UTPL
La Universidad Católica de Loja

Modalidad Abierta y a Distancia

Estadística para Ingenierias y Arquitectura

Guía didáctica

UTPL

Departamento de Ciencias de la Computación y Electrónica

Sección departamental de Inteligencia Artificial

Estadística para Ingenierías y Arquitectura

Guía didáctica

Autor:

Torres Díaz Juan Carlos

ESTA_2029

Asesoría virtual
www.utpl.edu.ec

Estadística para Ingenierias y Arquitectura
Guía didáctica
Juan Carlos Torres Díaz

Universidad Técnica Particular de Loja

Diagramación y diseño digital:

Ediloja Cía. Ltda.
Telefax: 593-7-2611418.
San Cayetano Alto s/n.
www.ediloja.com.ec
edilojainfo@ediloja.com.ec
Loja-Ecuador

ISBN digital - 978-9942-25-882-3

29 de septiembre, 2020

Índice

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

1. Datos de información

1.1. Presentación de la asignatura

1.2. Competencias genéricas de la UTPL

- Pensamiento crítico.

1.3. Competencias específicas de la carrera

- Construir modelos específicos de ciencias de la computación mediante esquemas matemáticos y estadísticos, para propiciar el uso y explotación eficiente de datos e información.

1.4. Problemática que aborda la asignatura en el marco del proyecto

Utilizar modelos de datos para almacenar información.

1.5. Proyecto integrador de saberes

Estudio de tecnologías de la información usadas en el entorno del estudiante y análisis de las implementaciones de tecnologías en contextos empresariales.

2. Metodología de aprendizaje

Se aplica como metodología el aprendizaje basado en problemas, en cada unidad se plantean problemas que para resolverse requieren que el estudiante trabaje de forma iterativa buscando la información y los conceptos necesarios para llegar a la meta; en cada unidad se presenta una práctica de la unidad que debe ser resuelta.

3. Orientaciones didácticas por resultados de aprendizaje

Primer bimestre

Resultado de aprendizaje 1

Entiende el alcance y utilidad de la estadística en la resolución de problemas.

Contenidos, recursos y actividades de aprendizaje

Semana 1

Unidad 1. Introducción a la estadística

El contenido referente a esta unidad puede encontrarlo en el documento “Unidad 1”, en él se presenta paso a paso cada uno de los temas.

1.1. Definiciones de estadística

Dé lectura al documento “Unidad 1” y siga las instrucciones que ahí se dan. Se recomienda que usted escriba en su cuaderno de notas su propia definición de estadística, recuerde que puede contar con mi apoyo. Si tiene preguntas, hágamelas saber y estaré contestándole lo más pronto posible, no olvide que estoy para apoyarle en su proceso de aprendizaje.

Observe el video y tome nota de las definiciones de estadística.

1.2. Tipos de estadística

La estadística abarca dos partes estadística descriptiva y estadística inferencial. En el “Documento 1” se detalla cada parte.

Ir a recursos

Observe el video y tome nota de las diferencias entre estadística descriptiva e inferencial.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

1.3. Tipos de variables

Este tema es muy sencillo, le sugiero revisar la sección “Tipos de variables” del texto básico. Como tarea anote en su cuaderno los tipos de variables y escriba dos ejemplos de cada tipo. Esto puede ser complementado con la lectura de la sección correspondiente del “Documento 1”.

En el video siguiente se dan ejemplos y explicaciones al respecto.

1.4. Niveles de medición de variables

Los niveles de medición son cuatro (4), revise cuáles son y qué tipos de escalas se manejan, puede revisar esto en el “Documento 1”.

Observe el video y tome nota de los ejemplos dados para los distintos niveles de medición.

Actividades de aprendizaje recomendadas

Para reforzar el contenido de una manera práctica le recomiendo lo siguiente.

Busque en Internet información sobre las encuestas de elección de alcalde de su ciudad (elecciones anteriores) y determine cuáles eran los candidatos que ocupaban los primeros lugares en las preferencias de la ciudadanía. Con base en esto determine

¿Los resultados de las encuestas eran producto de aplicar estadística descriptiva o inferencial?, ¿por qué?

Autoevaluación 1

Desarrolle la autoevalución; le permitirá determinar el nivel de aprovechamiento del tema estudiado.

En base a la siguiente encuesta conteste las preguntas: 1. *Edad _____* 2. *Sexo M ( ) F( )* 3. *Tiene teléfono móvil SÍ ( ) NO( )* 4. *Cuánto gasta al mes en telefonía móvil ________* 5. *¿Cómo se calificaría usted mismo?* • *Escribo muchos mensajes al día* • *Escribo pocos mensajes al día* • *Casi no escribo mensajes*

Escriba Verdadero (V) o Falso (F) a las siguientes afirmaciones, según corresponda:

1. (   ) La pregunta 1 utiliza una variable que se mide a nivel de intervalo.
2. (   ) La pregunta 1 utiliza una variable que se mide a nivel ordinal.
3. (   ) La pregunta 2 utiliza una variable que se mide a nivel nominal.
4. (   ) La pregunta 3 utiliza una variable que se mide a nivel de razón.
5. (   ) La pregunta 4 utiliza una variable que se mide a nivel de razón.
6. (   ) La pregunta 5 utiliza una variable que se mide a nivel ordinal.

7. (   ) La inferencia es utilizar un estadístico para utilizarlo como parámetro.

8. (   ) Una muestra es equivalente a la población.

9. (   ) Una variable continua puede aceptar valores decimales.

10. (   ) Si se pregunta por el número de hijos de una familia nos referimos a una variable continua.

Ir al solucionario

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Resultado de aprendizaje 2 | Representa datos utilizando gráficas.

Contenidos, recursos y actividades de aprendizaje

Semana 2

Unidad 2. Distribuciones de frecuencia y representaciones gráficas

En esta sección es necesario que instale las herramientas Tableau Public y rStudio. En los siguientes videos están las instrucciones de descarga.

La descarga de Tableau Public es en el siguiente enlace

La dirección para la descarga de R y RStudio es https//cran.r-project.org/bin/windows/

https//www.rstudio.com/products/rstudio/download/

Tableau Public es un software que permite visualizar datos mientras que RStudio es un software estadístico que permite hacer todo tipo de cálculos.

Una de las tareas más comunes es la organización de conjuntos de datos y presentarlos de una manera entendible; para explicar este tema partamos de un ejemplo sencillo.

Seguidamente de lectura al “Documento 2” en el que se detalla el contenido.

Ir a recursos

2.1. Distribución de frecuencias

Esta es una de las áreas más interesantes y útiles para comprender la información que se nos presenta en la vida cotidiana. Para abordar el tema usted debe revisar en el texto básico el tema Distribución de frecuencias y el documento 2 en el que se encuentran algunos ejemplos.

Observe también el video en el que se indica paso a paso el proceso de construir distribuciones de frecuencia.

Ejemplos adicionales se encuentran en “Documento 2”.

Resumen

Vamos a hacer un resumen

Para elaborar una distribución de frecuencias los pasos recomendados son los siguientes.

1. Ordenar los valores y encontrar el máximo, mínimo y rango.
2. Determinar el número de clases con la fórmula 2k ≥ n.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

3. Calcular el tamaño de clase, la fórmula utilizada es

$$\textbf{Tamaño de clase} = \frac{\text{Rango}}{k}$$

4. Elaboración de la tabla de frecuencias.

Actividad

Utilizando Tableau construya una tabla de frecuencias para la variable “marca” del teléfono de archivo de datos “encuesta”

2.2. Distribución de frecuencias relativas

Antes de empezar esta sección es necesario entender el término “relativo”, según la real academia significa “que guarda relación con alguien o con algo.”. En efecto, cuando hablamos de frecuencias relativas estamos hablando de ellas de una forma que permite que se comparen con las demás, para esto debemos expresarlas en porcentaje; de esa manera, el porcentaje que tenga cada clase nos dirá su tamaño respecto a las demás clases. En el Documento 2 se muestran ejemplos desarrollados del procedimiento, siga paso a paso los ejercicios y tome nota de los aspectos importantes.

Una explicación detallada de cómo obtener una frecuencia relativa y de cómo realizar la respectiva gráfica se halla en el siguiente video.

2.3. Representación gráfica de una distribución de frecuencias

Una vez usted tenga lista su tabla de distribución de frecuencias, puede dar un paso más para presentar los resultados de una mejor manera. En realidad este es un trabajo muy sencillo y hay que diferenciar dos tipos de gráficos

Diagrama de barras. Utilizado para representar datos cualitativos

Histograma. Utilizado para representar datos cuantitativos.

Siga paso a paso el ejemplo del "Documento 2" en donde se construye una tabla de frecuencias de las marcas de teléfono. Tome nota de los aspectos que considere importantes.

Observe también el siguiente video, en donde se explica paso a paso la construcción de las distintas gráficas.

Resumen

Para obtener las frecuencias relativas hay que seguir los siguientes pasos.

1. Agregamos una columna a la tabla de distribución de frecuencias.
2. Dividimos el valor de la frecuencia absoluta simple para el total de observaciones, este valor es decimal y menor que uno.
3. Para expresar en porcentaje este valor lo multiplicamos por 100.
4. Una distribución de frecuencias se puede representar con un diagrama de barras y con un histograma.

Actividad

Utilizando Tableau construya un histograma para la variable "edad" del archivo de datos "encuesta".

Corra línea a línea el código en R del archivo "histograma.r" y observe las gráficas resultantes.

Actividades de aprendizaje recomendadas

En este sitio web, observe los videos siguientes Introducción, La interfaz de Tableau, Conexión a los datos, Análisis visual.

Autoevaluación 2

Desarrolle la autoevalución; le permitirá determinar el nivel de aprovechamiento del tema estudiado.

Escriba Verdadero (V) o Falso (F) a las siguientes afirmaciones, según corresponda:

1. ( ) El rango se obtiene restando el valor mínimo del valor máximo de una variable.
2. ( ) Un intervalo de clase equivale a la cantidad de datos de una clase.
3. ( ) En una distribución de frecuencias intervienen siempre un número impar de clases.
4. ( ) Una frecuencia se calcula obteniendo el valor máximo de una variable.
5. ( ) Para calcular el número de clases se utiliza la formula k^2.
6. ( ) El tamaño de la clase se calcula con la fórmula rango/N.
7. ( ) El histograma se utiliza para graficar datos cuantitativos.

8. ( ) El diagrama de barras se utiliza para representar datos cuantitativos.

9. ( ) La frecuencia relativa expresa la cantidad de observaciones de forma porcentual.

10. ( ) La frecuencia relativa acumulada va acumulando los porcentajes de cada categoría.

Ir al solucionario

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Resultado de aprendizaje 3 | Describe la estructura de un conjunto de datos a través de sus medidas de tendencia central y de dispersión.

Contenidos, recursos y actividades de aprendizaje

Semana 3

Unidad 3. Medidas de tendencia central y medidas de dispersión

En esta sección ponemos especial énfasis en el resultado de aprendizaje "Describe la estructura de un conjunto de datos a través de sus medidas de tendencia central y de dispersión", planificado para el primer bimestre. Si podemos calcular las medidas centrales y las de dispersión de un conjunto de datos, entonces podemos describir de forma acertada ese conjunto de datos. Hasta ahora hemos levantado un conjunto de datos, los hemos organizado, representado en una tabla y a través de un gráfico. Ahora agregaremos información que permite describir aún más el conjunto de datos. Una forma básica y siempre necesaria para describir los datos es la información que indique el centro de esos datos, el centro puede indicarse a través de la media, la mediana o la moda.

3.1. Media

Lea el "Documento 3" centrándose en el apartado que se refiere a la media aritmética, tome nota de los aspectos más importantes y revise con atención los ejemplos que se plantean.

Ir a recursos

Observar y tomar nota del video.

3.2. Mediana

Lea el "Documento 3" centrándose en el apartado que se refiere a la mediana preste especial atención a los casos en los que se prefiere la mediana en lugar de la media aritmética.

Observar y tomar nota del video

3.3. Moda

La moda representa al valor que más se repite en un conjunto de datos. Revise el "Documento 3" tomando nota de lo encontrado.

Observar y tomar nota del video

Actividades de aprendizaje recomendadas

- Resuelva los ejercicios del 1 al 12 del texto básico, estos se encuentran en la sección correspondiente a la media aritmética.
- Resuelva los ejercicios del 17 al 24 del texto básico, estos se encuentran en la sección correspondiente a la moda.

Autoevaluación 3

Desarrolle la autoevalución; le permitirá determinar el nivel de aprovechamiento del tema estudiado.

Escriba Verdadero (V) o Falso (F) a las siguientes afirmaciones, según corresponda:

1. ( ) La diferencia entre media muestral y media poblacional está en la fórmula que se aplica.
2. ( ) La media poblacional es un estadístico.
3. ( ) La mediana se utiliza en lugar de la media cuando la dispersión es demasiado alta.
4. ( ) La moda equivale al valor que más repita dentro de un conjunto de datos.
5. ( ) Para calcular la mediana se requiere que los valores estén ordenados.
6. ( ) La mediana es la posición central de un conjunto de datos.
7. ( ) La mediana y media siempre tienen el mismo valor.
8. ( ) La medida más utilizada es la media aritmética.

9. ( ) Se puede decir que la media y la moda son comparables.

10. ( ) El promedio de edad de los ecuatorianos no es un estadístico, sino un parámetro.

Ir al solucionario

Semana 4

3.4. Medidas de dispersión

Es muy importante leer la sección ¿Por qué estudiar la dispersión?, para entender qué es y por qué es importante.

Las medidas de dispersión que más se utilizan son rango, desviación media, varianza, desviación estándar.

Lea el "Documento 4**",** en él encontrará el detalle del contenido referente a esta unidad. Céntrese en primer lugar en determinar para qué sirve la dispersión.

Ir a recursos

Observar y tomar nota del video.

Resumen

1. Las medidas que nos señalan el centro de los datos son la media, la mediana y la moda.
2. Las medidas que nos indican la dispersión de los datos son la varianza, desviación estándar y el rango.
3. El rango es la medida más sencilla.
4. La regla empírica muestra cómo se distribuyen los valores bajo la curva normal utilizando la desviación estándar como medida.

Actividades de aprendizaje recomendadas

- Resuelva los ejercicios del 41 al 46 del texto básico, estos se encuentran en la sección correspondiente a la dispersión.
- Resuelva también los ejercicios 54 y 55, estos se encuentran en la sección correspondiente al teorema de Chevyshe(V)

Autoevaluación 4

Desarrolle la autoevalución; le permitirá determinar el nivel de aprovechamiento del tema estudiado.

Escriba Verdadero (V) o Falso (F) a las siguientes afirmaciones, según corresponda:

1. (   ) El rango es una medida de dispersión.
2. (   ) La desviación estándar nos dice que tan dispersos están los datos.
3. (   ) Datos dispersos significa que los valores están distribuidos en todo el rango.
4. (   ) Si la mayoría de los datos están cerca de la media aritmética la dispersión es alta.
5. (   ) La varianza es más efectiva que el rango a la hora de calcular la dispersión.
6. (   ) Dispersión equivale a la distribución de los valores dentro de un rango.
7. (   ) La desviación estándar se calcula Con base en las diferencias entre cada valor y la media aritmética.
8. (   ) Dados los valores 4, 5, 6, la desviación estándar es 1.

9. (   ) Dados los siguientes valores 75, 78, 23, 56 el rango es 55.

10. (   ) El rango si eleva al cuadrado las desviaciones de los valores respecto de la media.

Ir al solucionario

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Resultado de aprendizaje 2 | Representa datos utilizando gráficas.

Contenidos, recursos y actividades de aprendizaje

Semana 5

Unidad 4. Descripción de datos

Siempre se ha dicho que una imagen puede expresar más que mil palabras y esto no está alejado de la verdad; cuando se trata de datos, estos se pueden describir utilizando distintas gráficas. En esta sección ponemos especial énfasis en el indicador de aprendizaje “Describe datos utilizando gráficas” planificada para el primer bimestre.

Apreciado(a) estudiante, esta unidad no tiene mayor complejidad, le animo a que la desarrolle y aprenda a representar gráficamente conjuntos de datos. En caso de tener dudas, escríbame o llame, estaré presto para ayudarle en su proceso de aprendizaje.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

4.1. Diagramas de puntos

Lea el “Documento 4” y tome nota de cómo construir un diagrama de puntos.

