$$

b) La suma de todas las probabilidades debe ser igual a 1, es decir:

$$\sum \sum p(x_i, y_j) = 1$$

9.4.1.2 Cálculo de probabilidades.- Para determinar la probabilidad de que x se encuentre en el par (a_1, b_1) e y en el par (a_2, b_2) , se tiene:

$$p(a_1 \leq x \leq b_1, a_2 \leq y \leq b_2) = \sum \sum p(x,y) \quad (9.9)$$

9.4.1.3 Funciones de Distribución Acumulada.- La Función de Distribución Acumulada se define de la siguiente manera:

$$p(X,Y) = p(0 \leq x \leq X, 0 \leq Y \leq b_2) = \sum \sum p(x_i, y_j) \quad (9.10)$$

9.4.2 Funciones de densidad bidimensional.

9.4.2.1 Condiciones para que una función sea de densidad.- Al igual que para el caso unidimensional, para que una función sea de densidad debe cumplir dos condiciones:

a) Las probabilidades deben ser mayores o iguales a cero, es decir.

$$f(x,y) = 0$$

b) La suma de todas las probabilidades debe ser igual a 1, es decir:

$$\int \int f(x_i, y_j) = 1$$

9.4.2.2 Cálculo de probabilidades.- Para determinar la probabilidad de que x se encuentre entre el par (a_1, b_1) e y entre el par (a_2, b_2) , se tiene:

$$p(a_1 \leq x \leq b_1, a_2 \leq y \leq b_2) = \int \int f(x, y) \quad (9.11)$$

9.4.2.3 Funciones de Distribución Acumulada.- La Función de Distribución Acumulada se define de la siguiente manera:

$$F(X, Y) = p(0 \leq x \leq X, 0 \leq Y \leq b_2) = \int \int f(x_i, y_j) \quad (9.12)$$

9.4.2.4 Funciones de densidad marginal.- Dada la función de densidad para una variable conjunta es posible definir las funciones de densidad marginales, esto significa considerar una variable prescindiendo de la otra.

Dado $f(x, y)$ en el recorrido: $l.i._x < x < l.s_x, l.i_y < y < l.s_y$, se define:

- La función de densidad marginal $f(x)$ a:

$$f(x) = \int f(x, y) dy \quad (9.13)$$

- La función de densidad marginal $f(y)$ a:

$$f(y) = \int f(x, y) dx \quad (9.14)$$

Para calcular probabilidades se procede de la siguiente manera:

$$p(a_1 \leq x \leq b_1) = \int f(x)dx \quad (9.15)$$

$$p(a_2 \leq y \leq b_2) = \int f(y)dy \quad (9.16)$$

9.5 Análisis de regresión y correlación lineal.- El análisis de regresión y de correlación permite la predicción o estimación el valor desconocido de una variable "y" denominada variable dependiente en el supuesto caso de que se conoce el valor de otra variable "x" que se conoce con el nombre de variable independiente, con la cual se relaciona.

Para el análisis de regresión, desde el punto de vista teórico si se dispone de una serie de pares de valores para la variable conjunta (x_i, y_j) es posible encontrar una función matemática o función de ajuste que relacione a ambas variables y que determine la ley de comportamiento que existe entre tales variables.

El proceso de ajustar un conjunto de datos pertenecientes a las variables (x, y) se resume en dos etapas:

- **Primera etapa:** Elaborar el diagrama de dispersión y definir una función matemática que mejor se ajuste a los datos.
- **Segunda etapa:** Establecer un método que permita determinar los valores que asumirán los parámetros de la función de ajuste o de regresión. En este sentido el más apropiado es el de mínimos cuadrados.

Entre las funciones matemáticas más conocidas para relacionar dos variables se tiene:

- Lineal simple: $y = a + bx$
- Lineal inversa: $y = a + b/x$
- Lineal logaritmica: $y = a + b \ln x$
- Exponencial: $y = ab^x$.
- Potencial: $y = ax^b$.

El principal objetivo del análisis de correlación es medir el grado de relación entre la variable independiente y la variable dependiente.

Para efectuar el análisis de correlación se calculan dos coeficientes:

a) Coefficiente de determinación.- El coeficiente de determinación mide la proporción de variabilidad que ha sido estadísticamente explicada, respecto a la variabilidad total, mediante la ecuación de regresión. Es decir:

$$R = VE/VT = 1 - VNE/VT \quad (9.17)$$

Donde:

- $VE = \text{Variabilidad explicada} = \Sigma(Y_c - \bar{y})$
- $VNE = \text{Variabilidad no explicada} = \Sigma(Y_o - Y_c)$
- $VT = \text{Variabilidad total} = \Sigma(Y_o - \bar{y})$

Además:

$$VT = VE + VNE \quad (9.18)$$

Los valores que toma están siempre comprendidos en el intervalo:

$0 \leq R \leq 1$. De manera ideal se desea tener un valor de $R = 1$, puesto que entonces la variabilidad no explicada sería igual a cero, y que toda la variación puede explicarse por la presencia de las variables independientes en la ecuación de regresión.

b) Coefficiente de correlación.- El coeficiente de correlación indica el grado de relación que existe entre las variables independientes con la variable dependiente. Se calcula de la siguiente manera:

$$r = \sqrt{R} \quad (9.19)$$

El valor de r fluctúa entre $0 \leq r \leq 1$, cuando r es igual a 1 la relación es perfecta y cuando el valor de r es igual a cero, se dice que no existe relación entre las variables consideradas.

Para el caso de un modelo lineal simple, el valor r varía entre -1 y 1, siendo el signo de r el mismo que el del coeficiente de la variable independiente.

=====

UNIVERSIDAD MAYOR DE SAN SIMON
FACULTAD DE CIENCIAS Y TECNOLOGIA
LIC. EN INFORMATICA
ING. DE SISTEMAS
ING. INDUSTRIAL

CAPITULO IX

"DISTRIBUCIONES BIDIMENSIONALES"
(MAT-233/MAT-280)

SEMESTRE I/98

Docente: Ing. Roberto Manchego C.

Cochabamba, Julio de 1998

