

UNIDAD 3

ESTADÍSTICA DESCRIPTIVA



OBJETIVO

El estudiante será capaz de recopilar, presentar, describir, analizar y elaborar gráficos e interpretar compilaciones de datos, asimismo podrá presentar conclusiones sobre uno o dos elementos de información que caractericen la totalidad de éstos, con el fin de identificar la medida en que los datos se agrupan o dispersan en torno a un valor central.

TEMARIO

3.1 EXPERIMENTACIÓN

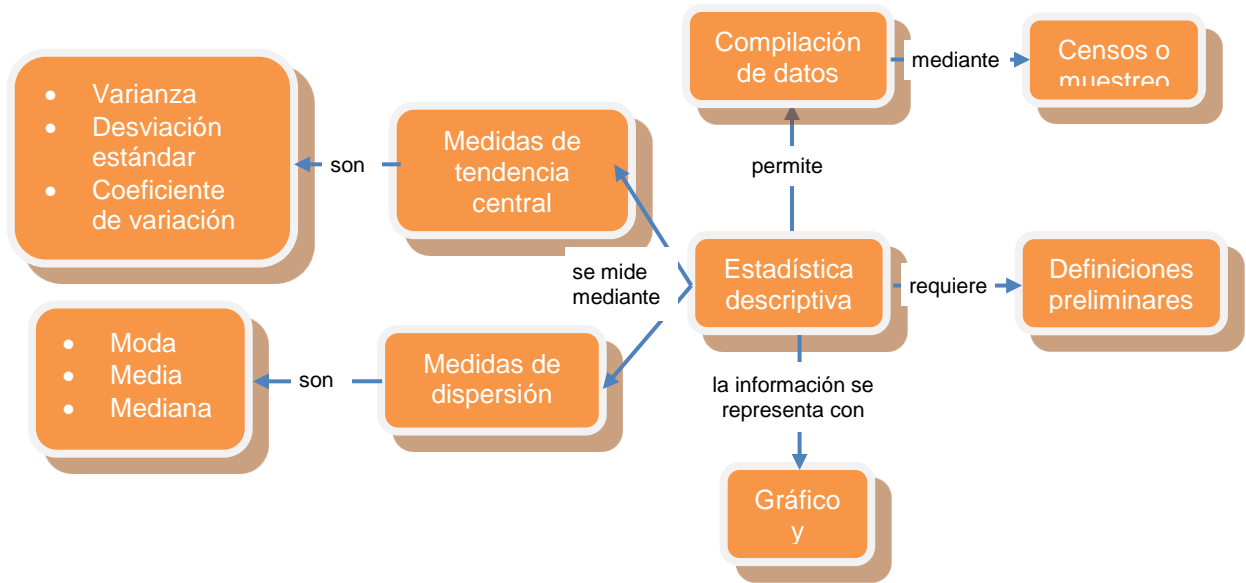
3.2 CONCEPTOS

3.3 PRESENTACIÓN DE DATOS

3.4 MEDIDAS DE TENDENCIA CENTRAL Y DISPERSIÓN PARA DATOS NO AGRUPADOS

3.5 MEDIDAS DE TENDENCIA CENTRAL Y DISPERSIÓN PARA DATOS AGRUPADOS

MAPA CONCEPTUAL



INTRODUCCIÓN

Resulta importante conocer la estadística descriptiva sin importar la profesión que se haya elegido, ya que es una herramienta que permite mostrar a los estudiantes o profesionistas cómo organizar y resumir datos.

La ordenación y la recopilación de datos puede presentar variabilidad o incertidumbre para los estudios encomendados y el solo hecho de conocerlos permite en su momento hacer correcciones ante alguna anomalía, de ahí la importancia de realizar los cálculos de las medidas de tendencia central o dispersión.

Un gran apoyo para la estadística descriptiva es la presentación de los datos por medio de tablas o de gráficas, simplemente porque sintetizan un gran acervo de información y es más sencillo valorar todos los datos de un fenómeno en general con ellas.

3.1 EXPERIMENTACIÓN

Para lograr entender la experimentación en estadística, sobre todo dentro de la descriptiva, es importante mencionar que la probabilidad es la base del estudio de la estadística ya que se interesa por estudiar las muestras de la población.

Para Wayne W. Daniel, estadística⁴ es “la recopilación, organización, y resumen de los datos y la obtención de inferencias acerca del conjunto de datos cuando se observa una parte de ello”; de acuerdo con lo anterior, se puede decir que la estadística es la presentación de los datos pero que parte desde su obtención, recolección e interpretación y la toma de decisiones para la población de acuerdo con el resultado obtenido por la muestra trabajada.

La estadística se divide en dos áreas de estudio, la estadística descriptiva, que se estudia en este apartado, y la inferencial que se estudiará en la siguiente unidad y será definida posteriormente. La descriptiva estudia la parte de la ordenación y la presentación de datos de manera sencilla, por lo tanto, describe las variables con las que se trabaja y para este caso son las del tipo aleatorio, las cuales ya fueron explicadas en la primera unidad.

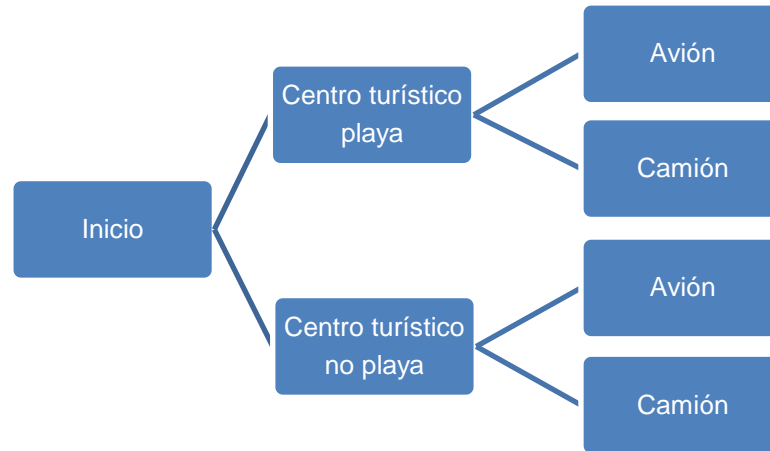
En estadística existen dos tipos de estudio para la casualidad: el experimental y el observacional; en el caso del primero se puede medir el estudio, por tanto hay manipulación en el experimento y en el segundo caso, no existe manipulación en la experimentación.

La experimentación depende del grado de conocimiento que se tiene, gracias a la información obtenida en estadística. Los métodos de experimentación que se emplean son el diagrama de árbol y los principios básicos de conteo, estudiados anteriormente.

En algunas ocasiones, el conteo de los elementos de forma individual se puede realizar con apoyo del diagrama de árbol (éste es la representación gráfica de las posibilidades de un evento). El diagrama de árbol se realiza partiendo de una rama que contiene cada una de las posibilidades, y cabe la posibilidad de crear nuevas ramas.

⁴ Daniel Wayne W., *Bioestadística*, 3a. ed., Limusa, 1993, p. 18.

Ejemplo: suponga que una agencia de viajes ofrece dos destinos turísticos, uno con playa y otro sin playa, y en cada uno de estos destinos se ofrece la trasportación, área o terrestre; entonces el diagrama de árbol se representa de la siguiente manera:



ACTIVIDAD DE APRENDIZAJE

Realizar el diagrama de árbol en los siguientes ejercicios:

- En un restaurante se ofrece un desayuno en tres tiempos, el primer tiempo es jugo o fruta, el segundo tiempo es huevo, molletes o chilaquiles y el tercer tiempo es crepa o helado. Representar el diagrama de árbol respectivo.
- En la rifa de fin de año de una empresa, se sortean 2 regalos, una televisión y un DVD, y los departamentos que entran a la rifa son Contabilidad, con 3 trabajadores; Ventas, con 6 trabajadores; Compras, con 2 trabajadores, y Finanzas, con 4 trabajadores. Representar el diagrama de árbol respectivo.

3.2 CONCEPTOS

Para poder entender la representación de los datos y su cálculo, es necesario esclarecer algunos conceptos.

La *población* es el conjunto de personas, animales o cosas que tienen una característica en común y que son susceptibles de medirse; pero sucede

que existen poblaciones que suelen ser muy grandes y es complicado trabajar con ellas, primero porque implica tener costos elevados y segundo, se requiere una mayor inversión en tiempo para el estudio, o bien es más difícil la observación de los elementos estudiados, es por esta razón que se requiere trabajar con la *muestra*, la cual es un subconjunto de la población que se estudia.

Lo que se estudia en cada muestra son las variables, las cuales son de dos tipos: cualitativas y cuantitativas. Las primeras representan las cualidades de los datos y se clasifican por medio de categorías, por ejemplo, si realizamos un estudio de adiestramiento canino, una variable de tipo cualitativo puede ser la raza del perro. Las variables de tipo cuantitativo son medibles, y se representan de forma numérica, por ejemplo el número de perros que participa en el estudio.

En estadística se trabaja con variables, cuando éstas son cuantificadas dentro de una población, reciben el nombre de *parámetros*, y cuando es dentro una muestra, reciben el nombre de *estimadores*.

Existen diferentes tipos de poblaciones y muestras. La población en estadística puede ser finita o infinita, la primera es la que tiene una cantidad exacta, es decir, una cifra limitada y por tanto los elementos se pueden numerar; la segunda es incuantificable, es decir, no tiene fin y no sabemos exactamente cuántos elementos contiene, por ejemplo, el pelo de un gato o la arena del mar.

El muestreo puede ser de tipo probabilístico y no probabilístico. El primero, hace mención a su nombre, es cuando todos los elementos de la población tienen la misma probabilidad de pertenecer a la muestra; en el segundo, el investigador selecciona los elementos de la muestra de acuerdo con sus criterios.

El muestreo probabilístico puede ser, a su vez, estratificado, simple o de conglomerados. El muestreo estratificado se divide por estratos homogéneos de la población de acuerdo con las mismas características. El muestreo simple consiste en elaborar una lista de la población y mediante números aleatorios se

selecciona cada uno de los elementos de la muestra. Finalmente, el muestreo por conglomerados consiste en elaborar subgrupos homogéneos de elementos de grupos ya existentes o previamente definidos.

El muestreo no probabilístico puede ser por juicios, donde la muestra es elegida por el experto con los elementos más representativos y usualmente se emplea para muestras de tamaño pequeño.

ACTIVIDAD DE APRENDIZAJE

Elaborar un cuadro comparativo de los diferentes tipos de muestreo que incluya tres ejemplos de cada uno.

3.3 PRESENTACIÓN DE DATOS

En estadística descriptiva se trabaja en la obtención, la ordenación y la presentación de datos, buscando que la información sea confiable y se exprese de manera simple y sencilla.

Los datos se pueden presentar por medio de cuadros numéricos o gráficos.

La presentación de datos por cuadros numéricos de la información, a su vez, se hace por representación tabular. Es decir, es presentada en filas y columnas, con título (que es el objeto del cuadro), columna principal, lugar donde se anotan categorías, encabezado de las columnas, explicación del objeto de las columnas, cuerpo donde se anota la información y notas de pie en las que se aclara la información.

Ejemplo: una consultoría que se dedica a realizar estimadores, calculó la siguiente información de la Encuesta Nacional de Empleo: el número de personas pertenecientes a la licenciatura concluida en Economía, en hombres es de 49,755 y en mujeres 16,568; la población económicamente activa en hombres es de 43,874 y en mujeres 15,959 y, finalmente, la población no económicamente activa en hombres es de 5,881 y en mujeres 609, lo cual

resulta en un total de 66,323; 59,833 y 6,490, respectivamente; lo anterior se representa de la siguiente manera:

Población con licenciatura concluida en Economía, según condición de actividad y sexo			
Sexo / Condición de actividad	Población con licenciatura concluida (1) <i>Personas</i>	Población económicamente activa (2) <i>Personas</i>	Población no económicamente activa (3) <i>Personas</i>
Hombres	49,755	43,874	5,881
Mujeres	16,568	15,959	609
Total	66,323	59,833	6,490

Elaboración propia.