Se recomienda leer la sección “Diagramas de puntos” del texto básico y observar el diagrama del ejemplo que se desarrolla.

4.2. Gráficas de tallo y hojas

Un diagrama de tallo y hojas es muy fácil de construir. Le recomiendo leer la sección correspondiente del “Documento 4”.

4.3. Cuartiles

Los cuartiles son posiciones dentro de un conjunto de datos. Por ejemplo, si tenemos un conjunto de 100 valores, los cuartiles serán la posición 25, la posición 50, la posición 75 y la posición 100. Note que estamos refiriéndonos a la posición, en otras palabras el cuartil va a ser cualquier valor que se encuentre en esa posición.

Los cuartiles dividen al conjunto ordenado de datos en cuatro partes.

Observe el siguiente video en donde se explica paso a paso el proceso de cálculo.

4.4. Deciles

Los deciles dividen al conjunto de observaciones en 10 partes, los conceptos son los mismos que se aplican para el cálculo de cuartiles, existen pequeñas diferencias en las fórmulas que se aplican.

Observe el siguiente video en donde se explica paso a paso el proceso de cálculo.

4.5. Percentiles

Con los percentiles ocurre algo similar, dividen el conjunto de datos en 100 partes. El proceso para calcular cada uno es

Primer percentil P1 = (n+1) (1/100)

Segundo percentil P2 = (n+1) (2/100)

Tercer percentil P3 = (n+1) (3/100)

Y así sucesivamente para cada percentil del conjunto de datos.

Observe el siguiente video en donde se explica paso a paso el proceso de cálculo.

4.6. Descripción de la relación entre dos variables

Este es un tema muy sencillo, le recomiendo leer la sección correspondiente en el texto básico y observar los ejemplos que se plantean.

En el "Documento 4" se presenta un ejemplo en el que se relaciona la edad en meses y el peso en libras, revíselo.

Resumen

- Los diagramas de puntos se utilizan para representar de forma gráfica un conjunto de datos, tienen la característica de resumir los datos sin perder información.

- Un diagrama de caja permite apreciar la dispersión de los datos.
- Tanto los deciles, como los cuartiles y los percentiles son medidas de dispersión de los datos.

Actividades de aprendizaje recomendadas

- Resuelva los ejercicios del 11 al 14 del texto básico, estos se encuentran al final de la sección que trata “Otras medidas de dispersión”.
- Con los cuartiles se puede desarrollar diagramas de caja. Vamos a complementar este capítulo con una lectura de la sección “Diagramas de caja” del texto básico. De la lectura debe obtener lo siguiente.
    - Observar atentamente los dos ejemplos que se plantean.
    - Determinar para qué se utilizan los diagramas de caja.
- Resuelva los ejercicios 23, 24 y 25 del texto básico, estos se encuentran al final de la sección que trata “Descripción de la relación entre dos variables”.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Autoevaluación 5

Desarrolle la autoevalución; le permitirá determinar el nivel de aprovechamiento del tema estudiado.

Escriba Verdadero (V) o Falso (F) a las siguientes afirmaciones, según corresponda:

1. (   ) Un diagrama de punto ofrece más información que una tabla de distribución de frecuencias.
2. (   ) Un diagrama de tallo y hojas permite condensar los datos de las observaciones sin perder información.
3. (   ) Los cuartiles dividen en 10 partes a un conjunto de observaciones.
4. (   ) Cuando se calcula un decil, no se determina el valor, sino la posición en la que está el valor que nos interesa.
5. (   ) Un percentil cualquiera es en realidad una posición.
6. (   ) Los diagramas de caja permiten determinar la dispersión de los datos.
7. (   ) Un diagrama de caja se diagrama desde el primer al tercer cuartil.
8. (   ) El cuartil 1 es equivalente al percentil 25.

9. ( ) El decil 4 es equivalente al percentil 50.

10. ( ) El percentil 100 es equivalente al cuarto cuartil.

Ir al solucionario

Índice

Primer bimestre

Segundo bimestre

Solucionario

Referencias bibliográficas

Recursos

Resultado de aprendizaje 4 y 5

- Calcula probabilidades de sucesos de variables aleatorias.
- Diferencia entre sucesos dependientes e independientes.

Contenidos, recursos y actividades de aprendizaje

Semana 6

Unidad 5. Probabilidades y distribución de probabilidad discreta

En esta sección ponemos especial énfasis en el indicador de aprendizaje “Calcula probabilidades de sucesos de variables aleatorias” planificada para el primer bimestre. Se trata de procedimientos sencillos que lo invito a revisar y en caso de tener dudas, escríbame sus inquietudes y procederé a apoyarle en su proceso de aprendizaje.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Para tratar este tema, le recomiendo leer la sección "¿Qué es probabilidad?" en el texto básico, ponga énfasis en entender el concepto de probabilidad y, si es necesario, plantear un par de ejemplos. Tome nota de los conceptos de experimento, evento y resultado.

5.1. ¿Qué es la probabilidad?

La probabilidad de un suceso está asociado a cuán posible es que ocurra o no ocurra, en esta unidad vamos a estudiar las probabilidades y cómo se distribuyen.

El "Documento 5" aporta también con algunas ideas al respecto, por lo que recomiendo revisarlo y tomar nota.

Ir a recursos

Observe el video y tome nota de los conceptos.

5.2. Tipos de probabilidad

Al igual que en la sección anterior, aquí es conveniente que usted empiece dando una lectura al texto básico en la sección que corresponde.

El tema se explica también en el "Documento 5".

5.3. Reglas para calcular probabilidades

Para iniciar este tema le recomiendo leer la sección correspondiente en el texto básico. En cada caso vaya tomando nota y desarrollando

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

nuevamente los ejemplos que se presentan, de esta forma usted podrá relacionar los conceptos con las fórmulas y con las situaciones del diario vivir.

5.3.1. Reglas de adición

Revisar los ejercicios y explicaciones dadas en el "Documento 5".

Regla especial de la adición

P(A o B) = P (A) + P (B)

Regla general de la adición

P(A o B) = P (A) + P (B) – P (A y B)

5.3.2. Reglas de multiplicación

Revisar los ejercicios y explicaciones dadas en el "Documento 5".

Regla especial de multiplicación

P (A y B) = P(A) * P(B)

Regla general de multiplicación

P (A y B) = P(A) * P(B | A)

5.4. Distribuciones de probabilidad discreta

Esta sección se basa en el capítulo 6 del texto básico, y se centra en distribuir los resultados de un experimento de forma tal que estos puedan observarse como probabilidades. Una distribución de probabilidad, entonces, muestra todos los resultados posibles de un experimento y la probabilidad de que cada resultado ocurra.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Es importante que de lectura al “Documento 5” que lo guía en lo que tiene que ver con distribución de probabilidad, variables aleatorias y en el cálculo de media y desviación estándar de una distribución de probabilidad.

Actividades de aprendizaje recomendadas

- Resuelva los ejercicios del 1 al 6 del texto básico, estos se encuentran al final de la sección que trata “Media, varianza y desviación estándar de una distribución de probabilidad discreta”.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Autoevaluación 6

Desarrolle la autoevalución; le permitirá determinar el nivel de aprovechamiento del tema estudiado.

Escriba Verdadero (V) o Falso (F) a las siguientes afirmaciones, según corresponda:

1. ( ) El concepto de probabilidad hace referencia a la cuantificación de un evento que pudiera presentarse o no.
2. ( ) La probabilidad se puede calcular a través del cociente entre los resultados posibles para los resultados favorables a un evento.
3. ( ) Se dice que dos o más eventos resultan ser mutuamente excluyentes cuando la presencia de uno impide que otro se presente al mismo tiempo
4. ( ) La probabilidad empírica también se conoce como probabilidad relativa ya que representa la fracción de eventos similares que sucedieron en el pasado.
5. ( ) Un evento es el conjunto de uno o más resultados de un experimento.
6. ( ) Existen tres enfoques para asignar probabilidades objetivo, subjetivo y clásico.
7. ( ) La probabilidad clásica parte del supuesto de que los resultados de un experimento son igualmente posibles.

8. ( ) En las reglas de multiplicación se estima la probabilidad de que la ocurrencia de dos eventos sea simultanea

9. ( ) En la regla general de la adición los eventos deben ser mutuamente excluyentes.

10. ( ) La regla del complemento se emplea para determinar la probabilidad de que un evento ocurra restando de 1 la probabilidad de un evento que no ha ocurrido.

Ir al solucionario

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Índice

Primer bimestre

Segundo bimestre

Solucionario

Referencias bibliográficas

Recursos

Semana 7

Unidad 1 a 3

Repaso de los contenidos del primer bimestre

Semana 8

Unidades 4 y 5

Repaso de los contenidos del primer bimestre

Segundo bimestre

Resultado de aprendizaje 6

Aplica la distribución normal a conjuntos de datos y calcula intervalos de confianza y áreas bajo la curva normal.

Contenidos, recursos y actividades de aprendizaje

Semana 9

Unidad 6. Distribución de probabilidad normal

En esta sección ponemos especial énfasis al resultado de aprendizaje “Aplica la distribución normal a conjuntos de datos y calcula áreas bajo la curva normal” planificado como parte de este primer bimestre; le recomiendo leer inicialmente el “Documento

6” y el texto a partir de la sección “La familia de distribuciones de probabilidad normal”.

Ir a recursos

Inicialmente es necesario entender que es distribución normal y algunas de sus características, le recomiendo, por tanto, observar el video y tomar nota de los aspectos más importantes.

6.1. Distribución normal y familia de distribuciones de probabilidad normal.

Dé lectura al “Documento 6”, revise detenidamente los conceptos y ejemplos; el objetivo es determinar ¿qué es una familia de curvas normales?, ¿qué es el valor z?, ¿cómo se calcula z?, ¿cómo se busca z en la tabla de distribución normal?

6.2. Cálculo de áreas bajo la curva normal

Si ha revisado la sección anterior, este subtema es bastante sencillo, debe revisar el documento 6 y determinar cómo calcular un área bajo la curva, el proceso es sencillo; si tiene dudas en el siguiente video puede revisar cada uno de los pasos.

El “Documento 6” tiene varios ejercicios de distinto tipo, le recomiendo que los revise todos, ponga especial atención a la gráfica en donde se señala el área que se está calculando, comprender esto es vital para entender las operaciones de sumas o restas de áreas.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Resumen

1. Una distribución normal tiene el 50% a la izquierda de la media y el 50% a la derecha de la media.
2. Hay una regla empírica que señala que el 99,7% se encuentra a tres desviaciones estándar antes y después de la media.
3. Las diferentes curvas normales se estandarizan convirtiendo el rango de fluctuación de la variable a desviaciones estándar.
4. Las áreas que están bajo la curva se pueden determinar utilizando una tabla de distribución normal, para encontrar el área es necesario contar con el valor de z.

Actividades de aprendizaje recomendadas

- Resuelva los ejercicios del 7 al 12 del texto básico, estos se encuentran al final de la sección que trata "Distribución de probabilidad normal estándar".
- Resuelva los ejercicios del 23 al 29 del texto básico, estos se encuentran al final de la sección "Determinación de áreas bajo la curva".

Autoevaluación 7

Desarrolle la autoevalución; le permitirá determinar el nivel de aprovechamiento del tema estudiado.

Escriba Verdadero (V) o Falso (F) a las siguientes afirmaciones, según corresponda:

1. (   ) La distribución de probabilidad normal es simétrica respecto a la mediana.
2. (   ) La distribución de probabilidad normal tiene forma de campana.
3. (   ) Una variable con distribución normal puede adoptar una infinidad de valores dentro de un rango específico.
4. (   ) La distribución normal se aplica a valores numéricos continuos.
5. (   ) La variable marca de un vehículo puede ser representada con una distribución normal.
6. (   ) La distribución normal estándar en realidad representa a una familia de distribuciones.
7. (   ) El valor de z representa la media aritmética.
8. (   ) La regla general dice que 68 de cada 100 observaciones se encuentran a dos desviaciones estándar antes y después de la media.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

9. (   ) el 100% de las observaciones se encuentran a 3 desviaciones estándar antes y después de la media.

10. (   ) La tabla de distribución normal en las intersecciones de fila y columna contienen áreas en porcentaje.

Ir al solucionario

Índice

Primer bimestre

Segundo bimestre

Solucionario

Referencias bibliográficas

Recursos

Resultado de aprendizaje 7	Determinar tamaños adecuados de muestra para poblaciones y para proporciones.

Contenidos, recursos y actividades de aprendizaje

Semana 10

Unidad 7. Muestreo

En esta unidad usted aprenderá a determinar una muestra, su tamaño y un método para obtenerla, también comprenderá los conceptos necesarios que señalan de qué manera una muestra representa fielmente a una población. El objetivo es resultado de aprendizaje “Entender qué es muestrear y en qué situación se aplica un tipo determinado tipo de muestreo”. Le invito a que revise cuidadosamente el contenido y vaya tomando nota de los aspectos importantes que encuentre. En caso de tener dudas le recuerdo que puede contar con su tutor a través de una consulta en el Entorno Virtual de Aprendizaje.

Para empezar, le invito a revisar el “Documento 7” y seguir las indicaciones que se dan allí.

Ir a recursos

7.1. Razones para muestrear

> En el apartado “Métodos de muestreo” del texto básico se señalan cinco razones de peso que le solicito lea, es muy importante que tenga claro por qué es más conveniente levantar información de un grupo de elementos que de todo el universo.

7.2. Tipos de muestreo

En el texto básico se señalan cuatro tipos de muestreo, lea detenidamente cada uno de los tipos y proponga dos ejemplos de cada uno; en los ejemplos asegúrese de que exista una razón que justifique el uso de ese tipo de muestreo.

7.3. Error de muestreo

Para empezar este apartado le recomiendo leer en el “Documento 7” el tema referente a “Error de muestreo”. Determine qué es el error de muestreo y cómo se calcula.

7.4. Distribución de probabilidad

Aquí nos vamos a centrar en definir y entender qué es, cómo se obtiene y qué nos indica una distribución de probabilidad. En el "Documento 8" se señala el concepto de distribución de probabilidad y de las características de una distribución de probabilidad. Se recomienda revisar también los ejemplos descritos.

Ir a recursos

Resumen

Muestrear es un proceso por el que se selecciona una parte del total de elementos de una población. Se escoge solo una parte porque preguntar o encuestar a todos supone mayores costos y tiempo.

Los resultados de analizar solo una muestra implican que las diferencias con todo el universo (población) no van a ser significativas; es decir, si se pregunta al número adecuado de elementos de la muestra, sus resultados van a ser bastante semejantes a los de la población.

Actividades de aprendizaje recomendadas

- Resolver los ejercicios 1, 2 y 3 que se refieren a muestreo
- Asegúrese de tener claro el concepto de error de muestreo. Proponga un ejemplo.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Autoevaluación 8

Desarrolle la autoevalución; le permitirá determinar el nivel de aprovechamiento del tema estudiado.

Escriba Verdadero (V) o Falso (F) a las siguientes afirmaciones, según corresponda:

1. (   ) La distribución de probabilidad lista todos los posibles resultados de un experimento.
2. (   ) Si se tiene en una bolsa 4 bolitas (negra, blanca, roja y azul), al extraer solo una, la probabilidad de que sea blanca es ¼.
3. (   ) Si se desea conocer cómo va un candidato a la alcaldía de la ciudad para plantear sus estrategias de campaña, es mejor encuestar a la muestra en lugar de a toda la población.
4. (   ) Muestreo es un procedimiento que calcula un dato.
5. (   ) Error de muestreo es la diferencia que existe entre la media de la población y una medida de tendencia central anterior de toda la población.
6. (   ) Si se desea conocer la simpatía que tiene un candidato a presidente se puede estimar una medida de tendencia central.
7. (   ) Si se desea conocer el salario promedio de los ciudadanos de una región se puede estimar una proporción.

8. ( ) Un experimento puede tener solamente tres posibles resultados, listar estos resultados y la probabilidad que tenga cada uno se denomina tabla de distribución de probabilidad.

9. ( ) Muestra representativa significa que debe representar a cada sector de la población.

10. ( ) La diferencia entre estimar una media y estimar una proporción está en el fenómeno que se está investigando, si se desea conocer un valor absoluto o un porcentaje.

Ir al solucionario

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Resultado de aprendizaje 8 | Aplica el teorema del límite central.

Contenidos, recursos y actividades de aprendizaje

Semana 11

7.5. Tamaño de una muestra

El tamaño de la muestra es la cantidad de elementos a los que se debe estudiar para tener una estimación adecuada de una población. Los conceptos necesarios para abordar esta unidad se encuentran en el "Documento 7", por lo que le recomiendo que lo revise en la parte pertinente

7.5.1. Tamaño de la muestra para calcular una media poblacional

En este caso se busca determinar un valor numérico que estime una medida de la población; es decir, se busca variables como la edad, el peso, el salario, etc., de una población. Revise el ejercicio del Documento 7 En una ciudad pequeña se desea determinar el salario medio de los trabajadores. El error máximo que se va a permitir es de 100 dólares y el nivel de confianza de 95%. Se conoce que en estudios anteriores aplicados en la ciudad la desviación estándar fue 120 ¿Cuántas personas debe abarcar la muestra?

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

7.5.2. Tamaño de la muestra para calcular una proporción de una población

Siempre que se hace levantamiento de opinión pública (¿qué opina de esto?, o ¿cuál es su candidato favorito?, etc.) es necesario determinar una proporción de la población que se utilice como muestra con la que se trabaje. Revise el Documento 7 y analice el ejemplo que se presenta y tome nota de la fórmula que se aplica.

Actividades de aprendizaje recomendadas

Resolver los ejercicios 19 a 24 que se plantean al final de la sección "Tamaño de la muestra para calcular la proporción de una población" del texto básico.

Autoevaluación 9

Desarrolle la autoevalución; le permitirá determinar el nivel de aprovechamiento del tema estudiado.

Escriba Verdadero (V) o Falso (F) a las siguientes afirmaciones, según corresponda:

1. ( ) Una muestra no siempre debe ser representativa.
2. ( ) En la fórmula para calcular el tamaño de una muestra para una media poblacional, z equivale al nivel de confianza.
3. ( ) En la fórmula para calcular el tamaño de una muestra para una media poblacional, E se refiere a un valor que no puede pasar del 5%.
4. ( ) Si se desea conocer a cuantas personas se debe encuestar para medir las preferencias de un candidato se debe aplicar la fórmula para calcular el tamaño de la muestra para la media.
5. ( ) En la fórmula para calcular el tamaño de una muestra para una proporción el valor de p equivale a 0,5.
6. ( ) El tamaño de una muestra depende del tamaño de la población.

7. (   ) Dado el ejemplo de conocer las preferencias de los ciudadanos para la elección de alcalde de la ciudad, una muestra es representativa si se obtienen todas las encuestas en el sector central de la ciudad

8. (   ) Al calcular el tamaño de la muestra para una proporción, los valores de p y q deben valer 0,5 para maximizar el tamaño de la muestra.

9. (   ) Si se desea conocer el tamaño de una muestra para calcular el salario promedio de los trabajadores de una ciudad se tiene que aplicar la fórmula para calcular el tamaño de una muestra para una proporción.

10. (   ) En una ciudad pequeña se desea determinar el salario medio de los trabajadores. El error máximo que se va a permitir es de 140 dólares y el nivel de confianza de 95%. Se conoce que en estudios anteriores aplicados en la ciudad la desviación estándar fue 100. El tamaño adecuado para esa muestra es 700.

Ir al solucionario

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Semana 12

Unidad 8. Distribución muestral y teorema central del límite

En esta sección vamos a trabajar en comprender en qué consiste la distribución de la media muestral y el teorema central del límite, tener claro este tema permite comprender por qué el muestreo es representativo de una población y por qué a mayor tamaño de la muestra mayor tendencia a una forma normal para cualquier tipo de población. Se trata de un tema interesante que se desglosa en el "Documento 8", como siempre, en caso de tener dudas cuente con su tutor para resolverlas.

8.1. Distribución muestral de la media

Revise el "Documento 8" en la parte correspondiente y siga las instrucciones allí dadas, Es necesario entender el ejercicio de distribución de las muestras y hacer una comparación de las gráficas resultantes de la distribución de frecuencias y la distribución de la media muestral. Entender esta diferencia es lo más importante en esta sección.

En el “Documento 8” se presentan algunos ejemplos, se sugiere revisar todos ellos y de ser necesario acudir al texto básico para ahondar.

8.2. Teorema central del límite

Este teorema señala que si de una población se toman todas las muestras posibles de un determinado tamaño, la distribución muestral de la media aritmética presenta una tendencia a asemejarse a una distribución normal. Esta semejanza es mayor mientras mayor sea el tamaño de la muestra. Esto se profundiza en el “Documento 8”. Revise los ejemplos que se plantean.

Resumen

- La distribución muestral de la media consiste en todas las posibles medias de todas las muestras de determinado tamaño.
- La media de las medias muestrales es igual a la media poblacional.
- La distribución muestral de la media tiene forma de campana y se perfecciona conforme aumenta el tamaño de la muestra.
- El teorema central del límite señala que si se toman todas las muestras de un determinado tamaño, la distribución muestral de la media se aproxima a una forma normal conforme el tamaño de la muestra aumenta.
- Para cubrir los contenidos de este capítulo requiere conocer histogramas, medias poblacionales y muestrales, tablas de frecuencia, frecuencias y frecuencias relativas.

Actividades de aprendizaje recomendadas

- Resolver los ejercicios que se encuentran en el texto básico en las páginas 232 y 233.
- Resolver los ejercicios 13 y 14 de la página 239 del texto básico. Los procedimientos que se aplican son los mismos que se han aplicado en los ejercicios anteriores de esta guía.

Autoevaluación 10

Desarrolle la autoevalución; le permitirá determinar el nivel de aprovechamiento del tema estudiado.

Escriba Verdadero (V) o Falso (F) a las siguientes afirmaciones, según corresponda:

1. ( ) A mayor tamaño de la muestra menor nivel de error muestral.
2. ( ) La distribución de la media muestral tiende a mostrarse como una campana de forma normal.
3. ( ) Dado el siguiente conjunto de datos 4, 6, 5, 7, 4, 6. Si se toman todas las muestras posibles de tamaño 3 la distribución de la media muestral adopta una forma normal perfecta.
4. ( ) Dado el siguiente conjunto de datos 4, 6, 5, 7, 4, 6. Si se toman todas las muestras posibles de tamaño 3 la distribución de la media muestral adopta una forma normal mejor que la forma que adopta la distribución para todas las muestras posibles de tamaño 4.
5. ( ) Dado el siguiente conjunto de datos 4, 6, 5, 7, 4, 6. Si se va a desarrollar la distribución de la media muestral y se busca que adopte la mejor forma normal posible, es mejor trabajar con muestras de tamaño 2 en lugar de muestras de tamaño 4.

6. (   ) Si no conoce cuál es la forma de la distribución de una población, o sabe que esta no es normal, lo recomendable es tomar una muestra grande.

7. (   ) Mientras más grande es la muestra, tiende menos a una distribución normal.

8. (   ) La media de la distribución muestral de medias va a ser exactamente igual a la media de la población solo cuando se toman todas las muestras de tamaño n.

9. (   ) Si no se toman todas las muestras de tamaño n, la media de la distribución muestral de medias va tener una gran diferencia con la media poblacional.

10. (   ) El teorema central del límite señala que la distribución muestral de la media tiende a una forma normal conforme aumenta el tamaño de la muestra.

Ir al solucionario

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Resultado de aprendizaje 9 | Verifica hipótesis

Contenidos, recursos y actividades de aprendizaje

Semana 13

Unidad 9. Hipótesis y pruebas de hipótesis

En esta unidad aprenderemos a verificar hipótesis siguiendo cada uno de los pasos. Para esto, es necesario entender antes una serie de conceptos que complementan. Para iniciar es necesario que trabaje el "Documento 9" y el texto básico según se indique.

Ir a recursos

Al terminar la lectura asegúrese de tener claros los conceptos hipótesis, prueba de hipótesis, pasos para resolver una hipótesis, hipótesis nula, hipótesis alternativa, nivel de significancia y estadístico de prueba. Estos conceptos ya se encuentran desarrollados en el texto y es necesario conocerlos en detalle. Quiero motivarle a que en caso de tener dudas me escriba, mi trabajo es ayudarle a lograr sus objetivos de aprendizaje y por lo tanto estaré presto a responder sus dudas.

9.1. Pasos para verificar una hipótesis

Los pasos para verificar una hipótesis se señalan claramente en el texto básico y se puede observar un ejemplo que se va desarrollando conforme se van explicando los pasos. El "Documento 9" le guiará paso a paso para lograr los objetivos que básicamente consisten en conocer los pasos que se deben seguir en el proceso de validación de hipótesis. Preste mucha atención a cómo se plantean hipótesis.

9.2. Operadores de la hipótesis alternativa H1

En el "Documento 9" se presenta una tabla con los operadores que se utilizan en las hipótesis, analice detenidamente esta tabla y revísela siempre que se presente un ejemplo de hipótesis nula y alternativa.

Al finalizar deberá estar en capacidad de plantear correctamente los operadores relacionales de las hipótesis.

Actividades de aprendizaje recomendadas

- Observar el video para resolver dudas que puedan existir, en este video se presentan más ejercicios y proceso para resolverlos.
- Del libro de texto resolver los ejercicios 5, 6, 7 y 8 que se encuentran en la sección correspondiente. El enunciado del ejercicio 5 inicia así “El fabricante de neumáticos radiales con cinturón de acero x-15 para camiones señala que el millaje medio que cada uno recorre antes de que se desgasten las cuerdas es de 60.000 millas…”.

Autoevaluación 11

Desarrolle la autoevalución; le permitirá determinar el nivel de aprovechamiento del tema estudiado.

Escriba Verdadero (V) o Falso (F) a las siguientes afirmaciones, según corresponda:

1. (   ) Las hipótesis se refieren siempre a un estadístico de la muestra y nunca a un parámetro de la población.
2. (   ) Verificar la hipótesis consiste en determinar si la hipótesis es una afirmación razonable.
3. (   ) El nivel de significancia es el error que se está dispuesto a permitir aceptando.
4. (   ) En el paso 3 para verificar una hipótesis se utiliza z cuando se desconoce la desviación estándar de la población.
5. (   ) La regla decisión es un pregunta acerca de un rango de valores, aceptando H0 cuando el valor de Z está dentro de ese tango de valores.
6. (   ) Un valor z puede ser calculado o puede ser buscado en la tabla. El valor crítico se lo calcula.
7. (   ) La prueba de una cola implica que el error se ubica en los dos lados de la curva.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

8. ( ) La prueba de dos colas requiere que el nivel de significancia se distribuya.

9. ( ) Dado el siguiente enunciado "… determine si el volumen de ventas se ha incrementado". En la hipótesis nula se debe colocar el operador >=

10. ( ) Dado el siguiente enunciado "… se desea conocer si el monto de ventas inicial es diferente del monto de ventas final". En la hipótesis nula se debe colocar el operador =

Ir al solucionario

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Semana 14

9.3. Verificación de hipótesis cuando se conoce la desviación estándar de la población

Si se conoce la desviación estándar de la población se utiliza el estadístico Z. Esto se amplía en el "Documento 9". Al finalizar la sección debe tener claro el procedimiento a seguir cuando conoce la desviación estándar de la población.

9.4. Verificación de hipótesis cuando NO se conoce la desviación estándar de la población

Hay ejercicios en los que ya no se utiliza el estadístico Z, sino el estadístico t. Esto se debe a que se desconoce la desviación estándar de la población.

En esos casos e procedimiento es similar y solo se cambia los valores z por valores t. Para más detalles revise el "Documento 9" en la parte pertinente.

Actividades de aprendizaje recomendadas

- Observar el video para resolver dudas que puedan existir, en este video se presentan más ejercicios y proceso para resolverlos.
- Lea la teoría respectiva de la temática Prueba de Hipótesis de una muestra.
- Resolver los ejercicios 1 al 4 de las páginas 293 y 294.
- Resolver los ejercicios 9 al 14 de la página 298 y 299.

Autoevaluación 12

Desarrolle la autoevalución; le permitirá determinar el nivel de aprovechamiento del tema estudiado.

Escriba Verdadero (V) o Falso (F) a las siguientes afirmaciones, según corresponda:

1. ( ) La hipótesis se refiere a un parámetro poblacional.
2. ( ) En la verificación de hipótesis se compara siempre un parámetro de la población con un estadístico de una muestra.
3. ( ) El estadístico Z se utiliza cuando se conoce la desviación estándar de la población.
4. ( ) Nivel de significación equivale al nivel de confianza que se tiene.
5. ( ) Un error de tipo I equivale al nivel de significancia.
6. ( ) El valor del estadístico de prueba no se calcula, se lo obtiene directamente de la tabla con la información de la muestra.
7. ( ) Un error de tipo 2 equivale a rechazar la hipótesis nula cuando esta es falsa.
8. ( ) La regla de decisión nos indica cuando aceptar o rechazar la H0.

9. ( ) El valor crítico se calcula aplicando una fórmula.

10. ( ) Si en el enunciado de un problema se pregunta si “es mayor que” se debe colocar > en H1.

Ir al solucionario

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Semana 15

Unidades correspondientes al segundo bimestre

Semana 16

Unidades correspondientes al segundo bimestre

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

4. Solucionario

Supongamos que se tiene la siguiente encuesta
En base a la siguiente encuesta conteste las preguntas: 1. Edad _____ 2. Sexo M ( ) F 3. Tiene teléfono móvil SÍ ( ) NO( ) 4. Cuánto gasta al mes en telefonía móvil ________ 5. ¿Cómo se calificaría usted mismo? Escribo muchos mensajes al día. Escribo pocos mensajes al día. Casi no escribo mensajes.

Autoevaluación 1		
Pregunta	**Respuesta**	**Retroalimentación**
1.	F	La pregunta 1 utiliza una variable que se mide a nivel de intervalo.
2.	F	La pregunta 1 utiliza una variable que se mide a nivel ordinal.
3.	V	La pregunta 2 utiliza una variable que se mide a nivel nominal.
4.	F	La pregunta 3 utiliza una variable que se mide a nivel de razón.
5.	V	La pregunta 4 utiliza una variable que se mide a nivel de razón.
6.	V	La pregunta 5 utiliza una variable que se mide a nivel ordinal.
7.	V	La inferencia es utilizar un estadístico para utilizarlo como parámetro.
8.	V	Una muestra es equivalente a la población.

Autoevaluación 1		
Pregunta	**Respuesta**	**Retroalimentación**
9.	V	Una variable continua puede aceptar valores decimales.
10.	F	Si se pregunta por el número de hijos de una familia nos referimos a una variable continua.

Ir a la autoevaluación

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Autoevaluación 2		
Pregunta	Respusta	Retroalimentación
1.	V	El rango se calcula restando el mínimo del máximo valor.
2.	F	El intervalo de clase constituye el rango de la clase.
3.	F	No hay una regla para indicar cuantos valores intervienen.
4.	V	Equivale al número de veces que se repite un suceso.
5.	F	La fórmula que se emplea es 2^K.
6.	F	La fórmula que se emplea es (rango / k).
7.	V	El histograma grafica datos numéricos.
8.	F	De preferencia representa datos cualitativos.
9.	V	Relativo se refiere a porcentual.
10.	V	Acumula los porcentajes.

Ir a la autoevaluación

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Autoevaluación 3		
Pregunta	**Respuesta**	**Retroalimentación**
1.	F	La diferencia entre media muestral y media poblacional está en la fórmula que se aplica.
2.	F	La media poblacional es un estadístico.
3.	V	La mediana se utiliza en lugar de la media cuando la dispersión es demasiado alta.
4.	V	La moda equivale al valor que más repita dentro de un conjunto de datos.
5.	V	Para calcular la mediana se requiere que los valores estén ordenados.
6.	V	La mediana es la posición central de un conjunto de datos.
7.	F	La mediana y media siempre tienen el mismo valor.
8.	V	La medida más utilizada es la media aritmética.
9.	F	Se puede decir que la media y la moda son comparables.
10.	V	El promedio de edad de los ecuatorianos no es un estadístico, sino un parámetro.