La representación gráfica, es la otra alternativa para presentar los datos. Un gráfico en general debe contener el título, claro y completamente definido, además las unidades en que se expresa la medición, información suficiente que permita cotejar la validez de los argumentos y la fuente de la información.

Los principios básicos de un gráfico son simplicidad, fidelidad, ayuda y respuesta del objetivo planteado.

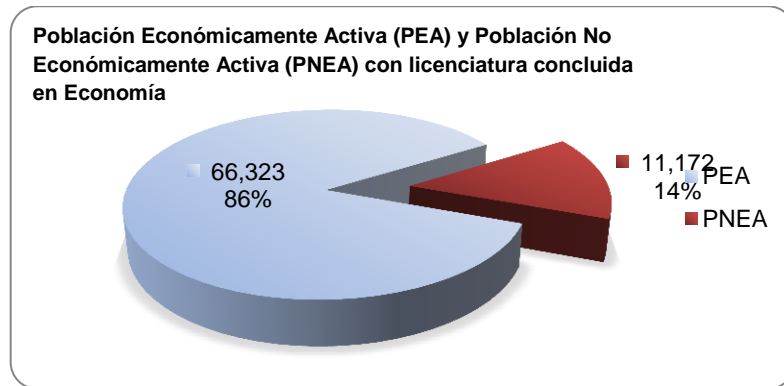
Las reglas de los gráficos incluyen la participación de la población en cada una de las observaciones y la ordenación de éstas.

Existen gráficos que facilitan la representación de los datos dependiendo de las variables de estudio. Para las variables cualitativas son representativas las de pastel, barras simples, barras múltiples, barras compuestas; para las cuantitativas, funcionan muy bien histogramas, polígonos de frecuencia, ojivas y diagramas de caja y bigote.

- a) Gráfica de pastel. Representa las frecuencias relativas (la frecuencia relativa es un determinado valor con respecto al número total de datos) o porcentuales. Dibujar la gráfica de pastel es sencillo, solamente se traza un círculo y se divide según el número de

frecuencias, lo ideal es representar menos de siete datos en este tipo de gráficos.

Ejemplo: una consultoría calculó que la población económicamente activa de estudiantes egresados de la carrera de economía es de 86% y la población no económicamente activa es de 14%:

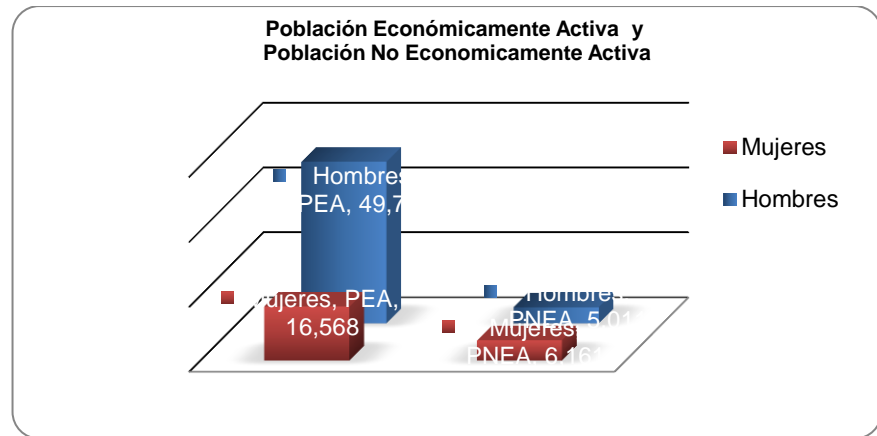


Elaboración propia ⁵

b) Gráfica de barras. Representa datos del tipo cualitativos que se han resumido en una distribución de frecuencias; en el eje horizontal se expresan los indicadores o nombres de cada clase y en el eje vertical la escala de frecuencia.

Ejemplo: una consultoría representó la PEA y la PNEA de hombres y de mujeres en el segundo trimestre del 2010:

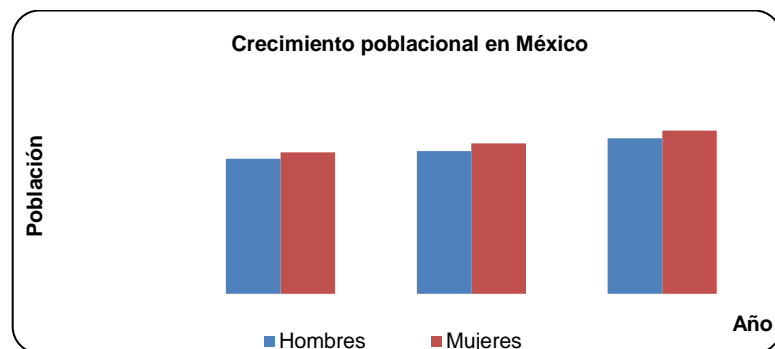
⁵ Fuente: Encuesta Nacional de Ocupación y Empleo, segundo trimestre de 2010.



Elaboración propia.⁶

c) Gráfica de barras múltiples. Sirve para representar frecuencias observadas en clasificaciones dobles, es decir, cuando son más de dos criterios de clasificación porque representan dos variables o más de acuerdo con el número de clases (las barras se colocan juntas de acuerdo con su clasificación).

Ejemplo: una consultoría representó el crecimiento poblacional en México de hombres y mujeres en tres periodos:



Elaboración propia.⁷

d) Gráfica de barras compuestas. Representa frecuencias relativas en forma porcentual, y clasificaciones dobles; se divide de manera

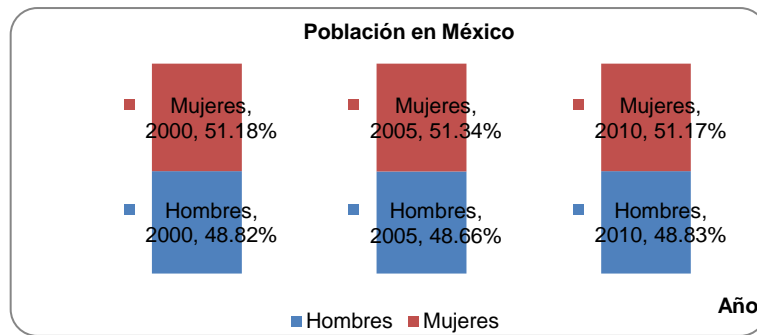
⁶ Fuente: Encuesta Nacional de Ocupación y Empleo, segundo trimestre de 2010.

⁷ Fuente: INEGI.

⁵ Fuente: INEGI.

proporcional respecto de los porcentajes correspondientes a las clases del otro criterio de clasificación.

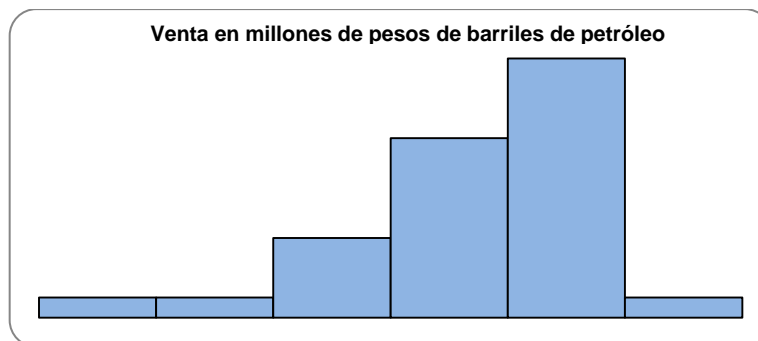
Ejemplo: una consultoría representó la población en México de hombres y mujeres en tres periodos:



Elaboración propia.⁸

e) Histograma. Resume la distribución de frecuencias. En el eje horizontal se representa la variable de interés y en el vertical la distribución de frecuencia. A diferencia del gráfico de barras no hay separación entre los rectángulos formados por las clases y éstos se separan con la línea vertical.

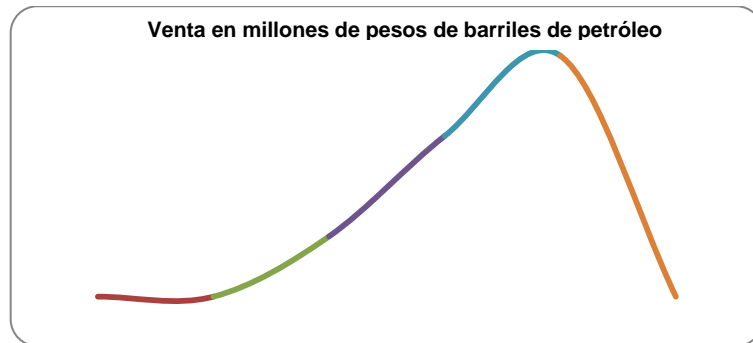
Ejemplo: en la siguiente gráfica se representa la venta en millones de pesos de barriles de petróleo de datos agrupados, tema que se estudiará en los próximos apartados en la parte teórica:



Elaboración propia.

- f) Polígono de frecuencia. Se forma por medio del histograma ya que se plotea el punto medio de la clase; se utiliza para las frecuencias de variables cuantitativas continuas y sólo se utilizan los segmentos de la recta.

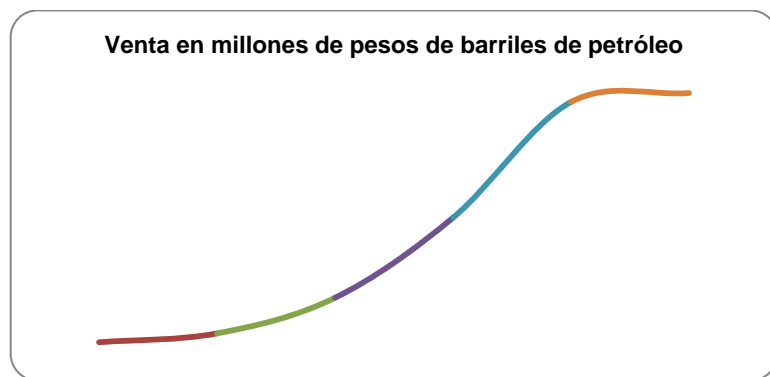
Ejemplo: de la gráfica anterior se forma el polígono de frecuencia:



Elaboración propia.

- g) Ojiva. También se conoce como gráfico de frecuencia acumulada, representa distribuciones de frecuencias de variables continuas acumuladas y para su trazo se utilizan los segmentos de la rectas y se plotea al final de cada clase.

Ejemplo: con base en el histograma se formula la ojiva:



Elaboración propia.

ACTIVIDAD DE APRENDIZAJE

Elaborar un cuadro comparativo en el que se anoten las similitudes y las diferencias de los diferentes tipos de gráficos.

3.4 MEDIDAS DE TENDENCIA CENTRAL Y DISPERSIÓN PARA DATOS NO AGRUPADOS

En el apartado anterior se estudiaron los gráficos y fue posible apreciar que la mayor densidad de frecuencia se encuentra en la parte central de éstos, es decir, de acuerdo con el tipo de gráfico, los valores altos de la gráfica se encuentran en medio y de ahí el interés de conocerlos.

Los valores medios se obtienen con las medidas de tendencia central: media aritmética, media geométrica y media armónica, moda y mediana.

1. Media aritmética. Es la suma de los valores de ciertos números de cantidades dividido entre el número total de elementos; se expresa por medio de la siguiente fórmula:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{N}$$

Ejemplo: el departamento de informática de una empresa da mantenimiento a las computadoras una vez por semana, este servicio por cada una de las semanas es el siguiente: 29, 27, 26, 28 y 30. Calcular el promedio de mantenimiento de computadoras.

Sustituyendo:

$$\begin{aligned}\bar{X} &= \frac{29 + 27 + 26 + 28 + 30}{5} \\ &= \frac{140}{5} \\ &= 28\end{aligned}$$

Respuesta: el promedio de mantenimiento en cinco semanas en la oficina es de 28 computadoras.

2. Media geométrica. Es la raíz de n del producto de n términos, se utiliza para calcular tasas de crecimiento y se expresa con la siguiente fórmula:

$$\text{media geométrica} = \sqrt[n]{X_1 X_2 \dots X_n}$$

Ejemplo: el crecimiento de la población de niños con déficit de atención en los últimos cuatro años es de 8, 17, 21 y 22%, respectivamente. Calcular la media geométrica anual de crecimiento.