Ir a la autoevaluación

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Autoevaluación 4		
Pregunta	**Respuesta**	**Retroalimentación**
1.	V	El rango es una medida de dispersión.
2.	V	La desviación estándar nos dice que tan dispersos están los datos.
3.	V	Datos dispersos significa que los valores están distribuidos en todo el rango.
4.	F	Si la mayoría de los datos están cerca de la media aritmética la dispersión es alta.
5.	V	La varianza es m,ás efectiva que el rango a la hora de calcular la dispersión.
6.	V	Dispersión equivale a la distribución de los valores dentro de un rango.
7.	V	La desviación estándar se calcula con base en las diferencias entre cada valor y la media aritmética.
8.	V	Dados los valores 4, 5, 6, la desviación estándar es 1.
9.	V	Dados los siguientes valores 75, 78, 23, 56 el rango es 55.
10.	F	El rango si eleva al cuadrado las desviaciones de los valores respecto de la media.

Ir a la autoevaluación

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Autoevaluación 5		
Pregunta	**Respuesta**	**Retroalimentación**
1.	V	Un diagrama de punto ofrece más información que una tabla de distribución de frecuencias.
2.	V	Un diagrama de tallo y hojas permite condensar los datos de las observaciones sin perder información.
3.	F	Los cuartiles dividen en 10 partes a un conjunto de observaciones.
4.	V	Cuando se calcula un decil, no se determina el valor sino la posición en la que está el valor que nos interesa.
5.	V	Un percentil cualquiera es en realidad una posición.
6.	V	Los diagramas de caja permiten determinar la dispersión de los datos.
7.	V	Un diagrama de caja se diagrama desde el primer al tercer cuartil.
8.	V	El cuartil 1 es equivalente al percentil 25.
9.	F	El decil 4 es equivalente al percentil 50.
10.	V	El percentil 100 es equivalente al cuarto cuartil.

Ir a la autoevaluación

Autoevaluación 6		
Pregunta	**Respuesta**	**Retroalimentación**
1.	V	El concepto de probabilidad hace referencia a la cuantificación de que un evento que pudiera presentarse o no.
2.	F	La probabilidad se puede calcular a través del cociente entre los resultados posibles para los resultados favorables a un evento.
3.	V	Se dice que dos o más eventos resultan ser mutuamente excluyentes cuando la presencia de uno impide que otro se presente al mismo tiempo.
4.	V	La probabilidad empírica también se conoce como probabilidad relativa ya que representa la fracción de eventos similares que sucedieron en el pasado.
5.	V	Un evento es el conjunto de uno o más resultados de un experimento.
6.	V	Existen tres enfoques para asignar probabilidades: objetivo, subjetivo y clásico.
7.	V	La probabilidad clásica parte del supuesto de que los resultados de un experimento son igualmente posibles.
8.	V	En las reglas de multiplicación se estima la probabilidad de que la ocurrencia de dos eventos sea simultánea.
9.	F	En la regla general de la adición los eventos deben ser mutuamente excluyentes.
10.	V	La regla del complemento se emplea para determinar la probabilidad de que un evento ocurra restando de 1 la probabilidad de un evento que no ha ocurrido.

Ir a la autoevaluación

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Autoevaluación 7		
Pregunta	**Respuesta**	**Retroalimentación**
1.	V	La distribución de probabilidad normal es simétrica respecto a la mediana.
2.	V	La distribución de probabilidad normal tiene forma de campana.
3.	V	Una variable con distribución normal puede adoptar una infinidad de valores dentro de un rango específico.
4.	V	La distribución normal se aplica a valores numéricos continuos.
5.	F	La variable marca de un vehículo puede ser representada con una distribución normal.
6.	V	La distribución normal estándar en realidad representa a una familia de distribuciones.
7.	F	El valor de z representa la media aritmética.
8.	F	La regla general dice que 68 de cada 100 observaciones se encuentran a dos desviaciones estándar antes y después de la media.
9.	V	El 100% de las observaciones se encuentran a 3 desviaciones estándar antes y después de la media.
10.	V	La tabla de distribución normal en las intersecciones de fila y columna contienen áreas en porcentaje

Ir a la autoevaluación

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Autoevaluación 8		
Pregunta	**Respuesta**	**Retroalimentación**
1.	V	La distribución de probabilidad lista todos los posibles resultados de un experimento.
2.	V	Si se tiene en una bolsa 4 bolitas (negra, blanca, roja y azul), al extraer solo una, la probabilidad de que sea blanca es ¼.
3.	V	Si se desea conocer cómo va un candidato a la alcaldía de la ciudad para plantear sus estrategias de campaña, es mejor encuestar a la muestra en lugar de a toda la población.
4.	V	Muestreo es un procedimiento que calcula un dato.
5.	F	Error de muestreo es la diferencia que existe entre la media de la población y una medida de tendencia central anterior de toda la población.
6.	F	Si se desea conocer la simpatía que tiene un candidato a presidente se puede estimar una medida de tendencia central.
7.	F	Si se desea conocer el salario promedio de los ciudadanos de una región se puede estimar una proporción.
8.	V	Un experimento puede tener solamente tres posibles resultados, listar estos resultados y la probabilidad que tenga cada uno se denomina tabla de distribución de probabilidad.
9.	V	Muestra representativa significa que debe representar a cada sector de la población.
10.	V	La diferencia entre estimar una media y estimar una proporción está en el fenómeno que se está investigando, si se desea conocer un valor absoluto o un porcentaje.

Ir a la autoevaluación

Autoevaluación 9		
Pregunta	**Respuesta**	**Retroalimentación**
1.	F	Una muestra no siempre debe ser representativa.
2.	V	En la fórmula para calcular el tamaño de una muestra para una media poblacional, z equivale al nivel de confianza.
3.	F	En la fórmula para calcular el tamaño de una muestra para una media poblacional, E se refiere a un valor que no puede pasar del 5%.
4.	F	Si se desea conocer a cuántas personas se debe encuestar para medir las preferencias de un candidato se debe aplicar la fórmula para calcular el tamaño de la muestra para la media.
5.	V	En la fórmula para calcular el tamaño de una muestra para una proporción el valor de p equivale a 0,5.
6.	V	El tamaño de una muestra depende del tamaño de la población.
7.	F	Dado el ejemplo de conocer las preferencias de los ciudadanos para la elección de alcalde de la ciudad, una muestra es representativa si se obtienen todas las encuestas en el sector central de la ciudad.
8.	V	Al calcular el tamaño de la muestra para una proporción, los valores de p y q deben valer 0,5 para maximizar el tamaño de la muestra.
9.	F	Si se desea conocer el tamaño de una muestra para calcular el salario promedio de los trabajadores de una ciudad se tiene que aplicar la fórmula para calcular el tamaño de una muestra para una proporción.
10.	F	En una ciudad pequeña se desea determinar el salario medio de los trabajadores. El error máximo que se va a permitir es de 140 dólares y el nivel de confianza de 95%. Se conoce que en estudios anteriores aplicados en la ciudad la desviación estándar fue 100. El tamaño adecuado para esa muestra es 700.

Ir a la autoevaluación

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Autoevaluación 10		
Pregunta	**Respuesta**	**Retroalimentación**
1.	V	A mayor tamaño de la muestra menor nivel de error muestral.
2.	V	La distribución de la media muestral tiende a mostrarse como una campana de forma normal.
3.	V	Dado el siguiente conjunto de datos 4, 6, 5, 7, 4, 6. Si se toman todas las muestras posibles de tamaño 3 la distribución de la media muestral adopta una forma normal perfecta.
4.	F	Dado el siguiente conjunto de datos 4, 6, 5, 7, 4, 6. Si se toman todas las muestras posibles de tamaño 3 la distribución de la media muestral adopta una forma normal mejor que la forma que adopta la distribución para todas las muestras posibles de tamaño 4.
5.	F	Dado el siguiente conjunto de datos 4, 6, 5, 7, 4, 6. Si se va a desarrollar la distribución de la media muestral y se busca que adopte la mejor forma normal posible, es mejor trabajar con muestras de tamaño 2 en lugar de muestras de tamaño 4.
6.	V	Si no conoce cuál es la forma de la distribución de una población, o sabe que esta no es normal, lo recomendable es tomar una muestra grande.
7.	F	Mientras más grande es la muestra, tiende menos a una distribución normal.
8.	V	La media de la distribución muestral de medias va a ser exactamente igual a la media de la población solo cuando se toman todas las muestras de tamaño n.
9.	V	Si no se toman todas las muestras de tamaño n, la media de la distribución muestral de medias va tener una gran diferencia con la media poblacional.
10.	V	El teorema central del límite señala que la distribución muestral de la media tiende a una forma normal conforme aumenta el tamaño de la muestra.

Ir a la autoevaluación

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Autoevaluación 11		
Pregunta	**Respuesta**	**Retroalimentación**
1.	F	Las hipótesis se refieren siempre a un estadístico de la muestra y nunca a un parámetro de la población.
3.	V	Verificar la hipótesis consiste en determinar si la hipótesis es una afirmación razonable.
3.	V	El nivel de significancia es el error que se está dispuesto a permitir aceptando.
4.	V	En el paso 3 para verificar una hipótesis se utiliza z cuando se desconoce la desviación estándar de la población.
5.	V	La regla decisión es una pregunta acerca de un rango de valores, aceptando H0 cuando el valor de Z está dentro de ese tango de valores.
6.	F	Un valor z puede ser calculado o puede ser buscado en la tabla. El valor crítico se lo calcula
7.	F	La prueba de una cola implica que el error se ubica en los dos lados de la curva.
8.	V	La prueba de dos colas requiere que el nivel de significancia se distribuya.
9.	F	Dado el siguiente enunciado "… determine si el volumen de ventas se ha incrementado". En la hipótesis nula se debe colocar el operador >=
10.	V	Dado el siguiente enunciado "… se desea conocer si el monto de ventas inicial es diferente del monto de ventas final". En la hipótesis nula se debe colocar el operador =

Ir a la autoevaluación

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Autoevaluación 12		
Pregunta	**Respuesta**	**Retroalimentación**
1.	V	La hipótesis se refiere a un parámetro poblacional.
2.	V	En la verificación de hipótesis se compara siempre un parámetro de la población con un estadístico de una muestra.
3.	V	El estadístico Z se utiliza cuando se conoce la desviación estándar de la población.
4.	F	Nivel de significación equivale al nivel de confianza que se tiene.
5.	V	Un error de tipo I equivale al nivel de significancia.
6.	F	El valor del estadístico de prueba no se calcula, se lo obtiene directamente de la tabla con la información de la muestra.
7.	F	Un error de tipo 2 equivale a rechazar la hipótesis nula cuando esta es falsa.
8.	V	La regla de decisión nos indica cuando aceptar o rechazar la H0.
9.	F	El valor crítico se calcula aplicando una fórmula.
10.	V	Si en el enunciado de un problema se pregunta si “es mayor que” se debe colocar > en H1.

Ir a la autoevaluación

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

5. Referencias Bibliográficas

Lind, D.; Marchal, W. y Wathen, S. (2015). *Estadística aplicada a los negocios y economía*. Loja, Ecuador: McGraw Hill-UTPL.

Lind, D.; Marchal, W. y Wathen, S. (2015). *Pruebas de hipótesis de una muestra.* [Presentación de PowerPoint]. Loja, Ecuador: McGraw Hill-UTPL.

6. Recursos

Documento 1.

Introducción a la estadística

1.1. Definiciones de la estadística

Antes de definir el término estadística es recomendable que usted lea las secciones “Introducción” y “¿Por qué se debe estudiar estadística?” del texto básico, en estas secciones el autor nos hace una introducción muy interesante que sirve como punto de partida. Seguidamente es necesario leer la sección “¿Qué se entiende por estadística?” en el texto básico, en donde se define el alcance de esta ciencia. Se recomienda que usted escriba en su cuaderno de notas su propia definición, recuerde que puede contar con mi apoyo. Si tiene preguntas, hágamelas saber y estaré contestándole lo más pronto posible, recuerde que estoy para apoyarle en su proceso de aprendizaje.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

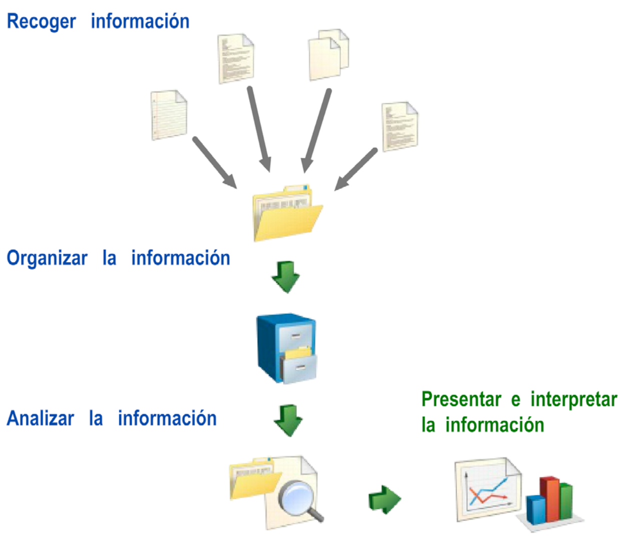

Figura 1. Definición de estadística. *Fuente:* XXXX.

En el texto, en la sección señalada existen varios ejemplos de estadística, sin embargo, estos no son de la realidad local. A continuación, le planteo dos ejemplos de la realidad ecuatoriana.

1. La tasa de desempleo en el Ecuador es del 5%.

Esto significa que de cada 100 ecuatorianos 5 no tienen empleo formal.

2. El crecimiento económico del país en el 2020 será del 4%.

Esto significa que el total de bienes y servicios que produce el país será 4% mayor que en el año 2019. Por ejemplo, si en el 2019 el PIB ecuatoriano fue de cien mil millones, en el año 2020 será de ciento cuatro mil millones.

Estos ejemplos son de tipo económico, pero hablemos de temas con los que estamos más relacionados.

1. El 10% de los padres de familia nunca preguntan en la escuela por el desempeño de sus hijos.

Esto significa que, de cada 100 padres de familia, 10 nunca preguntan por el desempeño de sus hijos en la escuela.

2. El 97% de ecuatorianos tiene teléfono móvil.

Esto significa que, de cada 100 ecuatorianos, 97 tienen teléfono celular y 3 no lo tienen.

El estudio de la estadística es importante por los siguientes aspectos.

- Diariamente nos encontramos con información numérica que debemos interpretar.
- La información es nuestro insumo para tomar decisiones que afectan nuestra vida diaria.
- Independientemente del tipo de trabajo que realicemos, la estadística es útil para analizar datos y tomar ciertas decisiones importantes.

De forma general podemos resumir que el tener conocimientos de estadística nos puede ayudar en dos niveles: 1) en el campo profesional para desempeñar mejor nuestro trabajo; y, 2) en el diario vivir para comprender las distintas situaciones que ocurren en la sociedad, ya sea a nivel político, económico, cultural, etcétera.

1.2. Tipos de estadística

La estadística abarca dos partes: estadística descriptiva y estadística inferencial.

La descriptiva se encarga de recoger, organizar y presentar datos de una manera entendible.

La estadística inferencial se encarga de determinar alguna característica de una población con base en los datos de una muestra de esa población; también se puede decir que se pueden obtener conclusiones generales para una población basándose en los datos de una muestra.

Un ejemplo de estadística inferencial es el siguiente: se pregunta a un segmento de la población (a una muestra) cuánto gasta mensualmente en telefonía celular. Si la muestra es representativa, el promedio de los valores obtenidos en la encuesta es una aproximación bastante real del promedio de toda la población.

Aquí es necesario que consideremos dos conceptos adicionales: población y muestra.

Hay una imagen en la página 8 del texto básico que, sin necesidad de palabras, nos va a aclarar qué es una población y qué es una muestra; sin embargo, es conveniente que tenga claro un concepto más formal:

Una **población** se refiere a todos los elementos o personas sobre las que estamos buscando información. Por su parte, la **muestra** es solamente un conjunto de esos elementos.

La inferencia consiste en extender los cálculos de la muestra a toda la población. Por ejemplo, se calcula el gasto mensual en telefonía celular de la muestra y se asume que ese promedio corresponde a toda la población.

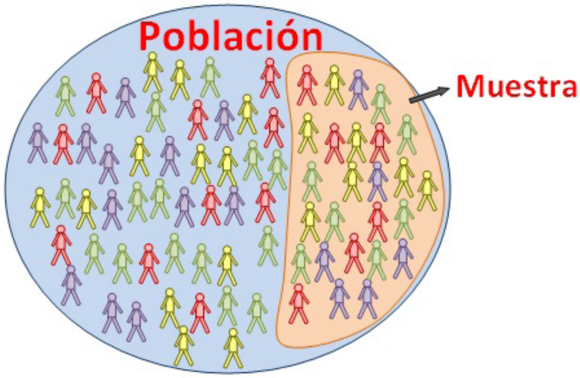

Figura 2. Población y muestra. *Fuente:* Universo fórmulas. (ND). Muestra estadística

Aquí es importante señalar que la muestra va a contar siempre con estadísticos, estos pueden ser la media o la desviación estándar; en tanto que la media o la desviación estándar de una población toman el nombre de parámetros. Es conveniente tener claro la diferencia entre estadístico y parámetro.

Hay razones que nos conducen a trabajar solamente con una muestra y no con toda la población. Algunas de estas razones se exponen en la página 7 del texto básico. En general, cuando la muestra es representativa de la población, los resultados alcanzados en la muestra son estadísticamente equivalentes con los de la población.

1.3. Tipos de variables

Este tema es muy sencillo, le sugiero revisar la sección “Tipos de variables” del texto básico. Como tarea anote en su cuaderno los tipos de variables y escriba dos ejemplos de cada tipo.

Al hablar de variables conviene señalar que estas guardan los datos que se recogen. Los datos pueden obtenerse de fuentes primarias y de fuentes secundarias de información. Las fuentes primarias pueden ser: encuestas, entrevistas, los datos que se obtienen en un laboratorio, etc. Las fuentes secundarias se refieren a periódicos, revistas científicas, informes de investigación, etcétera.

1.4. Niveles de medición de variables

Cuando hablamos de niveles de medición nos referimos al tipo de cálculos que se pueden hacer con los datos que se recogen. Ejemplo: el valor que cada persona gasta mensualmente en telefonía celular es un tipo de dato numérico, pero se mide a un nivel de razón. Nivel de razón significa que se puede hacer cálculos aritméticos con el dato.

Al revisar la sección correspondiente a “niveles de medición” del texto básico determine los niveles de medición. Luego lea este resumen que planteo a continuación.

1. Nivel nominal: sirve para dar nombre a algo (nominal); por ejemplo: ¿cuál es la marca del carro que conduce? La persona que contesta puede escoger cualquier color, este dato es de tipo nominal.

2. Nivel ordinal: es igual al nivel nominal, pero se le adiciona un orden. Ejemplo: a los pacientes que llegan con dolor a un hospital se les pregunta si tienen mucho dolor, dolor moderado, poco dolor o ningún dolor. No existe una cuantificación del dolor de cada categoría, pero sí se puede saber que “mucho dolor” es mayor que “poco dolor”. Por esa razón estos datos toman el nombre de ordinales.

3. Nivel de intervalo: igual al nivel nominal, pero se puede cuantificar la diferencia entre una categoría y otra. Ejemplo: las tallas de ropa; se sabe que la talla 38 es 2,5 pulgadas más grande de la talla 36.

4. Nivel de razón: se refiere básicamente a datos numéricos con los que se puede hacer operaciones aritméticas.

Ir al contenido

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Documento 2.

Distribuciones de frecuencia y representaciones gráficas

> En esta sección ponemos especial énfasis en el indicador de aprendizaje “Construye una distribución de frecuencias para presentar un conjunto de datos” planificada para el primer bimestre. Quiero invitarle a que se organice y vaya tomando nota de los diferentes elementos que se tratan en esta unidad; en caso de tener dudas, estoy presto para apoyarle, recuerde que el tiempo que invierta en este cometido, redundará en su beneficio personal.

Una de las tareas más comunes es la organización de conjuntos de datos y presentarlos de una manera entendible; para explicar este tema partamos de un ejemplo sencillo.

Se encuesta a los estudiantes de la modalidad a distancia de la UTPL (a 20 estudiantes) y se les hace las siguientes cuatro preguntas.

1. Sexo:___
2. Edad:___
3. Gasto mensual en teléfono:____
4. Marca del teléfono:____
5. Los resultados de aplicar la encuesta son los siguientes:

Tabla 1. Datos obtenidos de la encuesta a 20 personas

Número	Sexo	Edad	Gasto mensual	Marca
1	H	21	12	Nokia
2	M	27	10	Nokia
3	H	32	12	Samsung

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Número	Sexo	Edad	Gasto mensual	Marca
4	H	45	11	Nokia
5	M	22	10	HTC
6	H	25	5	Nokia
7	M	31	5	iPhone
8	M	22	10	Nokia
9	H	37	12	iPhone
10	M	54	15	Sony
11	H	45	14	Samsung
12	M	33	8	Sony
13	M	31	10	Nokia
14	H	30	5	Alcatel
15	M	27	14	Sony
16	H	25	8	Samsung
17	M	19	10	Sony
18	H	40	10	iPhone
19	H	17	5	Alcatel
20	M	39	12	HTC

Fuente: XXXXXX

Analicemos inicialmente la variable gasto que representa el valor gastado mensualmente en telefonía.

Paso 1: ordenamos los valores

5, 5, 5, 5, 8, 8, 10, 10, 10, 10, 10, 10, 11, 12, 12, 12, 12, 14, 14, 15

Paso 2: establecemos los valores mínimo y máximo.

Mínimo: 5

Máximo: 15

Rango = Máximo – Mínimo

Rango = 10

De aquí se puede determinar que la variable gasto va a tomar los valores desde 5 hasta 15; es decir el rango en el que puede variar va desde 5 hasta 15.

Paso 3: se cuenta la frecuencia de cada valor de la variable gasto.

Tabla 2. Tabla de frecuencias del gasto mensual en telefonía celular

Gasto mensual	Frecuencia
5	4
8	2
10	6
11	1
12	4
14	2
15	1
Total	**20**

Fuente: XXXXXX

Ampliando la explicación de este paso, se toma cada uno de los valores y se cuenta cuántas personas tienen el mismo valor. En el primer caso, hay 5 personas que gastan 5 dólares mensuales en telefonía; en el segundo caso, hay 2 personas que gastan 8 dólares mensuales y así sucesivamente.

Observe nuevamente los resultados de la encuesta y organicemos ahora los datos de la variable marca, esta variable se refiere a la marca de teléfono que tienen los estudiantes encuestados. Esta variable es de tipo cualitativo.

Tabla 3. Tabla de frecuencias de la marca de los teléfonos

Marca	Frecuencia
Alcatel	2
HTC	2

Marca	Frecuencia
iPhone	3
Nokia	6
Samsung	3
Sony	4
Total	**20**

Fuente: XXXXXX

En este caso al tratarse de una variable cualitativa no fue necesario ordenar los valores debido a que cada valor es una categoría. Se procedió directamente a contar cuantos teléfonos de cada marca hubieron y el resultado se lo coloca en la tabla.

Lo que acabamos de hacer es válido para conjuntos pequeños de datos, hay casos en los que la cantidad de información es más grande y es necesario optar por los procedimientos que estudiaremos a continuación.

2.1. Distribución de frecuencias

Esta es una de las áreas más interesantes y útiles para comprender la información que se nos presenta en la vida cotidiana. Para abordar el tema usted debe revisar en el texto básico el tema "Distribución de frecuencias" y luego continuar con la lectura de esta guía didáctica.

¿Terminó la lectura?, ¡felicitaciones! Ahora es necesario que tome nota de los siguientes conceptos.

- Rango
- Intervalo de clase
- Límite de clase
- Número de clases
- Tamaño de clase
- Frecuencia
- Marca de clase

En el texto básico se presentan ejemplos detallados, en esta guía vamos a complementar esto con los siguientes ejercicios.

Se consulta la edad a 40 estudiantes de la UTPL. Los resultados son los siguientes:

Tabla 4. Edades de la muestra de estudiantes UTPL

24	27	32	36	22	25	31	22	37	34
34	33	31	30	27	25	29	40	26	39
29	24	27	31	36	24	25	31	22	36
40	41	33	31	30	26	25	23	24	27

Fuente: XXXXXX

1. Máximo, mínimo y rango

Se procede a ordenar los valores y luego se obtiene el máximo y mínimo.

El valor máximo es: 22.

El valor mínimo es: 45.

El rango o recorrido de la variable es = Valor máximo – Valor mínimo.

El rango o recorrido de la variable es = 45 – 22.

El rango o recorrido de la variable es = 23.

2. Número de clases

El número de clases se calcula siguiendo las explicaciones del texto básico.

$2^k \geq n$

$2^k \geq 40$

Probamos valores para k hasta que el resultado de la potencia sea mayor o igual a 40:

2^K K=1 2

2^K K=2 4

2^K K=3 8

2^K K=4 16

2^K K=5 32

2^K K=6 64

64 es mayor que 40, por lo que el valor de k apropiado es 6.

3. Tamaño de clase

La fórmula utilizada es:

$$\text{Tamaño de clase} = \frac{\text{Rango}}{K}$$

$$\text{Tamaño de clase} = \frac{23}{6}$$

Tamaño de clase = 3,83 se redondea el valor al superior 4.

4. Elaboración de la tabla de frecuencias

Para elaborar la tabla, se establecen las categorías y se cuentan los valores que corresponden a cada clase. En este caso, vamos a contar las edades que pertenecen a cada clase.

Tabla 5. Tabla de distribución de frecuencias de las edades

Clase	Frecuencia
22 a 25	10
26 a 29	6
30 a 33	10
34 a 37	7
38 a 41	3
42 a 45	4
Total	**40**

Fuente: XXXXXX

Importante: cada clase tiene un ancho de clase igual a 4; las clases son mutuamente excluyentes, esto quiere decir que si un valor pertenece a una clase no puede pertenecer a otra; la frecuencia es el número de edades que están dentro del ancho de la clase (rango).

Hagamos otro ejemplo en el que la cantidad de valores sea mayor. Suponga que hemos pasado una encuesta a los estudiantes de la UTPL en donde se le formulan varias preguntas, una de las preguntas consulta sobre el nivel de ingresos de la familia de estudiante; es decir, cada persona que contesta la encuesta debe, entre otras cosas, responder el valor de los ingresos mensuales (sueldo) de su familia.

Los valores recolectados son los siguientes.

Tabla 6. Tabla de datos con los resultados de la encuesta

700	940	900	1000	250
640	850	610	900	560

467	720	780	650	670
235	390	680	710	458
1.560	1.450	540	1.480	582
278	1.600	620	1.510	593
450	1.200	540	230	620
357	410	610	250	494
456	900	620	480	729
230	1200	750	650	593
560	390	800	640	429
750	670	450	710	321
1.200	720	375	840	498
1.400	840	280	756	578
360	600	590	895	890
570	560	560	1.000	1.250
1.100	760	840	467	1.150
347	450	810	624	1.080
180	580	814	1.050	1.400
150	990	640	300	1.450

Fuente: XXXXXX

1. Máximo, mínimo y rango

Se procede a ordenar los valores y luego se obtiene el máximo y mínimo

El valor máximo es: 150.

El valor mínimo es: 1.600.

El rango o recorrido de la variable es = valor máximo – valor mínimo

El rango o recorrido de la variable es = 1.600 – 150.

El rango o recorrido de la variable es = 1.450.

2. Número de clases

El número de clases se calcula siguiendo las explicaciones del texto básico.

2k ≥ n

2k ≥ 100 (son 100 los valores recogidos con la encuesta)

Probamos valores para k hasta que el resultado de la potencia sea mayor o igual a 100:

2^K	K=1	2
2^K	K=2	4
2^K	K=3	8
2^K	K=4	16
2^K	K=5	32
2^K	K=6	64
2^K	K=7	128

128 es mayor que 100, por lo que el valor de k apropiado es 7.

3. Tamaño de clase

La fórmula utilizada es:

$$\text{Tamaño de clase} = \frac{\text{Rango}}{k}$$

$$\text{Tamaño de clase} = \frac{1450}{7}$$

Tamaño de clase = 207,14 Se redondea el valor al superior 208.

4. Elaboración de la tabla de frecuencias

Para elaborar la tabla, se establecen las categorías y se cuentan los valores que corresponden a cada clase. En este caso, vamos a contar los sueldos que pertenecen a cada clase.

Note que cada clase se construye con un ancho de clase de 208.

Tabla 7. Tabla de distribución de frecuencias

Clase	Frecuencia
150 – 358	13
359 – 567	22
568 – 776	32
777 – 985	14
986 – 1194	7
1195 – 1403	6
1404 - 1602	6
Total	**100**

Fuente: XXXXXX

En el texto básico existen muchos ejemplos del mismo tipo, le recomiendo seguir pasos que se han señalado aquí.

Resumen

Vamos a hacer un resumen: para elaborar una distribución de frecuencias los pasos recomendados son los siguientes.

1. Ordenar los valores y encontrar el máximo, mínimo y rango.
2. Determinar el número de clases con la fórmula $2k \geq n$.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

3. Calcular el tamaño de clase, la fórmula utilizada es:

$$\text{Tamaño de clase} = \frac{\text{Rango}}{k}$$

4. Elaboración de la tabla de frecuencias.

2.2. Distribución de frecuencias relativas

Antes de empezar esta sección es necesario entender el término “relativo”, según la real academia significa: “que guarda relación con alguien o con algo.”. En efecto, cuando hablamos de frecuencias relativas estamos hablando de ellas de una forma que permite que se comparen con las demás, para esto debemos expresarlas en porcentaje; de esa manera, el porcentaje que tenga cada clase nos dirá su tamaño respecto a las demás clases.

Vamos a obtener las frecuencias relativas de los ejemplos planteados anteriormente.

En el primer caso nos referimos a la tabla de frecuencias con la edad de los 40 estudiantes de la UTPL.

Tabla 8. Tabla de distribución de frecuencias de las edades de los encuestados

Clase	Frecuencia
22 a 25	10
26 a 29	6
30 a 33	10
34 a 37	7
38 a 41	3
42 a 45	4
Total	**40**

Fuente: XXXXXX

Vemos que la tabla tiene siete clases, para obtener la frecuencia relativa de cada clase hacemos lo siguiente.

Tabla 9. Tabla de distribución de frecuencias que incluye frecuencia relativa

Clase	Frecuencia	Frecuencia relativa
22 a 25	10	10/40
26 a 29	6	6/40
30 a 33	10	10/40
34 a 37	7	7/40
38 a 41	3	3/40
42 a 45	4	4/40
Total	**40**	**1**

Fuente: XXXXXX

Cada uno de los valores de la frecuencia (frecuencia absoluta simple) debe ser dividido para el total de observaciones (valores recogidos). El resultado en cada caso es un valor decimal menor que uno.

Tabla 10. Tabla de distribución de frecuencias con explicación de frecuencia relativa

Clase	Frecuencia	Frecuencia relativa	Significado
22 a 25	10	0,25	Equivale al 25%
26 a 29	6	0,15	Equivale al 15%
30 a 33	10	0,25	Equivale al 25%
34 a 37	7	0,175	Equivale al 17,5%
38 a 41	3	0,075	Equivale al 7,5%
42 a 45	4	0,1	Equivale al 10%
Total	**40**	**1**	**Equivale al 100%**

Fuente: XXXXXX

En la tabla anterior se ha incluido una columna final que explica el significado de los valores de la frecuencia relativa; esto equivale a multiplicar por 100 el valor de la frecuencia relativa.