Para multiplicar los crecimientos correspondientes se agrega el número uno a cada dato y cuando se tenga el resultado se elimina.

Sustituyendo:

$$\text{media geométrica} = \sqrt[4]{1.08 * 1.17 * 1.21 * 1.22}$$

$$= \sqrt[4]{1.8653}$$

$$= 1.1686$$

$$= 0.1686$$

$$= 0.1686 * 100$$

$$= 16.86\%$$

Respuesta: la tasa de crecimiento en los últimos cuatro años de déficit de atención es de 16.86%.

3. Media armónica. Es el recíproco de la media aritmética de los recíprocos de los números de la serie. Se expresa con la siguiente fórmula:

$$\frac{1}{H} = \frac{\frac{1}{X_1} + \frac{1}{X_2} \dots + \frac{1}{X_n}}{N}$$

Ejemplo: Un psicólogo entrega 3 reportes psicométricos en un día y otro entrega 2. Calcular el rendimiento representativo de los dos psicólogos.

Sustituyendo:

$$\frac{1}{H} = \frac{\frac{1}{3} + \frac{1}{2}}{2}$$

$$\frac{1}{H} = \frac{\frac{2+3}{6}}{2}$$

$$\frac{1}{H} = \frac{\frac{5}{6}}{2}$$

$$\frac{1}{H} = \frac{5}{12}$$

$$H = \frac{12}{5}$$

$$H = 2.4$$

Respuesta: el rendimiento representativo de entrega de reportes es de 2.4.

4. Mediana. Es el valor intermedio que divide un conjunto de datos previamente ordenados de menor a mayor.
5. Moda. Es el valor que más se repite de un conjunto de datos

Ejemplo 1: en un examen extraordinario de la materia de contabilidad se obtuvieron las siguientes calificaciones: 2, 7, 0, 9, 1, 9, 7, 8, 2, 9, 0, 8, 0, 8, 2, 0, 0, 7, 2 ¿cuál es la media y la moda de las calificaciones?

Para calcular la media, primero se ordenan los datos de menor a mayor:

0, 0, 0, 0, 0, 1, 2, 2, 2, 2, 7, 7, 7, 8, 8, 8, 9, 9, 9

El dato marcado divide en partes iguales la serie de datos, de cada lado se tienen 9 valores.

Para calcular la moda, se observa que el valor que más se repite es el número 0, porque aparece 5 veces.

Respuesta: La mediana de las calificaciones del extraordinario de contabilidad es 2, y la moda es 0.

Las medidas de dispersión hacen referencia a la variedad de valores de las observaciones. Si se diera el caso de que todos los valores fueran iguales entonces no existiría dispersión; por tanto, hablar de dispersión significa que los datos no son iguales. Entonces cuando los datos están muy cercanos entre sí existe una varianza pequeña y caso contrario, si están alejados, la varianza es grande.

Las medidas de dispersión que se emplean en el estudio de la estadística, son varianza, desviación estándar y coeficiente de variación.

1. Varianza. Son los valores de observaciones muy cercanos a la media que se calculan para conocer la dispersión de los datos, mediante la siguiente fórmula:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Ejemplo: suponga que el número de ventas de automóviles en una agencia durante los días hábiles de una semana, es 10, 11, 21, 23 y 26. Calcular la varianza de los datos anteriores.

Primero se debe calcular \bar{x} , y recordando la sección anterior de medidas de tendencia central se obtiene:

$$\bar{X} = \frac{10 + 11 + 21 + 23 + 26}{5}$$
$$\bar{X} = 18.2$$

Sustituyendo en la fórmula de la varianza:

$$s^2 = \frac{(10 - 18.2)^2 + (11 - 18.2)^2 + (21 - 18.2)^2 + (23 - 18.2)^2 + (26 - 18.2)^2}{5}$$
$$= \frac{(-8.2)^2 + (-7.2)^2 + (2.8)^2 + (4.8)^2 + (7.8)^2}{5}$$
$$= \frac{67.2 + 51.84 + 7.84 + 23.04 + 60.84}{5}$$
$$= \frac{210.76}{5}$$
$$= 42.152$$

Respuesta: la varianza de los datos representativos de la venta de automóviles es de 42.152.

2. Desviación estándar. Es la raíz cuadrada de la varianza, mide la separación que hay entre los datos y su fórmula es:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

Ejemplo: con base en el ejercicio anterior del número de ventas de automóviles en una agencia de automóviles, calcular la desviación estándar.

De acuerdo con lo desarrollado anteriormente, el resultado de la varianza es 42.152 y a este resultado se le calcula su raíz cuadrada:

$$s = \sqrt{42.152}$$

$$s = 6.4924$$

Respuesta: la desviación estándar de los datos representativos de la venta de automóviles es de 6.4924.

3. Coeficiente de variación. Se utiliza cuando se quiere comparar la dispersión de dos conjuntos de datos porque se desea conocer la variación relativa y se expresa la desviación estándar como un porcentaje de la media, quedando la siguiente fórmula:

$$CV = \frac{s}{\bar{X}}$$

Ejemplo: con base en los ejercicios que se han desarrollado anteriormente, calcular el coeficiente de variación.

De los desarrollos anteriores, el resultado de la media es 18.2 y la desviación estándar es 6.4924, que se sustituyen en la fórmula:

$$\begin{aligned}
 CV &= \frac{6.4924}{18.2} \\
 &= \frac{6.4924}{18.2} \\
 &= 0.3567 \\
 &= 0.3567 * 100 \\
 &= 35.67
 \end{aligned}$$

Respuesta: el coeficiente de variación de los datos representativos de la venta de automóviles es 35.67%.

ACTIVIDAD DE APRENDIZAJE

Determinar las medidas de tendencia central (la media aritmética, moda, mediana) y medidas de dispersión (varianza, desviación estándar y coeficiente de variación), de los siguientes conjuntos de datos:

1. 4, 5, 5, 8, 8, 7, 9.
2. 2, 7, 0, 9, 1, 9, 7, 8.
3. Unos encuestadores realizaron visitas a zonas de alta marginación y recorrieron las siguientes distancias: 5, 6, 9, 18, 5, 9, 11, 3, 9, 9 km, respectivamente. Calcular las medidas de tendencia central y dispersión.

3.5 MEDIDAS DE TENDENCIA CENTRAL Y DISPERSIÓN PARA DATOS AGRUPADOS

En algunas ocasiones se requieren realizar los cálculos de medidas de tendencia central y dispersión en datos que no se encuentran de manera bruta sino que sólo se tiene la distribución de frecuencias. La distribución de

frecuencias es el arreglo ordenado de los datos para sintetizar la información y éstos se presentan sin los valores individuales o las medidas descriptivas.

Cuando los datos se encuentran agrupados, las observaciones individuales pierden su identidad y en la distribución de frecuencias se tiene el número de observaciones que caen dentro de los intervalos de clase; éstos son los que se utilizan para agrupar el conjunto de observaciones y se dan intervalos contiguos de tal manera que cada valor pueda ubicarse en uno sólo.

Las medidas de tendencia central y dispersión de datos agrupados no se calculan de la misma forma que para los datos no agrupados.

Para las medidas de tendencia central se tiene:

Media:

$$\bar{X} = \frac{\sum_{i=1}^k mf}{\sum_{i=1}^k f}$$

Donde m = marca de clase, es el punto medio de cada intervalo de clase y se obtiene $\frac{Li + Lf}{2}$ y f = frecuencia; se presenta en una tabla en que se organizan los datos que caen en cada una de las clases y representa la agrupación de los datos.

Mediana:

$$Me = Li_{me} + \left(\frac{\frac{n}{2} - F_{-1}}{f_{me}} \right) Qm$$

Donde Li_{me} = límite inicial donde se ubica la mediana; n = número total de datos; F_{-1} = frecuencia acumulada (es la cantidad de datos que integra cada una de las clases) anterior de donde se encuentra la mediana; f_{me} = frecuencia donde se ubica la mediana, y Qm = tamaño del intervalo de las clases.

Moda:

$$Mo = Li_{mo} + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) Qm$$

Donde Li_{mo} = limite inicial donde se ubica la moda; Δ_1 y Δ_2 = diferencia de las frecuencias, antes y después, respectivamente, de donde se encuentra la moda.

Para las medidas de dispersión se tiene:

Varianza:

$$s^2 = \frac{\sum_{i=1}^k (m - \bar{x})^2 f}{n}$$

Desviación estándar:

$$s = \sqrt{\frac{\sum_{i=1}^k (m - \bar{x})^2 f}{n}}$$

Coefficiente de variación:

$$CV = \frac{s}{\bar{X}}$$

Ejemplo: en la siguiente tabla se presenta una serie de datos agrupados, calcular las respectivas medidas de tendencia central y dispersión:

Intervalos	Frecuencia (f)	Frecuencia acumulada (F)
21.5 - 26.5	1	1
26.5 - 31.5	1	2
31.5 - 36.5	4	6
36.5 - 41.5	9	15
41.5 - 46.5	13	28

46.5 – 51.5	1	29
-------------	---	----

Para calcular la media conforme la fórmula, se requiere conocer la marca de clase, por tanto en el cuadro se incorpora una nueva columna con el encabezado “marca de clase” m y se realiza el cálculo con $\frac{Li + Lf}{2}$:

Intervalos	Frecuencia (f)	Frecuencia absoluta (F)	Marca de clase (m)
21.5 - 26.5	1	1	24
26.5 - 31.5	1	2	29
31.5 - 36.5	4	6	34
36.5 – 41.5	9	15	39
41.5 – 46.5	13	28	44
46.5 – 51.5	1	29	49

Después se forma otra columna de mf , donde se multiplica la columna m con la columna f y se suman los resultados además de sumar f .

Intervalos	Frecuencia (f)	Frecuencia absoluta (F)	Marca de clase (m)	mf
21.5 - 26.5	1	1	24	24
26.5 - 31.5	1	2	29	29
31.5 - 36.5	4	6	34	136
36.5 – 41.5	9	15	39	351
41.5 – 46.5	13	28	44	572
46.5 – 51.5	1	29	49	49
	$\sum 29$			$\sum 1116$

Posteriormente se sustituyen los valores en la fórmula de la media:

$$\bar{X} = \frac{116}{29}$$

$$\bar{X} = 40.03$$

Respuesta: la media es 40.03.

Ahora se va a calcular la mediana, y se identifica en la fila de color amarillo donde de acuerdo con los 29 datos, el valor intermedio que los divide es 14.5 y el valor más cercano es 15 y se ubica en la columna F ; se realiza la sustitución de la fórmula:

$$Me = 36.5 + \left(\frac{\frac{29}{2} - 6}{9} \right) 5$$

$$= 36.5 + \left(\frac{14.5 - 6}{9} \right) 5$$

$$Me = 41.44$$

Respuesta: la mediana es 41.44.

Para calcular la moda se identifica en la fila de color rosa y la columna f , porque representa cuantos datos se encuentran en el intervalo y como se trata de la moda, son los datos donde hay una mayor cantidad; se realiza la sustitución de la fórmula:

$$Mo = 41.5 + \left(\frac{4}{4 + 12} \right) 5$$

$$Mo = 41.5 + \left(\frac{4}{16} \right) 5$$

$$Mo = 42.75$$

Respuesta: la moda es 42.75.

Para determinar la varianza y antes de sustituir los respectivos valores en la fórmula, se requiere incorporar nuevas columnas para obtener el dato deseado.