En el segundo ejemplo vamos a proceder de la misma manera:

Tabla 11. Tabla de distribución de frecuencias

Clase	Frecuencia
150 – 358	13
359 – 567	22
568 – 776	32
777 – 985	14
986 – 1194	7
1.195 – 1.403	6
1.404 – 1.602	6
Total	**100**

Fuente: XXXXXX

Agregamos una columna en la que vamos a calcular la frecuencia relativa para cada frecuencia absoluta simple:

Tabla 12. Tabla de distribución de frecuencias y frecuencia relativa

Clase	Frecuencia	Frecuencia relativa
150 – 358	13	13 / 1.000
359 – 567	22	22 / 1.000
568 – 776	32	32 / 1.000
777 – 985	14	14 / 100
986 – 1194	7	7 / 100
1195 – 1403	6	6 / 1.000
1404 - 1602	6	6 / 100
Total	**100**	**1**

Fuente: XXXXXX

El resultado es el siguiente.

Tabla 13. Tabla de distribución de frecuencias con frecuencia relativa

Clase	Frecuencia	Frecuencia relativa
150 – 358	13	0,13
359 – 567	22	0,22
568 – 776	32	0,32
777 – 985	14	0,14
986 – 1194	7	0,07
1195 – 1403	6	0,06
1404 - 1602	6	0,06
Total	**100**	**1**

Fuente: XXXXXX

Para expresar el resultado en porcentaje cada valor debe multiplicarse por 100.

Veamos ahora un ejemplo en el que colocamos en una tabla de frecuencias el tipo de dolor que tienen los pacientes que llegan a un hospital. Se trata de un muestreo sobre 60 pacientes que llegan y son atendidos.

Luego de contar cuántos pacientes tienen dolor severo, se lo escribe en la tabla:

Tabla 14. Tabla de frecuencias

Categoría de dolor	Frecuencia (número de pacientes)	Frecuencia relativa	Frecuencia acumulativa	Frecuencia relativa acumulada
Severo	4			
Moderado	8			
Leve	17			
Ninguno	31			

Fuente: XXXXXX

Para calcular la frecuencia relativa se procede a dividir el número de pacientes en cada caso (la frecuencia) para el total de pacientes (se divide para 60).

Tabla 15. Tabla de frecuencias con frecuencia relativa

Categoría de dolor	Frecuencia (número de pacientes)	Frecuencia relativa	Frecuencia acumulativa	Frecuencia relativa acumulada
Severo	4	= 4 / 60 **= 0,07**		
Moderado	8	= 8 / 60 **= 0,13**		
Leve	17	= 17 / 60 **= 0,28**		
Ninguno	31	= 31 / 60 **= 0,52**		

Fuente: XXXXXX

Para calcular la frecuencia acumulativa se empieza por la última fila de la tabla y se va sumando las frecuencias como se muestra en la tabla siguiente.

Tabla 16. Tabla de frecuencias con frecuencia acumulada

Categoría de dolor	Frecuencia (número de pacientes)	Frecuencia relativa	Frecuencia acumulativa	Frecuencia relativa acumulada
Severo	4	= 4 / 60 **= 0,07**	= 56 + 4 **= 60**	
Moderado	8	= 8 / 60 **= 0,13**	= 48 + 8 **= 56**	
Leve	17	= 17 / 60 **= 0,28**	= 31 + 17 **= 48**	
Ninguno	31	= 31 / 60 **= 0,52**	= 31	

Fuente: XXXXXX

Para obtener la frecuencia relativa acumulada se inicia por la última fila de la tabla y se va sumando las frecuencias relativas como se muestra a continuación.

Tabla 17. Tabla de frecuencias con frecuencia relativa acumulada

Categoría de dolor	Frecuencia (número de pacientes)	Frecuencia relativa	Frecuencia acumulativa	Frecuencia relativa acumulada
Severo	4	= 4 / 60 **= 0,07**	= 56 + 4 **= 60**	= 0,93 + 0,07 **= 1**
Moderado	8	= 8 / 60 **= 0,13**	= 48 + 8 **= 56**	= 0,80 + 0,13 **= 0,93**
Leve	17	= 17 / 60 **= 0,28**	= 31 + 17 **= 48**	= 0,52 + 0, 28 **= 0,80**
Ninguno	31	= 31 / 60 **= 0,52**	**= 31**	**= 0,52**

Fuente: XXXXXX

2.3. Representación gráfica de una distribución de frecuencias

Una vez usted tenga lista su tabla de distribución de frecuencias, puede dar un paso más para presentar los resultados de una mejor manera. En realidad este es un trabajo muy sencillo y hay que diferenciar dos tipos de gráficos.

Diagrama de barras: utilizado para representar datos cualitativos

Histograma: utilizado para representar datos cuantitativos.

En uno de los ejemplos que hemos realizado en esta guía preguntamos por la marca del teléfono de las personas, la tabla de frecuencias y el respectivo gráfico son los siguientes.

Tabla 18. Tabla de frecuencias de la marca de teléfono

Marca	Frecuencia
Alcatel	2
HTC	2
iPhone	3
Nokia	6
Samsung	3
Sony	4

Fuente: XXXXXX

Esta tabla se representa en la siguiente figura.

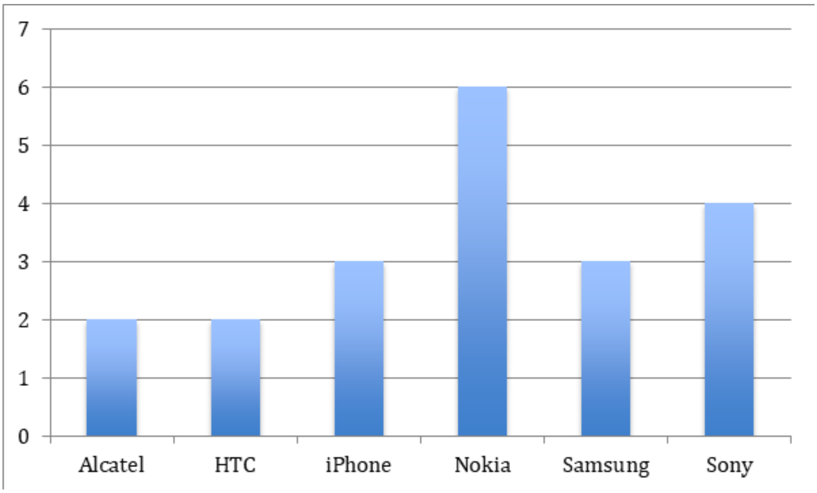

Figura 3. Diagrama de frecuencias. *Fuente:* XXXXXX

Note los siguientes aspectos:

a. Las barras están separadas debido a que cada marca de teléfono es una categoría, esto pasa con los datos cualitativos.
b. La altura de cada categoría está dada por la frecuencia, esta se mide con los valores del eje de las y.

c. El eje de las x tiene las categorías y el eje de las y los valores de las frecuencias.

d. Mirar un gráfico es más cómodo que mirar una tabla, en todo caso, la presentación de los resultados debe hacerse pensando en las personas que van a leer los resultados del trabajo.

Un histograma puede representar datos cuantitativos, aquí es conveniente diferenciar entre valores discretos y valores continuos.

En un histograma las barras están unidas entre sí, esto debido a que los valores de cada categoría son consecutivos.

En el eje de las x de un histograma se colocan los valores de las categorías o clases y en el eje de las y las frecuencias

Para complementar este tema le recomiendo la siguiente actividad: revise el ejercicio de la sección "Representación gráfica de una distribución de frecuencias" en el texto básico. En este ejercicio se puede apreciar claramente cómo se representa a través de un histograma los valores de una tabla de frecuencias.

Ir al contenido

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Documento 3.

Medidas de tendencia central y medidas de dispersión

En esta sección ponemos especial énfasis en el indicador de aprendizaje “Describe el centro y la dispersión de un conjunto de datos” planificada para el primer bimestre. Si podemos calcular las medidas centrales y las de dispersión de un conjunto de datos, entonces podemos describir de forma acertada ese conjunto de datos. No dude en escribirme las consultas que tenga, estoy presto para apoyarlo en su proceso de aprendizaje.

Hasta ahora hemos levantado un conjunto de datos, los hemos organizado, representado en una tabla y a través de un gráfico. Ahora vamos a agregar información que permite describir aún más el conjunto de datos.

Una forma básica y siempre necesaria para describir los datos es la información que indique el centro de esos datos, el centro puede indicarse a través de la media, la mediana o la moda.

3.1. Media

La media aritmética de un conjunto de datos nos indica el promedio. Este valor permite hacerse una idea clara de la naturaleza de los datos. Ejemplo: Al preguntar la edad de un grupo de personas, el promedio nos va a indicar si estamos hablando de: niños, adolescentes, adultos, adultos mayores. No necesariamente todas las personas van a pertenecer a la categoría del ejemplo, pero contamos con al menos una aproximación. Otros ejemplos que permiten describir un conjunto de datos con la media aritmética son los siguientes.

- El sueldo promedio en Ecuador es de 411 dólares.
- El promedio de edad de los estudiantes de la UTPL es de 27 años.

- El costo promedio por semestre de una carrera en la UTPL es de 1.000 dólares.

> Ahora le recomiendo leer en el texto básico la sección correspondiente a media poblacional y media de la muestra. Seguidamente lea el complemento que se presenta a continuación.

Tanto la media poblacional como la media de la muestra se calculan de la misma forma, la diferencia es conceptual; es decir, la diferencia se refiere al campo en el que se aplica cada una. La media poblacional constituye un parámetro poblacional y se calcula utilizando todos los elementos de la población; por su parte, la media muestral constituye un estadístico y se calcula utilizando solamente los elementos de la muestra.

Las propiedades de la media aritmética son 4 y se describen en el texto básico en el apartado "propiedades de la media aritmética". En esta guía se explora la primera propiedad que dice: "Todo conjunto de datos de intervalo o de nivel de razón tienen una media". Esto quiere decir que todos los valores numéricos tienen una media. Se puede obtener la media de datos a nivel de intervalo. Ejemplo: la talla de chaqueta de los estudiantes de un paralelo de la UTPL. Recuerde que la talla de ropa es una variable que se mide a nivel de intervalo. Otro ejemplo a nivel de razón es la edad. Se puede obtener la edad promedio de los habitantes de una ciudad. La edad es una variable que se mide a nivel de razón.

Actividad recomendada

Resuelva los ejercicios del 1 al 12 del texto básico, estos se encuentran en la sección correspondiente a la media aritmética.

Tanto la media ponderada como la media geométrica debe estudiarlas del texto básico, le recomiendo dar lectura a cada tema y luego observar cómo se resuelven los ejercicios. Algo que no puede dejar de hacer es tomar nota de cuando se utiliza cada una; es decir, usted debe tener claro cuando utilizar una media ponderada y cuando una media geométrica.

3.2. Mediana

Hay casos en los que la media no representa fielmente al conjunto de datos de la que se extrae. Existen conjuntos de datos que tienen una dispersión tan alta que hace que la media aritmética no sea la medida más apropiada. Suponga el siguiente ejemplo. Se le consulta la edad a un grupo de personas en una universidad. Los datos recolectados son los siguientes.

17, 18, 19, 21, 18, 23, 22, 77, 60, 70.

La edad promedio de este grupo de personas es de 34,5 años. Sin embargo, este promedio no refleja la realidad. Es posible que un error de muestreo haya incluido a 3 personas con edades bastante elevadas lo que hace que el promedio sea más alto.

En estos casos es mejor emplear la mediana en lugar de la media aritmética. La mediana encuentra el valor que está en la posición central de los datos previamente ordenados. Para presentar de mejor manera esta explicación le propongo revisar la siguiente imagen.

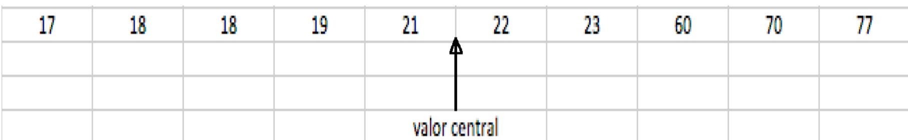

Figura 4. Mediana de un conjunto de valores. *Fuente:* XXXXXX

Si ordenamos los valores y encontramos el valor central tenemos 21,5, este valor representa mucho mejor las edades de las personas encuestadas.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Para complementar lo invito a leer la sección correspondiente en el texto básico y a tomar nota de las dos propiedades de la mediana. Estas propiedades son muy importantes por lo que le sugiero las anote en su cuaderno de notas.

3.3. Moda

La moda representa al valor que más se repite en un conjunto de datos. Cuando se trata de valores cuantitativos es conveniente tener los datos ordenados para poder identificar mejor el valor que más se repite.

Este es un tema bastante sencillo que se complementa con la lectura de la sección correspondiente en el texto básico.

Actividad recomendada

Resuelva los ejercicios del 17 al 24 del texto básico, estos se encuentran en la sección correspondiente a la moda.

3.4. Medidas de dispersión.

Es muy importante leer la sección "¿Por qué estudiar la dispersión?". Para entender qué es y por qué es importante.

Las medidas de dispersión que más se utilizan son:

- Rango.
- Desviación media.
- Varianza.
- Desviación estándar.
- Cuartiles, deciles y percentiles.

Una dispersión alta es un indicador de que los datos están distribuidos de manera uniforme a lo largo del rango, esto hace poco confiable a la media aritmética; por el contrario, una dispersión baja

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

es un indicador de que los datos están distribuidos alrededor de la media y, por lo tanto, la media aritmética es confiable.

Ahora le recomiendo dar lectura a toda la sección "Medidas de dispersión". Tome nota de los conceptos que vaya encontrando y de las fórmulas que se utilizan para calcular cada una de las medidas.

¿Recuerda el ejercicio que resolvimos para encontrar la mediana?, vamos a calcular el rango de estos datos.

17	18	18	19	21	22	23	60	70	77

El rango se obtiene restando del valor máximo el valor mínimo.

Rango = 77 – 17 = 60

Ahora vamos a determinar la desviación media, esto se describe en la siguiente tabla.

Tabla 1. Valor absoluto de la desviación media

Xi	Valor absoluto (Xi - media)	Valor absoluto (Xi - media)
17	17 - 34,5	17,5
18	18 - 34,5	16,5
18	18 - 34,5	16,5
19	19 - 34,5	15,5
21	21 - 34,5	13,5
22	22 - 34,5	12,5
23	23 - 34,5	11,5
60	60 - 34,5	25,5
70	70 - 34,5	35,5
77	77 - 34,5	42,5
Total		207

Fuente: XXXXXX

Desviación media = 207/10

Desviación media = 20,7

La varianza de estos datos se obtiene siguiendo la fórmula que consta en el texto básico.

Tabla 2. Desviación media al cuadrado

Xi	(Xi - media)	(Xi - media)	(Xi - media)2
17	17 - 34,5	-17,5	306,25
18	18 - 34,5	-16,5	272,25
18	18 - 34,5	-16,5	272,25
19	19 - 34,5	-15,5	240,25
21	21 - 34,5	-13,5	182,25
22	22 - 34,5	-12,5	156,25
23	23 - 34,5	-11,5	132,25
60	60 - 34,5	25,5	650,25
70	70 - 34,5	35,5	1.260,25
77	77 - 34,5	42,5	1.806,25
Total			**5.278,5**
Varianza = 5278,5 / 10			**527,85**

Finalmente, la desviación estándar del conjunto de datos equivale a obtener la raíz cuadrada de la varianza.

Desviación estándar = Raíz cuadrada de 527,85

Desviación estándar = 22,97

Actividad recomendada

Resuelva los ejercicios del 41 al 46 del texto básico, estos se encuentran al final de la sección que trata las medidas de dispersión.

El teorema de Chebyshev es muy importante y nos indica cómo se distribuyen los datos de naturaleza numérica. Antes de terminar esta unidad le recomiendo leer dicho teorema ubicado en la sección "Interpretación y usos de la desviación estándar". Después de leer, observe el gráfico correspondiente a la regla empírica, la cual sugiere que:

Si se toma la medida de los valores como punto de referencia, desde una desviación estándar antes de la media, hasta una desviación estándar después de la media se encuentran el 68% de los valores recogidos; desde dos desviaciones estándar antes de la media, hasta dos desviaciones estándar después de la media se encuentran el 95% de los valores; y, finalmente, desde tres desviaciones estándar antes de la media, hasta tres desviaciones estándar después de la media se encuentran el 99,7% de los valores.