Primero se realiza la operación $m - \bar{x}$ y se obtiene:

Intervalos	Frecuencia (<i>f</i>)	Frecuencia absoluta (<i>F</i>)	Marca de clase (<i>m</i>)	<i>mf</i>	$m - \bar{x}$
21.5 - 26.5	1	1	24	24	-16.03
26.5 - 31.5	1	2	29	29	-11.03
31.5 - 36.5	4	6	34	136	-6.03
36.5 - 41.5	9	15	39	351	-1.03
41.5 - 46.5	13	28	44	572	3.97
46.5 - 51.5	1	29	49	49	8.97

En segundo lugar, se realiza la operación $(m - \bar{x})^2$ y se obtiene:

Intervalos	Frecuencia (<i>f</i>)	Frecuencia absoluta (<i>F</i>)	Marca de clase (<i>m</i>)	<i>mf</i>	$m - \bar{x}$	$(m - \bar{x})^2$
21.5 - 26.5	1	1	24	24	-16.03	256.96
26.5 - 31.5	1	2	29	29	-11.03	121.66
31.5 - 36.5	4	6	34	136	-6.03	36.36
36.5 - 41.5	9	15	39	351	-1.03	1.06
41.5 - 46.5	13	28	44	572	3.97	15.76
46.5 - 51.5	1	29	49	49	8.97	80.46

En tercer lugar, se realiza la operación $(m - \bar{x})^2 f$ y se obtiene:

Intervalos	Frecuencia (<i>f</i>)	Frecuencia absoluta (<i>F</i>)	Marca de clase (<i>m</i>)	<i>mf</i>	$m - \bar{x}$	$(m - \bar{x})^2$	$(m - \bar{x})^2 f$
21.5 - 26.5	1	1	24	24	-16.03	256.96	256.96
26.5 - 31.5	1	2	29	29	-11.03	121.66	121.66
31.5 - 36.5	4	6	34	136	-6.03	36.36	145.44

36.5 – 41.5	9	15	39	351	-1.03	1.06	9.55
41.5 – 46.5	13	28	44	572	3.97	15.76	204.89
46.5 – 51.5	1	29	49	49	8.97	80.46	80.46
							Σ 818.97

Finalmente se sustituyen los valores de la sumatoria y el número total de datos:

$$s^2 = \frac{818.97}{29}$$

$$s^2 = \frac{818.97}{29}$$

$$s^2 = 28.24$$

Respuesta: la varianza es 28.24.

Para calcular la desviación estándar solamente son sustituidos los valores:

$$s = \sqrt{28.24}$$

$$s = 5.31$$

Respuesta: la desviación estándar es 5.31.

Y finalmente para calcular el coeficiente de determinación se sustituyen los respectivos valores en la fórmula:

$$CV = \frac{5.31}{40.03}$$

$$= 0.1326$$

$$= 0.1326 * 100$$

$$CV = 13.26\%$$

Respuesta: el coeficiente de variación es 13.26%.

ACTIVIDAD DE APRENDIZAJE

Calcular las medidas de tendencia central y dispersión de los siguientes datos agrupados:

1.

Intervalos de clase	<i>f</i>	<i>F</i>
3.5 – 4.5	3	3
4.5 - 5.5	5	8
5.5 – 6.5	2	10
6.5 – 7.5	1	11

2.

Intervalos de clase	<i>f</i>
21.5 – 26.5	9
26.5 – 31.5	5
31.5 – 36.5	3
36.5 – 41.5	8
41.5 – 46.5	6

3.

Intervalos de clase	Frecuencia	Frecuencia acumulada
54.5 – 57.5	2	2
57.5 - 60.5	5	7
60.5 – 63.5	6	13
63.5 – 66.5	5	18
66.5 - 69.5	2	20

AUTOEVALUACIÓN

5. Relacionar las siguientes columnas:

1. Consiste en elaborar subgrupos homogéneos de elementos de grupos ya existentes o previamente definidos. ()	a) concepto de estadística
2. Consiste en elaborar una lista de la población y a través de números aleatorios se selecciona a cada uno de los elementos de la muestra. ()	b) muestreo probabilístico
3. Es un ejemplo claro de muestreo no probabilístico. ()	c) muestreo de juicio
4. Es la recopilación, organización, y resumen de los datos. ()	d) muestreo de conglomerados
5. Es cuando todos los elementos de la población tienen la misma probabilidad de pertenecer a la muestra. ()	e) concepto de estadística descriptiva
6. Estudia la parte de la ordenación y la presentación de datos de manera sencilla. ()	f) muestreo simple

6. Subrayar la respuesta que corresponda con la afirmación:

e) Representa las frecuencias relativas o porcentuales trazando un círculo:

- gráfica de pastel
- gráfica de barras simple
- gráfica de barras múltiples
- gráfica de barras compuestas

- f) Representa frecuencias observadas en clasificaciones dobles, cuando existen más de dos criterios de clasificación:
- gráfica de pastel
 - gráfica de barras simple
 - gráfica de barras múltiples
 - gráfica de barras compuestas
- g) Resume la distribución de frecuencias, no hay separación entre los rectángulos formados por las clases y se separa con una línea vertical:
- polígono de frecuencia
 - histograma
 - gráfica de barras múltiples
 - ojiva
- h) Se plotea el punto medio de la clase, se utiliza para las frecuencias de variables cuantitativas continuas y sólo emplea segmentos de la recta:
- polígono de frecuencia
 - histograma
 - gráfica de barras múltiples
 - ojiva

7. En las siguientes afirmaciones o definiciones, indicar la palabra que falta:

- d) La _____ es la suma de los valores de ciertos números de cantidades dividido entre el número total de elementos.
- e) La _____ es la raíz de n del producto de n términos, se utiliza para calcular las tasas de crecimiento.
- f) La _____ es el valor intermedio que divide un conjunto de datos previamente ordenados de mayor a menor.
- g) Las medidas de _____ hacen referencia a la variedad de los valores de las observaciones.

- h) Las medidas de dispersión son: la _____, la _____ y el _____ .
- i) La _____ es la raíz cuadrada de la varianza y mide la separación que hay entre los datos.

Respuestas

1.

1. d)

2. f)

3. c)

4. a)

5. b)

6. e)

2.

a) gráfica de pastel

b) gráfica de barras múltiples

c) histograma

d) polígono de frecuencia

3.

a) media aritmética

b) media geométrica

c) mediana

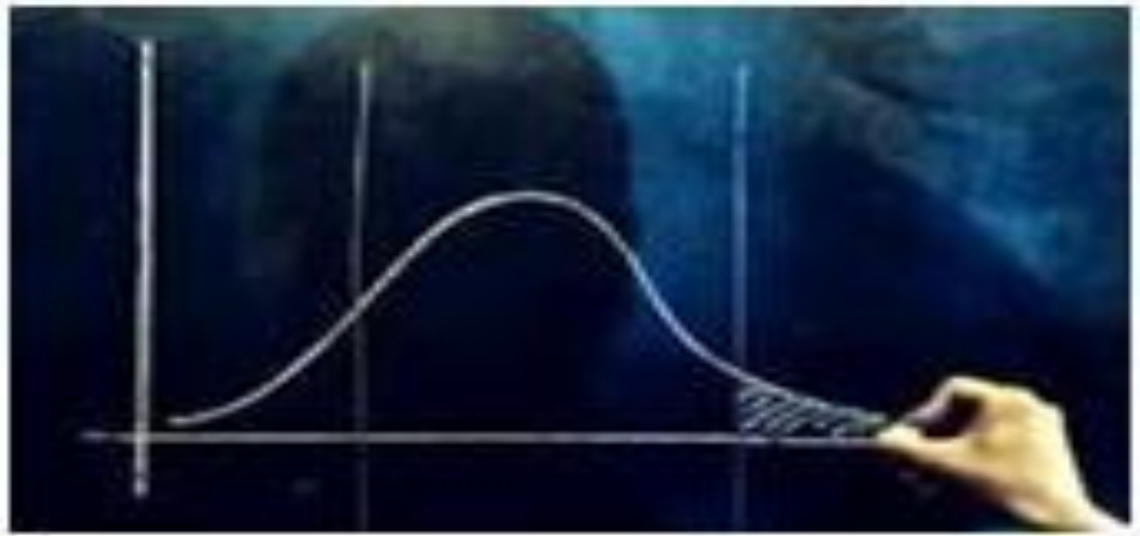
d) dispersión

e) varianza, desviación estándar y coeficiente de variación

f) desviación estándar

UNIDAD 4

ESTADÍSTICA INFERENCIAL



OBJETIVO

El estudiante será capaz de obtener información importante, para llevar a cabo la toma de decisiones acerca de la población de estudio a partir de las muestras, además de utilizar los procedimientos estudiados en la presente unidad en casos prácticos.

TEMARIO

4.1 DISTRIBUCIONES MUESTRALES

4.1.1 Muestreo aleatorio simple

4.1.2 Distribución de la media de la muestra

4.1.3 Distribución de la diferencia entre las medias de dos muestras

4.1.4 Distribución de la proporción de la muestra

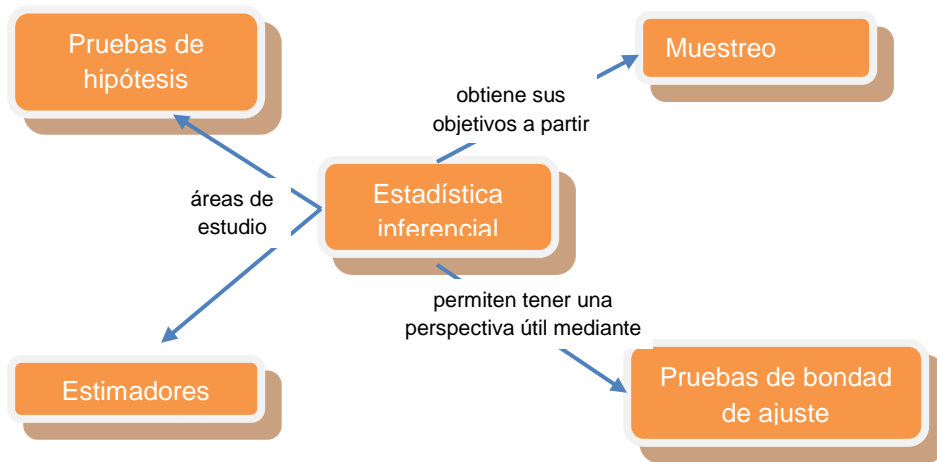
4.1.5 Distribución de la diferencia entre las proporciones de dos muestras

4.2 ESTIMADORES

4.3 PRUEBA DE HIPÓTESIS

4.4 PRUEBAS DE BONDAD DE AJUSTE

MAPA CONCEPTUAL



INTRODUCCIÓN

Es importante conocer la estadística inferencial porque permite a investigadores y administradores llevar a cabo la toma de decisiones en su población de estudio, con base en muestras que son extraídas de la población, pues permite disminuir costos innecesarios y ahorrar tiempos por el solo hecho de trabajar con una muestra.

El estudio de las distribuciones muestrales es el vínculo con lo ya revisado de estadística descriptiva hacia la realización de inferencias. La estadística inferencial tiene dos áreas de estudio los estimadores y las pruebas de hipótesis. Los estimadores favorecen el trabajo con una pequeña porción de datos, en lugar de todos los que conforman la población, mientras que las pruebas de hipótesis facilitan tomar alguna decisión en torno a la población conforme el cálculo de la muestra.

Es importante señalar que en ocasiones surgen discrepancias sobre los valores observados y los esperados en los cálculos, por lo que se requiere comprobar si la serie de datos está ajustada a las distribuciones de probabilidad ya conocidas, por medio de las pruebas de ajuste de bondad.

4.1 DISTRIBUCIONES MUESTRALES

La distribución muestral es la frecuencia de un estadístico muestral que se obtiene de todos los posibles valores calculados a partir de muestras del mismo tamaño extraídas al azar de la misma población. Las distribuciones muestrales son importantes para comprender la inferencia estadística, recordando que ésta es la que permite realizar conclusiones acerca de las poblaciones por medio de muestras representativas. Para lograr entender la experimentación en estadística, sobre todo dentro de la descriptiva, es importante mencionar que la probabilidad es la base del estudio de la estadística, ya que se interesa por estudiar las muestras de la población.

Para construir una distribución muestral se requiere:

1. Población de tamaño finito, donde es posible extraer las muestras.
2. Calcular el estadístico de interés
3. Enlistar los valores observados de la estadística.

En forma general, las distribuciones muestrales son un apoyo para conocer la media y la varianza.

ACTIVIDAD DE APRENDIZAJE

Investigar, enumerar y realizar un breve resumen de las distribuciones muestrales más importantes estudiadas en la estadística inferencial.

4.1.1 *Muestreo aleatorio simple*

Antes de poder dar una definición del muestreo aleatorio simple es importante conocer que la muestra probabilística es cuando cada uno de los elementos o miembros de una población tiene exactamente la misma probabilidad de pertenecer a la muestra.

Cuando se menciona el concepto de muestra aleatoria simple, se define que cada una de las muestras pertenecientes a la población tiene exactamente la misma probabilidad de ser seleccionada.

Hay dos diferentes tipos de muestreo: con reemplazo y sin reemplazo. En primer caso, cada miembro de la población está disponible para la extracción; por ejemplo, cuando se realiza alguna auditoria de ISO-9000 los auditores seleccionan al azar algunos trabajadores y al término de las preguntas que se les realizan, regresan a la población, estando nuevamente disponibles para otra extracción. El segundo caso, se ocupa más en la práctica y se define que cuando algún miembro de la población es extraído para la muestra, ya no vuelve a estar disponible para otra extracción; por ejemplo las degustaciones de algún producto nuevo de comida en los centros comerciales, una vez que un individuo se extrae una muestra, ya no puede conformar parte de una nueva extracción.

ACTIVIDAD DE APRENDIZAJE

Elaborar un listado de cinco ejemplos de muestreo con reemplazo y sin reemplazo, aplicables a la profesión que está estudiando.

4.1.2 *Distribución de la media de la muestra*

Pueden existir dos casos de muestreo: el primero, que pueda partir de una distribución normal y el segundo, a partir de una población que no tiene distribución normal.

La distribución normal se estudió en la unidad 2, y para el caso de no existir una distribución normal, cuando se tiene una muestra grande, se utiliza el teorema del límite central.

El teorema del límite central consiste en que dada una población no normal con medias y varianzas finitas, pero con distribución muestral de la media grande, éstas se distribuyen en forma aproximadamente normal .

En estadística, se considera que una muestra es grande, cuando es mayor de 30 elementos o contiene más de 5% de las observaciones en la población; sin embargo, con la aplicación del teorema del límite central se corrige, porque la distribución de muestreo de la media será aproximadamente normal.

Entonces, modificando la fórmula de Z , estudiada en la unidad 2, se obtiene la fórmula de la distribución normal de la muestra en la distribución normal unitaria, quedando:

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}$$

Donde la raíz cuadrada de la varianza de la distribución muestral, es decir $\frac{\sigma}{\sqrt{n}}$, se conoce como el error estándar $\sigma_{\bar{x}}$.

Ejemplo: supóngase que las calificaciones de los alumnos de una prueba diagnóstica están distribuidas de forma casi normal con una media de 185.6 y una desviación estándar de 12.7, ¿cuál es la probabilidad de que una muestra aleatoria de tamaño 10 de esta población, tenga una calificación media mayor a 190? Se aplica el teorema del límite central porque aunque la muestra es menor a 30, tiene una distribución no normal, entonces es aplicable el teorema.

Sustituyendo:

$$z = \frac{190 - 185.6}{\frac{12.7}{\sqrt{10}}}$$

$$z = \frac{4.4}{4.02}$$

$$z = 1.09$$

El valor de 1.09 se busca en la tabla de distribución normal empleada en la unidad 2 (véase tabla al final de la unidad, antes de la Autoevaluación).

De acuerdo con la tabla, el valor es 0.3621, al cual se le resta 0.5, ya que tiende a la derecha porque se quiere conocer la probabilidad de que la media sea mayor a 90; por tanto, el resultado es 0.1379.

Respuesta: la probabilidad de que la muestra tenga una calificación media mayor a 190 es de 13.79%.

ACTIVIDAD DE APRENDIZAJE

De acuerdo con los siguientes ejercicios, calcular las probabilidades, además de valorar para el desarrollo de los problemas si se aplica el teorema del límite central.

- 1 Si la media y la desviación de respuestas asertivas en un examen de conocimientos es de 120 y 13, respectivamente, ¿cuál es la probabilidad de que una muestra al azar de tamaño 40 proporcione una media entre 115 y 128?
- 2 De acuerdo con los datos del ejercicio anterior, ¿cuál es la probabilidad que la muestra al azar de tamaño 40, proporcione una media menor a 105?
 - i) Supóngase que los salarios por día de los empleados de una empresa están distribuidos de forma casi normal con una media y una desviación estándar de 58.6 y 6.5, respectivamente, y si se selecciona una muestra al azar de tamaño 16, ¿cuál es la probabilidad de que la media del salario por día de la muestra sea mayor a 90?

4.1.3 *Distribución de la diferencia entre las medias de dos muestras*

Suele haber casos en los que existe interés en calcular la diferencia entre las medias de dos poblaciones; incluso se puede darse el caso de que al existir dos medias diferentes, exista interés por saber cuánto difieren éstas. Este cálculo se puede realizar por medio de la siguiente fórmula:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Ejemplo: supóngase que hay dos grupos de personas que empaacan el producto de venta de una empresa; el primero, lo hace en promedio en 45 min, con una desviación estándar de 15 min, y el segundo lo hace en 30 min, con una desviación estándar de 20 min. Si un trabajador seleccionado al azar del primer grupo empaaca 35 productos y un trabajador seleccionado al azar del segundo grupo empaaca 40, ¿cuál es la probabilidad de que la duración media del empaque difiera entre los dos grupos por más de 20 min?

Sustituyendo:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{20 - (45 - 30)}{\sqrt{\frac{(15)^2}{35} + \frac{(20)^2}{40}}}$$

$$= \frac{20 - (15)}{\sqrt{16.4286}}$$

$$= \frac{5}{4.05}$$

$$= 1.23$$

El valor de 1.23 se busca en la tabla de distribución normal (véase tabla al final de la unidad, antes de la Autoevaluación).

De acuerdo con la tabla, el valor es 0.3907, al cual se le resta 0.5, ya que se quiere conocer la diferencia entre las medias de las dos muestras mayor a 20 min, por tanto, el resultado es .01093.

Respuesta: la probabilidad que la duración media del empaque difiera entre los dos grupos por más de 20 min es de 10.93%.

ACTIVIDAD DE APRENDIZAJE

Calcular la probabilidad de los siguientes ejercicios.

1. Dadas dos poblaciones con distribución no normal, la primera con media de 280 y varianza de 122, y la segunda con media de 350 y varianza de 188, ¿cuál es la probabilidad de que las muestras 1 y 2 (de tamaño 41 y 48, respectivamente), proporcionen una diferencia de medias menor a 17?
2. En un estudio de ingreso familiar se investigó a dos poblaciones: la primera con un ingreso medio familiar de 6,800, varianza de 2,800 y muestra de 32, y la segunda con un ingreso medio familiar de 6,250, varianza de 3,200 y muestra de 47, ¿cuál es la probabilidad de que el ingreso familiar medio difiera entre los dos grupos por menos de 1,000?

4.1.4 *Distribución de la proporción de la muestra*

Cuando se quiere calcular la probabilidad de la proporción de una muestra se emplea la siguiente fórmula:

$$z = \frac{\check{p} - p}{\sqrt{\frac{p(1-p)}{n_1}}}$$

Ejemplo: supóngase que el promedio de personas que realizan estudios de nivel medio superior en una comunidad de alta marginación económica, es

de 8%. Si se seleccionan al azar 150 personas, ¿cuál es la probabilidad de que la proporción de alta marginación sea mayor a 10%?

Sustituyendo:

$$\begin{aligned} z &= \frac{0.10 - .08}{\sqrt{\frac{.10(1 - 0.10)}{150}}} \\ &= \frac{0.10 - .08}{\sqrt{\frac{.10(0.90)}{150}}} \\ &= \frac{0.10 - .08}{\sqrt{\frac{0.09}{150}}} \\ &= \frac{0.02}{\sqrt{0.0006}} \\ &= \frac{0.02}{.0244} \\ &= 0.81 \end{aligned}$$

El valor de 0.81 se busca en la tabla de distribución normal (véase tabla al final de la unidad, antes de la Autoevaluación).

De acuerdo con la tabla, el valor es 0.291, al cual se le resta 0.5, ya que se quiere conocer que la probabilidad de que la proporción de alta marginación sea mayor a 10%, por tanto, el resultado es 0.209.

Respuesta: la probabilidad de que la proporción de alta marginación sea mayor a 10% es de 20.9%.

ACTIVIDAD DE APRENDIZAJE

Calcular la probabilidad de los siguientes ejercicios.

1. Si en una población de adultos mayores, 21% no recibe ningún tipo de pensión por retiro, ¿cuál es la probabilidad de que de una muestra al azar de tamaño 90 dé una proporción mayor a 30%?
2. De acuerdo con el ejercicio anterior, calcular la probabilidad de que una muestra al azar de tamaño 133 dé una proporción menor a 13%.

4.1.5 *Distribución de la diferencia entre las proporciones de dos muestras*

En ocasiones no sólo se tiene interés por conocer la probabilidad de una sola muestra si no de dos, por tanto se debe calcular la probabilidad de la diferencia en las proporciones de dos muestras independientes extraídas de cada una de las poblaciones; para ello se emplea la siguiente fórmula:

$$z = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

Ejemplo: supóngase que la proporción de personas que compran discos compactos pirata de una población grupo 1, es 0.50, mientras que en la otra población, grupo 2, la proporción es 0.33. ¿Cuál es la probabilidad de que muestras de 100 individuos, extraídas de cada una de las poblaciones, tengan una diferencia entre las proporciones de muestras tan grande como 0.30?

Los valores son sustituidos de la siguiente manera:

$$z = \frac{(0.30) - (0.50 - 0.33)}{\sqrt{\frac{.50(1-.50)}{100} + \frac{0.33(1-0.33)}{100}}}$$

$$z = \frac{(0.30) - (0.50 - 0.33)}{\sqrt{\frac{.50(.50)}{100} + \frac{0.33(0.67)}{100}}}$$

$$z = \frac{(0.30) - (0.17)}{\sqrt{\frac{0.25}{100} + \frac{0.2211}{100}}}$$

$$z = \frac{(0.30) - (0.17)}{\sqrt{\frac{0.25}{100} + \frac{0.2211}{100}}}$$

$$z = \frac{0.13}{\sqrt{0.004711}}$$

$$z = 1.89$$

El valor 1.89 se busca en la tabla de distribución normal (véase tabla al final de la unidad, antes de la Autoevaluación).

De acuerdo con la tabla, el valor es 0.4706, al cual se le resta 0.5, ya que se quiere conocer la diferencia entre las medias de las dos muestras que sea tan grande como 0.30, por tanto, el resultado es 0.0294.

Respuesta: la probabilidad de que la diferencia entre las medias de las dos muestras sea tan grande como 0.30 es de 2.94%.

ACTIVIDAD DE APRENDIZAJE

Calcular la probabilidad de los siguientes ejercicios.

1. En cierta población de mujeres, se sabe que la proporción que han sufrido algún tipo de violencia familiar es de 0.40 y se extrajo al azar una muestra de 100 individuos y otra de tamaño 110, con la proporción de 0.30 con el mismo problema. ¿Cuál es la probabilidad de que la muestra proporcione una diferencia entre las dos muestras mayor de 0.60?
2. En una empresa, 40% de los trabajadores no está titulado, una muestra extraída al azar de 81 personas y otra de 60 dio como diferencia 11%. Si

no existe diferencia entre los dos grupos en la proporción de los trabajadores no titulados, ¿cuál es la probabilidad de observar una diferencia como ésta o menor?

4.2 ESTIMADORES

Para abordar este tema, es importante recordar que la inferencia estadística es el procedimiento mediante el cual se trabaja con muestras y de acuerdo con los resultados arrojados se realizan las inferencias en las poblaciones.

La estadística inferencial se divide en dos áreas de estudio: los estimadores y las pruebas de hipótesis.

El estudio de los estimadores implica calcular a partir de los datos de la muestra, algún estadístico que permite una aproximación del parámetro correspondiente a la población de la cual se extrajo la muestra; para cada uno de los parámetros se pueden calcular dos diferentes tipos de estimadores: el puntual y el de intervalos.