La regla anterior supone que, por ejemplo, si tenemos una media de 100 y una desviación estándar de 10, desde 90 (100 menos una desviación) hasta 110 (100 más una desviación) se encuentran el 68% de los valores observados; desde 80 (100 menos dos desviaciones) hasta 120 (100 más dos desviaciones) se encuentran el 95% de las observaciones; y finalmente, desde 70 (100 menos tres desviaciones) hasta 130 (100 más tres desviaciones) se encuentra el 99,7% de las observaciones.

Ir al contenido

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Documento 4.

Descripción de datos

Siempre se ha dicho que una imagen puede expresar mucho más que mil palabras y esto no está alejado de la verdad, cuando se trata de datos, estos se pueden describir utilizando distintas gráficas. En esta sección ponemos especial énfasis en el indicador de aprendizaje "Describir datos utilizando gráficas" planificada para el primer bimestre.

Apreciado(a) estudiante, esta unidad no reviste mayor complejidad, le animo a que la desarrolle y aprenda a representar gráficamente conjuntos de datos. En caso de tener dudas, escríbame o llame, estaré presto para ayudarle en su proceso de aprendizaje.

Para describir datos se puede utilizar diagramas de puntos, gráficas de tallo y hoja, también se puede utilizar medidas de dispersión complementarias como: cuartiles, deciles o percentiles. Estas medidas nos permiten tener una idea preliminar de la composición de un conjunto de datos. En caso de que se quiera relacionar dos variables podemos utilizar gráficas que nos presentan una realidad de forma entendible.

4.1. Diagramas de puntos

Los diagramas de puntos permiten graficar un conjunto pequeño de datos, en este tipo de diagramas cada observación se representa con un punto que se dibuja sobre una recta (eje X).

> Se recomienda dar lectura a la sección "Diagramas de puntos" del texto básico y observar el diagrama del ejemplo que se desarrolla.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Para complementar vamos a resolver el siguiente ejemplo: Dado el siguiente conjunto de datos, que corresponde al valor mensual que gastan en telefonía un grupo de 20 personas.

5, 5, 5, 5, 8, 8, 10, 10, 10, 10, 10, 10, 11, 12, 12, 12, 12, 14, 14, 15,

Lo primero es construir la recta y etiquetarla con los valores y asignar un punto a cada uno de los valores.

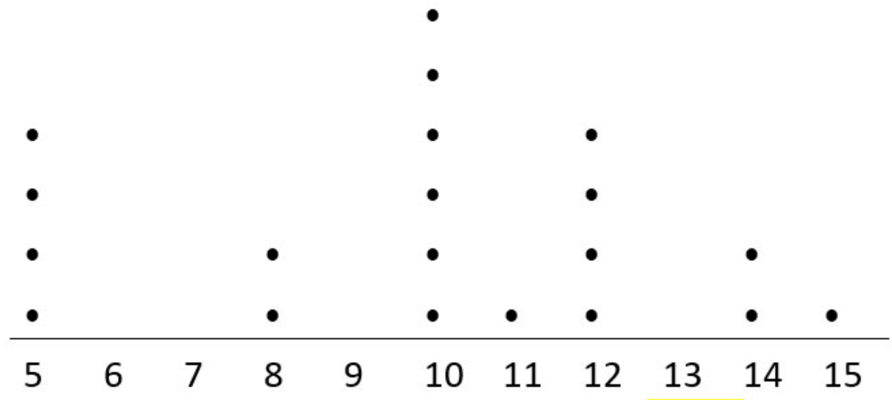

Figura 5. Diagrama de puntos. *Fuente*: XXXX.

4.2. Gráficas de tallo y hojas

Un diagrama de tallo y hojas es muy fácil de construir. Le recomiendo dar lectura a la sección correspondiente en el texto básico y luego analizar el siguiente ejemplo.

Para el siguiente conjunto de datos:

5, 5, 5, 5, 8, 8, 10,10, 10, 10, 10, 10, 11, 12, 12, 12, 12, 14, 14, 15

Vamos a elaborar el diagrama de la siguiente forma.

Para los valores menores que 10:

Tallo	Hojas					
0	5	5	5	5	8	8

Para los valores mayores o iguales a 10:

Tallo	Hojas													
1	0	0	0	0	0	0	1	2	2	2	2	4	4	5

Ahora unimos los dos:

Tallo	Hojas													
0	5	5	5	5	8	8								
1	0	0	0	0	0	0	1	2	2	2	2	4	4	5

Y ya tenemos muestro diagrama.

4.3. Cuartiles

Para tratar este tema le recomiendo dar lectura a la sección “Otras medidas de dispersión”. Tome nota de las fórmulas empleadas y observe atentamente los ejemplos.

Los cuartiles son posiciones dentro de un conjunto de datos. Por ejemplo, si tenemos un conjunto de 100 valores, los cuartiles serán la posición 25, la posición 50, la posición 75 y la posición 100. Note que estamos refiriéndonos a la posición, en otras palabras, el cuartil va a ser cualquier valor que se encuentre en esa posición.

Los cuartiles dividen al conjunto ordenado de datos en cuatro partes.

Para calcular las posiciones de los cuartiles se aplica el razonamiento siguiente.

Primer cuartil: Q1= (n+1) (25/100)

N es número de valores observados.

25/100 se utiliza porque el primer cuartil corresponde a la cuarta parte de los valores observados.

Segundo cuartil: Q2 = (n+1)(50/100)

N es el número de valores observados.

50/100 se utiliza porque el segundo cuartil corresponde a la mitad de los valores observados.

Tercer cuartil: Q3 = (n+1)(75/100)

N es el número de valores observados

75/100 se utiliza porque el tercer cuartil corresponde al 75% de los valores observados

Cuarto cuartil: Q4 = (n+1)(100/100)

N es el número de valores observados.

100/100 se utiliza porque el cuarto cuartil corresponde al 100% de los valores observados, esta posición siempre es la última de un conjunto ordenado de valores.

Ejemplos:

Dado el siguiente conjunto de datos:

Tabla 1. *Conjunto de datos ordenado*

150	321	450	540	593	640	720	814	940	1200
180	347	450	560	593	640	720	840	990	1250
230	357	456	560	600	650	729	840	1000	1400
230	360	458	560	610	650	750	840	1000	1400
235	375	467	560	610	670	750	850	1050	1450
250	390	467	570	620	670	756	890	1080	1450
250	390	480	578	620	680	760	895	1100	1480
278	410	494	580	620	700	780	900	1150	1510
280	429	498	582	624	710	800	900	1200	1560

150	321	450	540	593	640	720	814	940	1200
300	450	540	590	640	710	810	900	1200	1600

Obtener el primero, segundo y tercer cuartil.

Primer cuartil: Q1= (n+1) (25/100)

Primer cuartil: Q1= (101) (0,25)

Primer cuartil: Q1= 25,25

Esto significa que nuestro valor está en la posición 25 más 0,25 posiciones. Esto último quiere decir que hay que tomar la cuarta parte del valor que existe entre la posición 25 y 26.

Segundo cuartil: Q2 = (n+1)(50/100)

Segundo cuartil: Q2 = (101) (0,5)

Segundo cuartil: Q2 = 50,5

Nuestro segundo cuartil se encuentra en la posición 50 más el 0,5 de la diferencia entre la posición 50 y 51. En este caso, la posición 50 es 640 y la posición 51 tiene el mismo valor por lo que el segundo cuartil es 640.

Tercer cuartil: Q3 = (n+1)(75/100)

Tercer cuartil: Q3 = (101) (0,75)

Tercer cuartil: Q3 = 75,75

El tercer cuartil está en la posición 75 más tres cuartos de la diferencia entre la posición 75 y 76.

El valor en la posición 75 es 850 dólares y en la posición 76 es 890. La diferencia entre los dos es 40 dólares; el 75% de 40 dólares es 30.

Por lo tanto, al valor de 850 se le suma 30 y el resultado sería 880 como tercer cuartil.

4.4. Deciles

Los deciles dividen al conjunto de observaciones en 10 partes, los conceptos son los mismos que se aplican para el cálculo de cuartiles, existen pequeñas diferencias en las fórmulas que se aplican.

Primer Decil: D1= (n+1) (10/100)

10/100 es igual a 1/10 por lo que la fórmula queda así:

Primer decil: D1= (n+1) (1/10)

Segundo decil: D2= (n+1) (2/10)

Tercer decil: D3= (n+1) (3/10)

Y así sucesivamente para cada decíl del conjunto de datos.

4.5. Percentiles

Con los percentiles ocurre algo similar. La fórmula varía de la siguiente manera:

Primer percentil: P1= (n+1) (1/100)

Segundo percentil: P2= (n+1) (2/100)

Tercer percentil: P3= (n+1) (3/100)

Y así sucesivamente para cada percentil del conjunto de datos.

Actividad recomendada

Resuelva los ejercicios del 11 al 14 del texto básico, estos se encuentran al final de la sección que trata “Otras medidas de dispersión”.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Con los cuartiles se puede desarrollar diagramas de caja. Vamos a complementar este capítulo con una lectura de la sección “Diagramas de caja” del texto básico. De la lectura debe obtener lo siguiente.

a. Observar atentamente los dos ejemplos que se plantean.
b. Determinar para qué se utilizan los diagramas de caja.

4.6. Descripción de la relación entre dos variables

Este es un tema muy sencillo, le recomiendo dar lectura a la sección correspondiente en el texto básico y observar los ejemplos que se plantean.

Si considera necesario observe el siguiente ejercicio.

Suponga que se ha tomado el peso y la edad de 18 niños de una guardería de la ciudad. Los resultados son los siguientes:

Tabla 2. Peso y edad de un grupo de 18 niños

Edad en meses	Peso en libras
1	3,47
1	3,3
1	3,9
2	4,5
2	4,6
2	4,2
3	6,26
3	6,1
3	6,7
4	7,1
4	7,4
4	7,5
5	7,9
5	8
5	7,85

Edad en meses	Peso en libras
6	8,2
6	8,4
6	9

Fuente: XXXX.

Para graficar esta relación de las dos variables (edad en meses y peso en libras) se debe dibujar un plano, en el eje de las x vamos a colocar la edad y en el eje de las y vamos a colocar el peso.

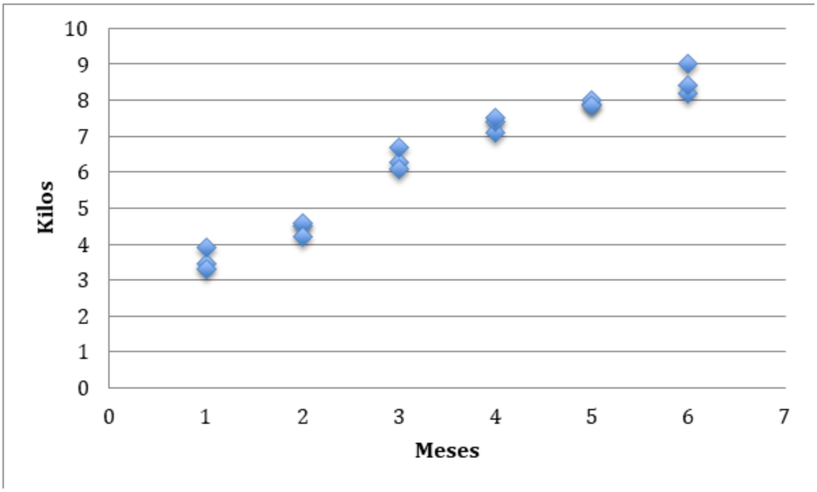

Figura 6. Relación edad peso en niños. *Fuente*: XXXX.

Observe que para el mes 1 se han dibujado los puntos en las posiciones 3,47, 3,3 y 3,9. De esta forma, se procede con todos los valores de la tabla.

Ir al contenido

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Documento 5.

Probabilidades y distribución de probabilidad discreta

En esta sección ponemos especial énfasis en el indicador de aprendizaje "Realiza una distribución de probabilidad" planificada para el primer bimestre. Se trata de procedimientos sencillos que lo invito a revisar y en caso de tener dudas, escríbame sus inquietudes y procederé a apoyarle en su proceso de aprendizaje.

5.1. ¿Qué es la probabilidad?

La probabilidad de un suceso está asociado a cuan posible es que ocurra o no ocurra, en esta unidad vamos a estudiar las probabilidades y como se distribuyen.

> Para tratar este tema le recomiendo dar lectura a la sección "¿Qué es probabilidad?" en el texto básico, ponga énfasis en entender el concepto de probabilidad y si es necesario plantear un par de ejemplos. Tome nota de los conceptos de experimento, evento y resultado.

Aquí diremos que probabilidad es el potencial (expresado en porcentaje) de que ocurra algo; también se puede decir que es una suposición de que algo ocurra expresada en porcentajes.

5.2. Tipos de probabilidad

Al igual que en la sección anterior, aquí es conveniente que usted empiece dando una lectura al texto básico en la sección que corresponde. Para complementar el texto básico vamos a organizar los tipos de probabilidad y a ampliar la explicación sobre probabilidad clásica y empírica.

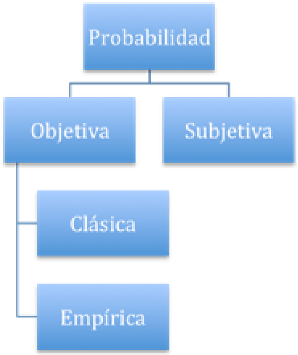

Figura 7. Clasificación de probabilidad. *Fuente*: XXXX.

La probabilidad clásica considera que en un experimento todos los resultados tienen la misma probabilidad de darse. Un ejemplo claro de esto es el siguiente:

Si se cuenta con una bolsa con 10 bolas de distintos colores (ningún color se repite) y se desea sacar una bola, el color que se va a obtener puede ser cualquiera de los 10 que están en la bolsa. La probabilidad de que sea blanco es igual a 1 dividido para el número de colores en la bolsa; es decir, 1 dividido para 10.

Esta probabilidad equivale al número de resultados favorables dividido para el total de resultados posibles.

En la probabilidad empírica interviene un dato muy importante: el número de veces que ya ha sido realizado un experimento y sus resultados. El término empírico en este caso quiere decir: basado en la experiencia, basado en experiencias (experimentos) anteriores.

En el texto existe un ejemplo muy didáctico que le recomiendo analizar por lo menos dos veces. Este ejemplo se refiere a la probabilidad de tener un vuelo al espacio de forma exitosa.

La probabilidad subjetiva no se basa en ninguna información histórica de un hecho o experimento, tampoco considera que los resultados pueden tener todos la misma probabilidad. En este tipo de probabilidad no se cuenta con información. La probabilidad subjetiva se basa en el conocimiento que el investigador tiene del experimento y con base en é,l este asigna un valor de probabilidad.

5.3. Reglas para calcular probabilidades

Para iniciar este tema le recomiendo dar lectura a la sección correspondiente en el texto básico. En cada caso vaya tomando nota y desarrollando nuevamente los ejemplos que se presentan, de esta forma usted podrá relacionar los conceptos con las fórmulas y con las situaciones del diario vivir.

A continuación, le presento un resumen de las reglas de adición.

5.3.1. Reglas de adición

Regla especial de la adición

P(A o B) = P (A) + P (B)

Se matriculan 200 estudiantes en asignaturas de una carrera de la UTPL.

Tabla 3. Matriculados en asignaturas

Matriculados en una asignatura	20	P = 20 / 200 P = 0,1
Matriculados en dos asignaturas	40	P = 40 / 200 P = 0,2
Matriculados en tres asignaturas	60	P = 60 / 200 P = 0,3
Matriculados en cuatro asignaturas	80	P = 80 / 200 P = 0,4
Total	**200**	**P = 1**

Fuente: XXXX.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Si se consulta a uno de esos 200 estudiantes, ¿cuál es la probabilidad de que este estudiante se haya matriculado en una o en dos asignaturas?

P (una o dos) = P (una) + P (dos)

P (una o dos) = 0,1 + 0,2

P (una o dos) = 0,3. El 0,3 equivale al 30% de probabilidades de que un estudiante se haya matriculado en una o dos asignaturas.

Regla general de la adición

P(A o B) = P (A) + P (B) – P (A y B)

Para este caso en el texto básico se presenta un ejemplo muy didáctico que se refiere a si una carta escogida puede ser rey o corazón. Le recomiendo observar detenidamente este ejemplo, ponga especial atención a la aplicación de la fórmula.