El estimador puntual es un solo valor numérico empleado para estimar el parámetro correspondiente de la población; los estimadores de este tipo se forman por medio de fórmulas (lo que ya se trabajó en este libro); por ejemplo, si se desea calcular la media muestral, se emplea la fórmula $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ y al unísono se estima la media de la población, es decir, μ .

Los estimadores cuentan con tres propiedades y la nomenclatura para representar un parámetro, es decir, el valor numérico que describe las características de la población, el cual es θ ; y para expresar el estimador, es decir, el valor numérico que describe las características de la muestra, se emplea $\hat{\theta}$.

Los estimadores tienen las siguientes propiedades:

- Si el lím $E(\hat{\theta}) = E(\theta)$
- Lím $\text{var}(\hat{\theta}) = 0$; propiedad de los insesgados.
- Cuando $n \rightarrow \infty$; y si n crece se vuelve insesgado.

- La varianza media o sesgo medio, también se conoce como varianza media uniforme, ya que se tiene una muestra aleatoria de una distribución de densidad $g(x_1, \dots, x_{1n})$ y se representa como $f(x, \theta)$ y siendo el estadístico $\hat{\theta} = u(x_1, \dots, x_{1n})$ un estimador con $E(\hat{\theta}) = \theta$ y $\text{var}(\hat{\theta})$ menor que cualquier otro estimador insesgado del parámetro para todos los posibles valores, es insesgada y de varianza mínima. La eficiencia, si $\hat{\theta}_1$ y $\hat{\theta}_2$ son 2 estimadores insesgados de θ y la varianza de $\hat{\theta}_1$ es menor a 0 o al menos a la varianza de $\hat{\theta}_2$.

Cuando $\hat{\theta}$ es insesgado de varianza mínima se llama estimador eficiente. El otro estimador es por intervalos; consta de dos valores numéricos, definidos por medio de un intervalo con un grado de confianza que incluye el parámetro que se está estimando. La forma general para calcular un intervalo de confianza es:

$$\text{Estimador} \pm \text{coeficiente de confianza} \times \text{error estándar}$$

El coeficiente de confianza puede ser cualquier valor, pero en la práctica los más utilizados son 90, 95 y 99% y aunque los valores de los diferentes valores de los coeficientes de confianza corresponden con los valores de la tabla z de distribución normal, con la que se ha trabajado en ejercicios anteriores, para facilitar el trabajo se presenta una tabla donde vienen especificados los coeficientes de confiabilidad de más uso.

Nivel de confianza	Z
99%	2.58
98%	2.33
96%	2.05
95%	1.96
90%	1.645
85%	1.43

Se pueden construir diferentes intervalos de confianza, los cuales se mencionan a continuación:

a) Intervalo de confianza para la media de una población.

$$\bar{x} \mp z_{(1-\frac{\alpha}{2})} \sigma_{\bar{x}}$$

Ejemplo: el departamento de selección de una empresa desea estimar, con 99% de confianza, la media de calificaciones de un cierto grupo de candidatos de ser seleccionados para un empleo. La distribución es aproximadamente normal y su desviación estándar es de 12. De una población se extrae una muestra de 16 individuos, la cual arrojó 88 como resultado en su media de calificaciones. Calcular el intervalo de confianza.

Solución:

$$88 \mp 2.58 * \frac{12}{\sqrt{16}}$$

$$88 \mp 2.58 * \frac{12}{4}$$

$$88 \mp 2.58 * 3$$

$$88 \mp 7.74$$

$$80.26, 95.74$$

Respuesta: de la media de las calificaciones de la población con 99% de confianza, está entre 80.26 y 95.74.

b) Intervalo de confianza para la diferencia entre las medias de dos poblaciones.

$$\bar{x}_1 - \bar{x}_2 \mp z_{(1-\frac{\alpha}{2})} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Ejemplo: el ingreso medio familiar de una muestra de 50 personas es de \$8,000, mientras que en otra muestra, también de 50 personas, es de \$6,000; las desviaciones estándar son de \$20 y \$50, respectivamente, ¿cuál es el intervalo de confianza de 99% para la diferencia de las dos poblaciones?

Solución:

$$8,000 - 6,000 \mp 2.58 \sqrt{\frac{20^2}{50} + \frac{50^2}{50}}$$

$$2,000 \mp 2.58\sqrt{8 + 50}$$

$$2,000 \mp 2.58 * 7.6157$$

$$2,000 \mp 19.64$$

$$1,980.35 \mp 2019.64$$

Respuesta: la diferencia entre las medias de dos poblaciones con 99% de confianza, está entre \$1,980.35 y \$2019.64.

c) Intervalo de confianza para la proporción de una población.

$$\check{p} \mp z_{(1-\frac{\alpha}{2})} \sqrt{\frac{\check{p}(1-\check{p})}{n}}$$

Ejemplo: se encuestó a 800 personas, de las cuales 600 afirmaron que pagan sus respectivos impuestos. Calcular el intervalo de confianza de 99%.

Solución:

$$0.75 \mp 2.58 \sqrt{\frac{0.75(1 - 0.75)}{800}}$$

$$0.75 \mp 2.58 \sqrt{\frac{0.75(0.25)}{800}}$$

$$0.75 \mp 2.58 \sqrt{\frac{0.1875}{800}}$$

$$0.75 \mp 2.58 * 0.1530$$

$$0.75 \mp 0.0394$$

$$0.7105, 0.7894$$

Respuesta: la proporción de una población con 99% de confianza, se encuentra entre 0.7105 y 0.7894.

- d) Intervalo de confianza para la diferencia entre las proporciones de dos poblaciones.

$$(\bar{p}_1 - \bar{p}_2) \mp z_{(1-\frac{\alpha}{2})} \sqrt{\frac{\check{p}_1(1-\check{p}_1)}{n_1} + \frac{\check{p}_2(1-\check{p}_2)}{n_2}}$$

Ejemplo: de una muestra de 200 personas, 100 de ellos cuentan con un ahorro en una institución bancaria y de otra muestra seleccionada de 120 personas, 60 de ellas tienen una cuenta de ahorro. Calcular el intervalo de confianza de 95%.

Solución:

$$(0.5 - 0.5) \mp 1.96 \sqrt{\frac{0.5(1-0.5)}{100} + \frac{0.5(1-0.5)}{120}}$$

$$0 \mp 1.96 * 0.0677$$

$$\mp 0.1326$$

$$-0.1326, 0.1326$$

- e) Respuesta: la confianza de la diferencia de 95% entre las proporciones de dos poblaciones, se encuentra entre -0.1326 y 0.1326.

ACTIVIDAD DE APRENDIZAJE

Calcular la probabilidad de los siguientes ejercicios e identificar el tipo de intervalo de confianza.

1. En una oficina gubernamental se quiere conocer el intervalo de confianza de 95% del tiempo de retraso en la llegada al trabajo; los datos arrojados del estudio elaborado son de 11.2 minutos en promedio, con una desviación estándar de 8 minutos, de una muestra extraída al azar de 50 trabajadores.
2. Calcular el intervalo de confianza de 99% de un estudio elaborado en un consultorio dedicado a impartir terapia de pareja. Se trabajó con dos muestras, una de 10 parejas y otra de 8. Todas las parejas contestaron un cuestionario acerca de los logros con su pareja. La calificación media de la primera muestra fue de 57 y de la segunda de 89; de acuerdo con la experiencia bajo situaciones similares, las calificaciones de parejas sin problemas al contestar el cuestionario muestran una distribución normal con desviaciones estándar de 13 y 17, respectivamente.
3. Calcular el intervalo de confianza de 99% de una encuesta que se realizó a 350 personas, en relación con su chequeo anual de salud, de las cuales 127 contestaron que se lo realizan.

4.3 PRUEBA DE HIPÓTESIS

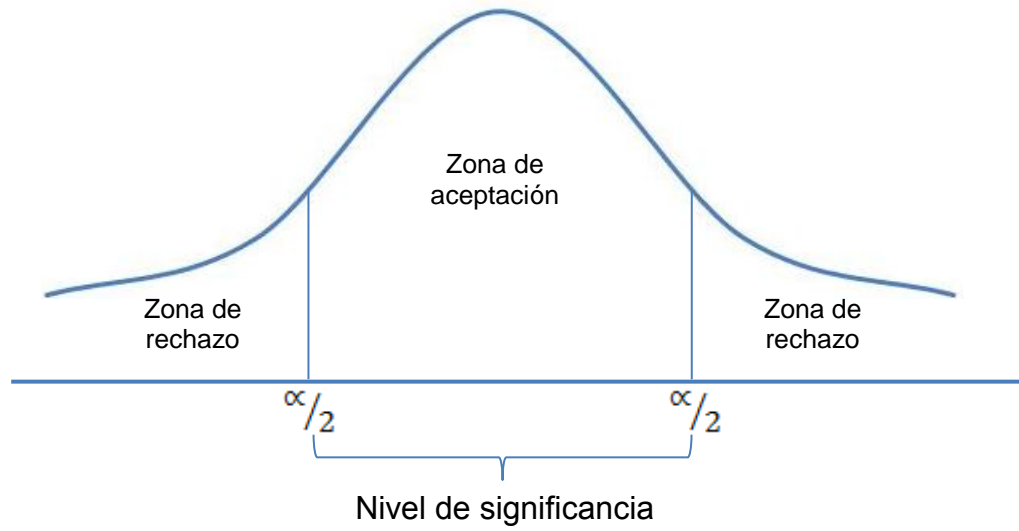
Las hipótesis son empleadas por los investigadores o los administradores que pretenden tomar una decisión en torno a su población, mediante la elaboración de ciertos cálculos con apoyo de la muestra.

La prueba de hipótesis es una afirmación acerca de los parámetros poblacionales y se valora por medio del desarrollo de los siguientes pasos:

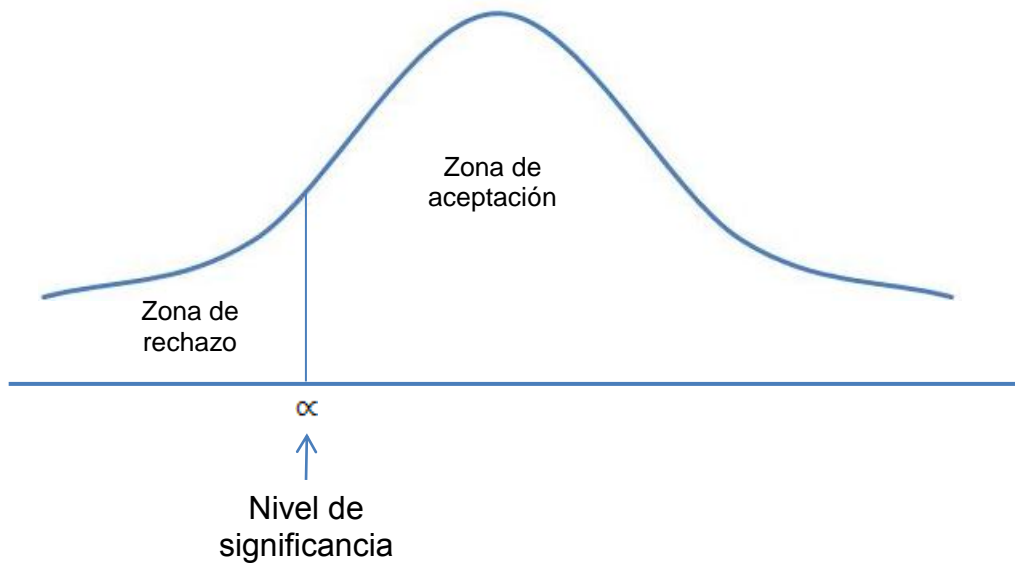
1. Datos. Son la base del desarrollo de la prueba de hipótesis.
2. Suposiciones. Hacen referencia a la independencia de las muestras y las varianzas.
3. Hipótesis. Se manejan dos hipótesis, la nula se nombra H_0 y la alternativa, H_a . La H_0 es la hipótesis de no diferencia y esto significa que es de conformidad; pero durante el desarrollo de la prueba de hipótesis se valora su aceptación o su rechazo. De acuerdo con lo anterior, la hipótesis nula se puede rechazar o se puede aceptar, si no es rechazada se argumenta que los datos sobre los cuales está basada la prueba no arrojan evidencia suficiente; si es rechazada, significa que no es compatible con la hipótesis nula, pero avala la otra hipótesis, es decir, la alternativa, H_a .
4. Estadístico de prueba. Es la enunciación que facilita el cálculo de los estadísticos por medio de la muestra.
5. Regla de decisión. Es cuando la distribución de la estadística se divide en dos grupos: la zona de aceptación y la de rechazo. La zona de aceptación es aquella región que tiene mayor probabilidad de suceder, cuando la hipótesis nula es verdadera. La zona de rechazo es la región que tiene menor probabilidad de suceder, si la hipótesis nula es verdadera.

Consiste en valorar la H_0 , la cual se rechaza si una vez calculado el estadístico de prueba, cae en la zona de rechazo, y se acepta, si el cálculo del estadístico de prueba cae en la zona de aceptación. Lo que permite valorar cuánto abarca la zona de aceptación y la zona de rechazo es el nivel de significancia; éste especifica el área debajo de la curva de la distribución estadística de prueba referente a la probabilidad de rechazar una H_0 ; por lo que siempre se buscan probabilidades pequeñas de rechazar una H_0 que sea verdadera. Los valores que en la práctica se trabajan más en el nivel de significancia son 0.01, 0.05 y 0.10.

Región de aceptación y rechazo de dos colas



Región de aceptación y rechazo de una sola cola



Región de aceptación y rechazo de una sola cola



En algunas ocasiones se pueden cometer errores, los cuales son de dos tipos. El de tipo I, se comete cuando se rechaza una H_0 verdadera y el de tipo II, es cuando se acepta una H_0 falsa.

6. Cálculo del estadístico de prueba. Es la elaboración del estadístico.
7. Decisión estadística. Es la valoración de aceptación o rechazo de la H_0 .
8. Conclusión. Si se acepta H_0 , se concluye que es verdadera y si se rechaza, se dice que H_a es verdadera.

La prueba de hipótesis es una herramienta que facilita a los investigadores y administrativos la toma de decisiones, sin embargo aunque su resultado no es definitivo, sí es una posibilidad que se debe tomar en cuenta además de la información enriquecedora que puedan poseer el investigador o el administrador.

Ejemplo 1: en una encuesta se entrevistó a 36 personas; uno de los detalles de la información obtenida son las veces que han salido de vacaciones durante toda su vida. El resultado que arrojó es que el promedio de salidas es de 22, con una desviación estándar de 6. El investigador desea saber si la encuesta realizada en la muestra, proporciona evidencia suficiente para indicar

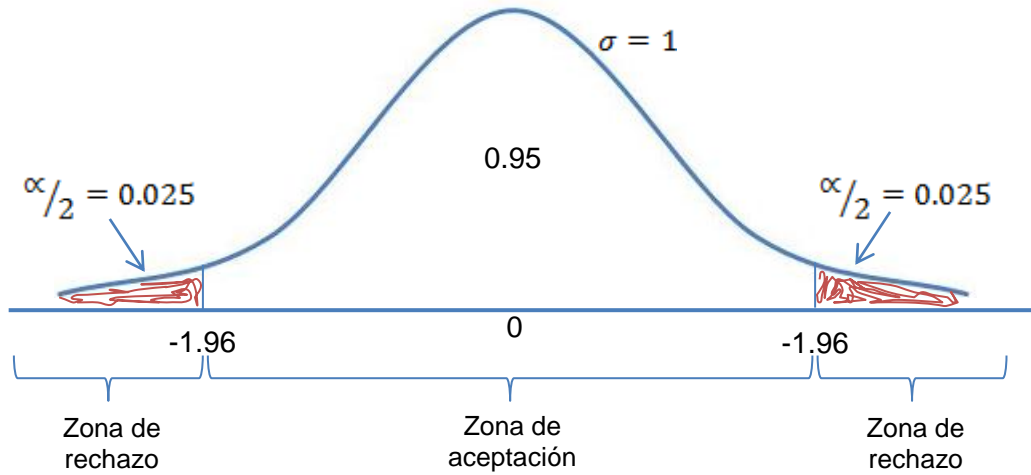
que la media de población sale en promedio 25 veces en su vida, si el grado de confianza es de 95%, es decir, un nivel de significancia de 0.05.

Solución:

1. Datos: $n = 36$, $\bar{x} = 22$.
2. Suposición: $\sigma = 6$.
3. Hipótesis: $H_0: \mu = 25$ y $H_a: \mu \neq 25$.
4. Estadístico de prueba: $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$
5. Regla de decisión: Para este caso, el grado de significancia es de 0.05, y se tienen dos zonas de rechazo por lo que α se divide entre dos, quedando $\frac{0.05}{2}$, por lo que el valor en cada cola es 0.025. Los valores del grado de confianza se obtienen por medio de la tabla de distribución normal, buscando los más próximos para la obtención de z (para facilitar el trabajo se presenta un tabla donde vienen especificados los coeficientes de confiabilidad de más uso).

Nivel de confianza	$Z_{\frac{\alpha}{2}}$	Z_{α}
99%	2.58	2.33
95%	1.96	1.645
90%	1.645	1.28

De manera más práctica, con apoyo de la tabla cuando el caso es de dos colas con un nivel de significancia de 95%, el valor es 1.96, y se representa de la siguiente manera:



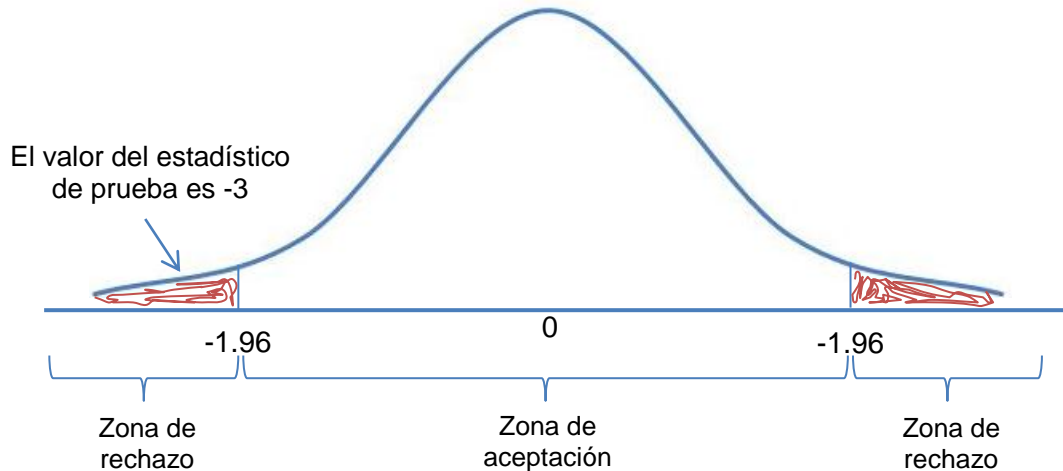
6. Cálculo del estadístico de prueba:

$$z = \frac{22 - 25}{\frac{6}{\sqrt{36}}}$$

$$= \frac{-3}{\frac{6}{6}}$$

$$= \frac{-3}{1} = -3$$

7. Decisión estadística: se rechaza H_0 , porque el valor de -3 se ubica en la zona de rechazo.



Por tanto, el valor calculado del estadístico de prueba es significativo a 95%.

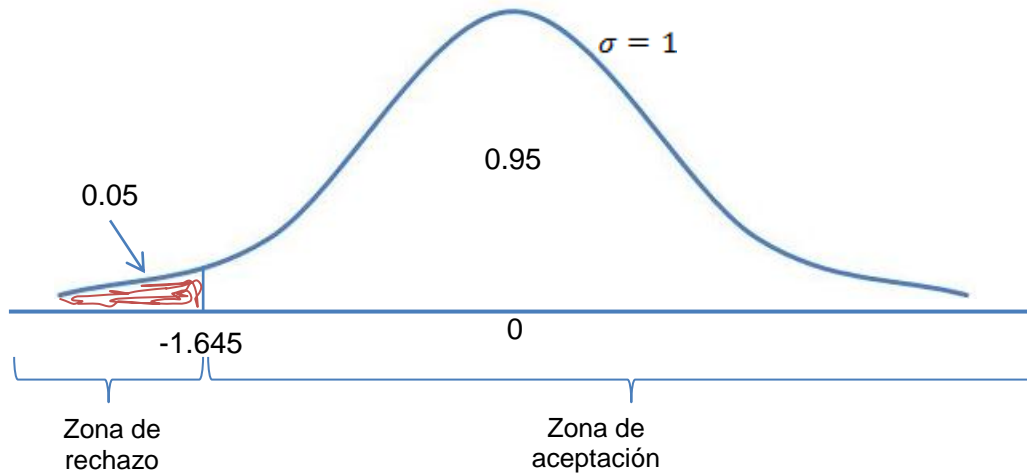
8. Conclusión: la media poblacional no sale en promedio 25 veces de vacaciones en su vida las personas, sólo hace falta valorar que el investigador esté de acuerdo con esta conclusión, recordando que no es una afirmación si no una posibilidad de acuerdo con la confiabilidad de 95%.

Ejemplo 2: conforme los datos del ejercicio anterior, valorar si la encuesta realizada a la muestra, proporciona la evidencia suficiente para indicar que la media de población sale en promedio en su vida menos de 25 veces.

Solución:

1. Datos: $n = 36$, $\bar{x} = 22$.
2. Suposición: $\sigma = 6$.
3. Hipótesis: $H_0: \mu \geq 25$ y $H_a: \mu < 25$
4. Estadístico de prueba: $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$
5. Regla de decisión: de manera más práctica, con apoyo de la tabla cuando el caso es de una cola con un nivel de

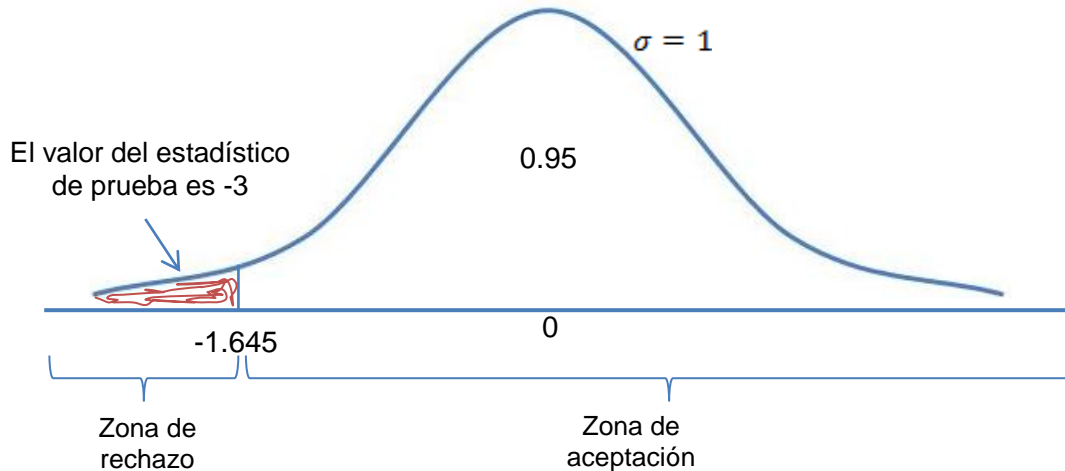
significancia de 95%, el valor es 1.645, y se representa de la siguiente manera:



6. Cálculo del estadístico de prueba:

$$\begin{aligned} z &= \frac{22 - 25}{\frac{6}{\sqrt{36}}} \\ &= \frac{-3}{\frac{6}{6}} \\ &= \frac{-3}{1} \\ &= -3 \end{aligned}$$

7. Decisión estadística: se rechaza H_0 , porque el valor de -3 se ubica en la zona de rechazo.



Por tanto, el valor calculado del estadístico de prueba es significativo a 95%.

8. Conclusión: la media poblacional sale en promedio menos 25 veces de vacaciones en su vida; sólo hace falta valorar que el investigador esté de acuerdo con esta conclusión, recordando que no es una afirmación si no una posibilidad de acuerdo con la confiabilidad de 95%.

ACTIVIDAD DE APRENDIZAJE

Calcular los siguientes ejercicios con ayuda de la prueba de hipótesis.

1. Se recolectó una muestra de 45 muertes registradas en el Estado de México y arrojó que el promedio de esperanza de vida es de 75.4 años, con una desviación estándar de 1.9 años. Se desea saber si la muestra proporciona suficiente evidencia para indicar que la población vive menos de 46 años, además se ocupará 99% de confianza, es decir un nivel de significancia de 0.01.
2. De acuerdo con el ejercicio anterior, valorar si la población vive en promedio 68 años, empleando 95% de confianza, es decir, un nivel de significancia de 0.05.

4.4 PRUEBAS DE BONDAD DE AJUSTE

Las pruebas de bondad de ajuste permiten examinar problemas con la finalidad de verificar si el conjunto de datos se puede ajustar o aseverar que proviene de una determinada distribución, es decir, permiten medir el grado de ajuste que existe entre la distribución determinada a partir de la muestra y la distribución teórica que se supone debe seguir esa muestra (los datos se ajustan a una determinada distribución que proviene de la muestra).

En muchas ocasiones no es posible conocer la distribución de probabilidad de la variable aleatoria que se está estudiando, por lo que se supone X , y se desea probar la hipótesis de que X sigue una distribución de probabilidad particular. Por ejemplo, podría ser de interés probar la hipótesis de que X sigue una distribución normal, una exponencial, etc., por lo que las pruebas de bondad de ajuste se pueden realizar en aquellos casos en que la distribución planteada en la hipótesis es de tipo normal, binominal, de Poisson, Chi cuadrada o cualquier otra distribución de las que ya se estudiaron en la unidad 2.

Las pruebas de bondad de ajuste brindan una perspectiva útil para poder evaluar la viabilidad o sustentabilidad de un modelo potencial para el suministro de datos.

La ventaja que se tiene al realizar pruebas de bondad de ajuste es que no hay una sola distribución en aplicaciones reales, sino que se plantean hipótesis en diferentes distribuciones.

Es muy importante entender el efecto del tamaño de la muestra, debido a que si muy pocos datos están disponibles, entonces una prueba de bondad de ajuste puede rechazar alguna distribución postulante; pero si hay muchos datos disponibles, entonces una prueba de ajuste de bondad puede rechazar todas las pruebas postulantes.

ACTIVIDAD DE APRENDIZAJE

De acuerdo con la bibliografía sugerida, escribir en una cuartilla cuál es la utilidad de las pruebas de bondad de ajuste.

Tabla de distribución normal

Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0	0.004	0.008	0.012	0.016	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.091	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.148	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.17	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.195	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.219	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.258	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.291	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.334	0.3365	0.3389
1	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.377	0.379	0.381	0.383
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.398	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.437	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.475	0.4756	0.4761	0.4767
2	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.483	0.4834	0.4838	0.4842	0.4846	0.485	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.489
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.492	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.494	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.496	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.497	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.498	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.499	0.499

AUTOEVALUACIÓN

8. Relacione las siguientes columnas.

<p>1. Es la frecuencia de un estadístico muestral que se obtiene de todos los posibles valores calculados a partir de muestras del mismo tamaño extraídos al azar de la misma población. ()</p> <p>2. Cada uno de los elementos de la población tienen la misma probabilidad de pertenecer a la muestra. ()</p> <p>3. Cada una de las muestras pertenecientes a la población tienen exactamente la misma probabilidad de ser seleccionadas. ()</p> <p>4. Cada miembro de la población está disponible para otra extracción después de utilizarse. ()</p> <p>5. Cuando cada miembro de la población es extraído para la muestra, pero ya no vuelve a estar disponible para otra extracción; este tipo de muestreo usualmente se utiliza en la práctica. ()</p>	<p>a) muestreo sin reemplazo</p> <p>b) muestreo con reemplazo</p> <p>c) muestra aleatoria simple</p> <p>d) muestra probabilística</p> <p>e) distribución muestral</p>
--	---

9. Subrayar la respuesta que corresponda con la afirmación.

- j) El teorema del límite central se utiliza cuando se tienen muestras de tamaño grande y son las que se consideran:

- mayores a 20 elementos
- menores a 20 elementos
- mayores a 30 elementos

k) Los estimadores y las pruebas de hipótesis son dos áreas de estudio de la:

- probabilidad
- estadística descriptiva
- estadística inferencial

l) Se calculan a partir de los datos de la muestra de algún estadístico que permite alguna aproximación del parámetro correspondiente de una población de la cual se extrajo la muestra:

- distribuciones normales
- estimadores
- pruebas de hipótesis

m) Es un solo valor numérico empleado para estimar un parámetro correspondiente de la población:

- estimador puntual
- estimador por intervalos
- prueba de hipótesis

n) Son dos valores numéricos y son definidos por medio de un intervalo con un grado de confianza e incluyen el parámetro estimado:

- estimador puntual
- estimador por intervalos
- prueba de hipótesis

10. En las siguientes afirmaciones o definiciones indicar la palabra que falta.

j) Las _____ son empleadas por los investigadores o administradores que pretenden tomar una decisión en

torno a su población, mediante la elaboración de ciertos cálculos con el apoyo de la muestra.

- k) Las _____ hacen referencia a la independencia de las muestras y las varianzas.
- l) El _____ es la enunciación que facilita calcular los estadísticos por medio de la muestra.
- m) El _____ es cuando se acepta una hipótesis nula falsa.
- n) La _____ es valorar la aceptación o el rechazo de la hipótesis nula.
- o) Las _____ permiten examinar problemas con la finalidad de verificar si el conjunto de datos se pueden ajustar o aseverar que provienen de una determinada distribución.

Respuestas

1.

- 1. e)
- 2. d)
- 3. c)
- 4. b)
- 5. a)

2.

- a) mayores a 30 elementos
- b) estadística inferencial
- c) estimadores
- d) estimador puntual
- e) estimador por intervalos

3.

- a) hipótesis
- b) suposiciones

- c) estadístico de prueba
- d) error tipo II
- e) decisión estadística
- f) pruebas de bondad

GLOSARIO

Axioma. Premisa que se considera evidente y se acepta sin requerir demostración previa.

Coefficiente de variación. Se utiliza cuando se quiere comparar la dispersión de dos conjuntos de datos porque se desea conocer la variación relativa y se expresa la desviación estándar como un porcentaje de la media.

Combinación. Selección de diferentes de artículos sin importar el orden de los objetos sin repetición.

Desviación estándar. Raíz cuadrada de la varianza, mide la separación que hay entre los datos.

Distribución muestral. Frecuencia de un estadístico muestral que se obtiene de todos los posibles valores calculados a partir de muestras del mismo tamaño extraídas al azar de la misma población.

Equiprobable. Caso en el que todos los eventos tienen la misma probabilidad de ocurrencia.

Error de tipo I. Cuando se rechaza una H_0 verdadera.

Error de tipo II. Cuando se acepta una H_0 falsa.

Espacio muestral. Conjunto de todos los posibles resultados que pueden ocurrir en un experimento.

Esperanza matemática. Valor esperado de una variable aleatoria

Estadística descriptiva. Estudia la parte de la ordenación y la presentación de datos de manera sencilla; por tanto, describe las variables con las que se trabaja.

Estadística inferencial. Consiste en extrapolar los resultados obtenidos en el análisis de los datos y a partir de ello predecir acerca de una población, con un margen de confianza conocido; se apoya en el cálculo de probabilidades.

Estimador puntual. Un solo valor numérico empleado para estimar el parámetro correspondiente de la población.

Evento. Suceso subconjunto del espacio muestral.

Experimento determinístico. Aquel en que independientemente de las condiciones naturales mediante las cuales se efectúa el experimento, los resultados siempre serán los mismos.

Experimento no determinístico. Aquel en el que las condiciones materiales bajo las cuales se efectúa el experimento, determinan la probabilidad de ocurrencia de los resultados.

Experimento. Proceso mediante el cual se obtienen resultados con la propiedad de la repetibilidad.

Hipótesis. Afirmación acerca de los parámetros poblacionales.

Media aritmética. Suma de los valores de ciertos números de cantidades dividido entre el número total de elementos.

Media armónica. Recíproco de la media aritmética de los recíprocos de los números de la serie.

Media geométrica. Raíz de n del producto de n términos; se utiliza para calcular tasas de crecimiento.

Mediana. Valor intermedio que divide un conjunto de datos previamente ordenados de menor a mayor.

Moda. Valor que más se repite en un conjunto de datos.

Muestra aleatoria simple. Caso en el que cada una de las muestras pertenecientes a la población tiene exactamente la misma probabilidad de ser seleccionada.

Muestra probabilística. Caso en el que cada uno de los elementos o miembros de una población tiene exactamente la misma probabilidad de pertenecer a la muestra.

Muestreo con reemplazo. Caso en el que cada miembro de la población está disponible para la extracción.

Muestreo sin reemplazo. Caso en el que cada miembro de la población está disponible sólo para una extracción.

Mutuamente excluyente. Implica la ocurrencia de un evento u otro, no dos a la vez.

Permutación. Arreglo ordenado de artículos, en el cual interesa el orden de los objetos sin repetición.

Probabilidad clásica. Probabilidad *a priori*, se calcula antes del evento.

Probabilidad condicional. Medida en que ocurre un evento en particular, dado que otro ya ha ocurrido o es seguro que ocurra.

Probabilidad frecuencial. Probabilidad *a posteriori* porque es modificada con base en información nueva.

Probabilidad subjetiva. Probabilidad de asignación basada en conjeturas y experiencia; puede ocurrir una vez y a lo máximo unas cuantas veces más.

Probabilidad. Asignación de un número entre cero y uno a cada resultado experimental que permite medir la certidumbre o incertidumbre.

Teorema. Afirmación que puede ser demostrada dentro de un sistema formal.

Variable. Objeto matemático que puede tomar diferentes valores, generalmente asociado con propiedades o características de las unidades de la muestra.

Variable aleatoria. Variable cuyo resultado depende de la muestra de una distribución de probabilidad.

Variables aleatorias continuas. Variables que permiten una infinidad de valores al azar dentro de un intervalo, considerándose continuas precisamente por la posibilidad de poder tomar cualquier valor dentro de una infinidad de valores.

Variables aleatorias discretas. Variables cuya naturaleza toma un número finito de valores enteros.

Varianza. Valores de observaciones muy cercanos a la media que se calcula para conocer la dispersión de los datos.

BIBLIOGRAFÍA

Webster, Allen L., *Estadística aplicada a los negocios y la economía*, México, McGraw-Hill, 2002.

Canavos, George, *Probabilidad y estadística*, México, McGraw-Hill, 1992.

Cuadras, Carles M., *Problemas de probabilidades y estadística I y II*, España, PPU, 1999.

Díaz, Godino, *Azar y probabilidad*, España, Síntesis, 1991.

Elorza, Haroldo, *Estadística para las ciencias sociales y del comportamiento*, México, Oxford University Press, 1999.

Engel, Arthur, *Probabilidad y estadística*, España, Consorci d'Editors Valencians, 1988.

Freud, John y Gary A. Simon, *Estadística elemental*, México, Prentice-Hall, 2003.

Fuenlabrada, Samuel, *Probabilidad y estadística*, México, McGraw-Hill, 2000.

Leonard J., Kazmier, *Estadística aplicada a la administración y la economía*, México McGraw-Hill, 1998.

Martín Pliego, Francisco Javier, *Introducción a la estadística económica y empresarial*, España, Editorial AC, 1994.

MendenHall, William, Richard Scheaffer y Dennis Wackely, *Estadística matemática con aplicaciones*, México, Grupo Editorial Iberoamericana, 1986.

Peña, Daniel y Juan Romo, *Introducción a la estadística para las ciencias sociales*, España, McGraw-Hill, 1997.

Quesada, Isidoro, *Curso y ejercicios de estadística*, España, Alhambra, 1989.

Stevenson, William J., *Estadística para administración y economía*, México, Harla, 1993.

Spiegel, Murray R., *Estadística*, México, McGraw-Hill, 1970.

Ritchey, Ferris J., *Estadística para las ciencias sociales*, México, McGraw-Hill, 2002.