5.3.2. Reglas de multiplicación

Regla especial de multiplicación

P (A y B) = P(A) * P(B)

Al revisar la estadística de estudiantes aprobados de la UTPL se determinó que el 75% de los estudiantes aprueban todas las asignaturas en las que se matriculan. Si se seleccionan dos estudiantes de forma aleatoria, ¿cuál es la probabilidad de que los dos hayan aprobado todas las asignaturas?

P (A y B) = P(A) * P(B)

P (A y B) = 0,75 * 0,75

P (A y B) = 0,5625. En este caso se multiplican las dos probabilidades y el resultado es equivalente al 56,25%; es decir el 56,25% es la probabilidad de que los dos estudiantes seleccionados hayan aprobado todas las asignaturas.

Invirtamos el ejemplo y determinemos la probabilidad de que los dos estudiantes no han aprobado todas las asignaturas.

P (A y B) = P(A) * P(B)

P (A y B) = 0,25 * 0,25

P (A y B) = 0,0625

Si la probabilidad de que un estudiante apruebe es de 75%, la probabilidad de que no apruebe es de 25%, que equivale a 0,25.

Esto quiere decir que, si se escogen dos estudiantes aleatoriamente, la probabilidad de que los dos no hayan aprobado todas la asignaturas es de 6,25%

Regla general de multiplicación

P (A y B) = P(A) * P(B | A)

En una caja existen 6 bolas blancas y 6 bolas negras. Si se toman dos bolas una después de otra sin reponer la primera. ¿Cuál es la posibilidad de que las dos sean negras?

P (A y B) = P(A) * P(B | A)

P (A y B) = (6/12) * (5/11) (en este caso quedan 5 negras y el total es 11)

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

P (A y B) = 0,5 * 0,4545

P (A y B) = 0,2272

La probabilidad alcanza un 22,72%.

5.4. Distribuciones de probabilidad discreta

> Esta sección se basa en el capítulo 6 del texto básico, y se centra en distribuir los resultados de un experimento de forma tal que estos puedan observarse como probabilidades. Una distribución de probabilidad entonces, muestra todos los resultados posibles de un experimento y la probabilidad de que cada resultado ocurra.

A continuación, le recomiendo dar lectura a la sección “¿Qué es una distribución de probabilidad?” del texto básico. De la lectura, escriba en su cuaderno de notas los conceptos que vaya encontrando y fíjese muy atentamente en el ejemplo del conteo de número de caras que salen en tres lanzamientos consecutivos de una moneda, anote la tabla resultante en la que se señalan los resultados y las probabilidades asociadas a cada resultado en la gráfica 6-1 del texto básico

Seguidamente lea la sección “Variables aleatorias” y tome nota de:

Concepto de variable aleatoria.

Variable aleatoria discreta (dos ejemplos).

Variable aleatoria continua (dos ejemplos).

Ahora en la sección “Media, varianza y desviación estándar de una distribución de probabilidad discreta” repase el ejercicio desarrollado

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

referente al número de autos vendidos por la agencia John Ragsdale y determine como se calcula la media de probabilidad y la varianza de dicha distribución.

Se trata de temas sencillos por lo que una vez analizado el ejemplo va a comprender sin mayores dificultades.

Ir al contenido

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Documento 6.

Distribución de probabilidad normal

> En esta sección ponemos especial énfasis en el indicador de aprendizaje "Calcula áreas bajo la curva normal" planificada para el segundo bimestre; le recomiendo leer primeramente la guía didáctica y el texto a partir de la sección "La familia de distribuciones de probabilidad normal".

La distribución de probabilidad normal se aplica a todos los datos de tipo numérico y continuo, por ejemplo, la estatura, el peso, el sueldo. Todas estas variables son numéricas y pueden ser decimales. En esta sección se aprende a diferenciar una distribución normal y a identificar sus características. Como aplicación práctica se aprende a determinar áreas bajo la curva normal.

6.1. Distribución normal y familia de distribuciones de probabilidad normal.

Una distribución normal tiene una característica fundamental. Los valores se agrupan alrededor de la media. Muchas veces es difícil entender este concepto, por lo que para aclararlo vamos a complementar la información del texto básico con la siguiente figura.

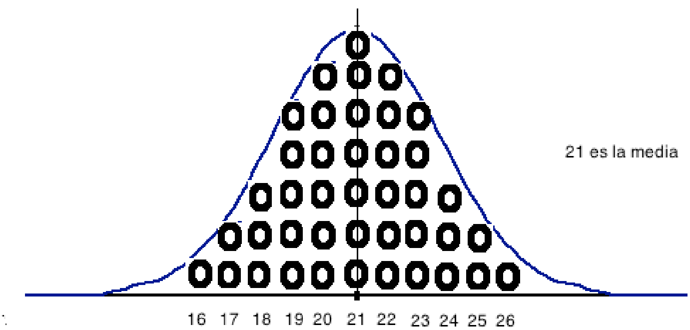

Figura 8. Distribución de valores bajo una curva para que sea normal.
Fuente: XXXX.

Suponga que se hizo una encuesta y se preguntó la edad de los estudiantes de un paralelo de la UTPL. Se obtiene la media aritmética y el resultado es 21 años. En la gráfica cada círculo representa un estudiante, la media se ubica al centro de la figura. Hay 7 estudiantes que tienen 21 años; 6 estudiantes con 22 años; 5 estudiantes con 23 años; 3 estudiantes con 24 años; 2 estudiantes con 25 años y 1 estudiante con 26 años. Algo similar ocurre para los valores inferiores a la media.

Note que la media es el valor que tiene más observaciones y los valores cercanos a la media tienen cada vez menos observaciones; es decir se distribuyen de manera uniforme alrededor de la media; si se cumple esta característica tenemos una distribución normal.

Ahora que hemos entendido que es una distribución normal, vamos a leer la sección "Familia de distribuciones de la probabilidad normal" para determinar las características de la distribución de probabilidad normal y tome nota de ellas.

Ahora es necesario que comprenda los siguientes aspectos.

1. Existen muchas curvas normales, una para cada problema o fenómeno. Ejemplos:

   Una curva para cuando la media de la edad es 21.

   Una curva para cuando la media de la edad es 22.

   Una curva para cuando la media de la edad es 23, etcétera.

2. Es necesario unificar todas las curvas en una sola.

3. Se requiere contar con una unidad de medida estándar para trabajar con una curva estándar. Para esto se utiliza Z que es una unidad de medida del número de desviaciones estándar.

Ahora lo invito a dar lectura a la sección "Distribución de probabilidad normal estándar", en esta sección va a encontrar.

- Qué es el valor z.
- La fórmula para calcular z.
- Cómo leer la tabla de distribución normal.
- Cómo ubicar un valor z en una curva normal.

Ahora vamos a complementar la explicación con lo siguiente: suponga que se preguntó la edad a 41 estudiantes, estos valores se presentan en la siguiente figura.

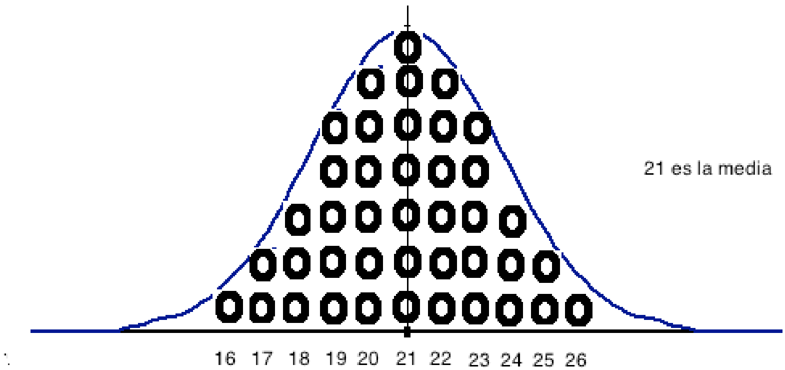

Figura 9. Distribución normal. *Fuente*: XXXX.

La media es 21 y la desviación estándar es 2. ¿Cuál es el valor z para la edad 23? Para ello se aplica la fórmula:

$Z=\frac{X-\mu}{\delta}$	$Z=\frac{23-21}{2}$	$Z=\frac{2}{2}$	$Z=1$

Esto quiere decir que el valor 23 años está a una desviación estándar después de la media. Esto quiere decir que z es el número de desviaciones estándar.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Ahora encontremos el valor z para la edad de 24 años.

$Z=\frac{X-\mu}{\delta}$	$Z=\frac{24-21}{2}$	$Z=\frac{3}{2}$	$Z=1,5$

La edad 24 años se encuentra a 1,5 desviaciones estándar después de la media.

Ahora encontremos el valor z para la edad 19.

$Z=\frac{X-\mu}{\delta}$	$Z=\frac{19-21}{2}$	$Z=\frac{-2}{2}$	$Z=-1$

La edad 19 años se encuentra una desviación estándar **antes** de la media.

Actividad recomendada

Resuelva los ejercicios del 7 al 12 del texto básico, estos se encuentran al final de la sección que trata “Distribución de probabilidad normal estándar”.

6.2. Cálculo de áreas bajo la curva normal

Antes de que de lectura al texto básico vamos dar lectura a la siguiente explicación.

Los valores de las edades (de las observaciones) se encuentran distribuidos dentro de la curva normal. Observe la siguiente figura.

Figura 10. Distribución simétrica de las observaciones. *Fuente*: XXXX.

La mitad de las edades (50%) se encuentra en la parte izquierda de la curva y la otra mitad (otro 50%) en la parte derecha de la curva.

La curva normal sirve, entre otras cosas, para aproximar probabilidades; es decir aproximar proporciones de observaciones que cumplan ciertas características. Por ejemplo:

Suponga que se hizo una encuesta y se preguntó la edad de 41 estudiantes de un paralelo de la UTPL. La media aritmética de la edad es 21 años y la desviación estándar 2 años.

¿Cuántas personas tienen 23 años o más?

¿Cuántas personas tienen entre 18 y 20 años?

Resolver este ejercicio paso a paso.

Lo primero es entender lo que realmente debemos calcular. En el primer caso (primera pregunta) de las 41 personas encuestadas debemos encontrar cuántas tienen 23 o 24 o 25 o 26 o más años de edad.

Para ello procedemos de la siguiente manera.

Calculamos los valores de z para 23 años.

$Z = \frac{X - \mu}{\delta}$	$Z = \frac{23 - 21}{2}$	$Z = \frac{2}{2}$	Z= 1

El resultado de z=1 lo buscamos en la tabla de distribución normal en el apéndice 3B al final del texto básico, el resultado de la tabla es 0,3413 que equivale al 34.13%. Pero este porcentaje equivale a la cantidad de personas cuya edad está comprendida desde 21 años (la media) hasta 23 años. Observe la sección manchada en la siguiente figura.

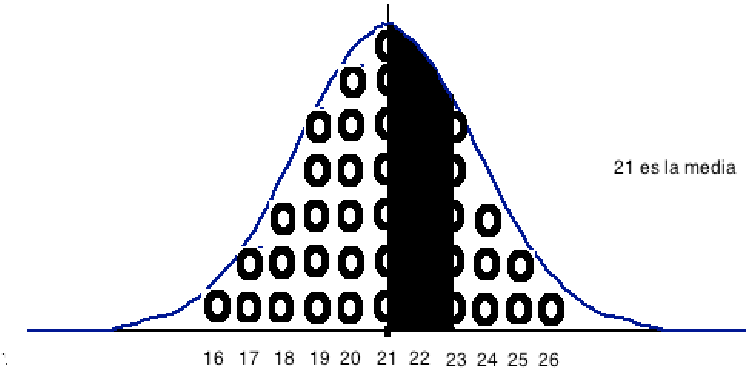

Figura 11. Área entre 21 y 23 *años. Fuente*: XXXX.

Ahora considere que vamos a hacer el cálculo en la parte derecha de la curva, esta contiene el 50% de las observaciones. Lo que hemos calculado es la parte sombreada, pero necesitamos saber a cuánto equivale la parte que no está sombreada que representa a las personas con 23 años o más.

Para obtener el valor lo que hacemos es restar de 50% el valor de la parte sombreada. Esto es:

Proporción = 0,5 – 0,3413.

Proporción = 0,1587 que equivale al 15,87%.

El resultado se puede apreciar en la siguiente gráfica.

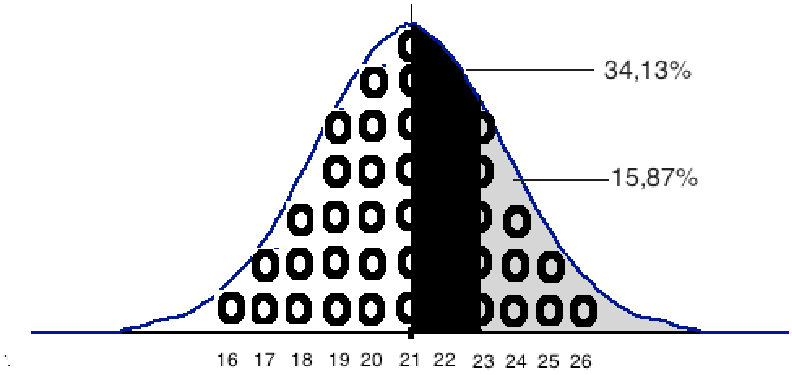

Figura 12. Distribución de áreas antes y después de 23. *Fuente*: XXXX.

Ahora vamos a trabajar con el texto básico, dando lectura a la sección "Determinación de áreas bajo la curva normal" y tomando nota de los hallazgos importantes. Una recomendación para la lectura es: volver a resolver los ejercicios que ya han sido resueltos en el texto básico e ir comparando los procedimientos y resultados.

Si cree necesario más ejemplos a continuación resolvemos otro ejercicio.

Dada una curva normal con media de 500 y una desviación estándar de 50, calcule el área entre 555 y 600.

El paso 1 consiste en entender que, en la curva normal, la mitad de la distribución se halla a la izquierda de la media y la otra mitad a la derecha de la media.

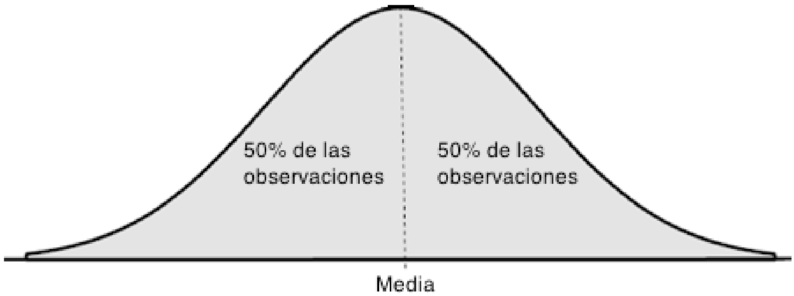

Figura 13. Distribución simétrica. *Fuente*: XXXX.

Como paso 2, observe la imagen y ubique los límites 555 y 600. Es el área entre esos dos valores lo que queremos determinar.

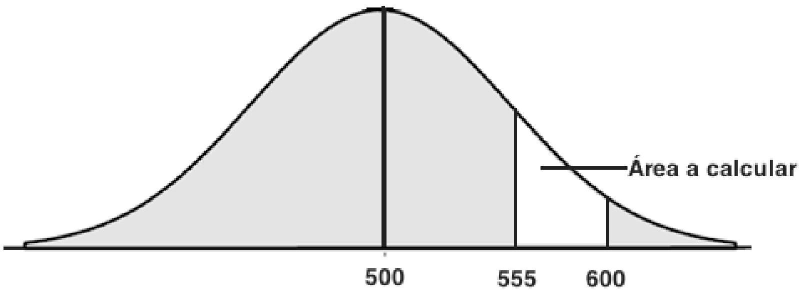

Figura 14. Áreas entre 555 y 600. *Fuente*: XXXX.

El valor de área que buscamos se encuentra en la tabla de distribución normal. Pero para poder determinarla necesitamos conocer los valores de Z para 555 y para 600.

$Z=\frac{X-\mu}{\delta}$	$Z=\frac{555-500}{50}$	$Z=\frac{55}{50}$	Z= 1.10
$Z=\frac{X-\mu}{\delta}$	$Z=\frac{600-500}{50}$	$Z=\frac{100}{50}$	Z= 2

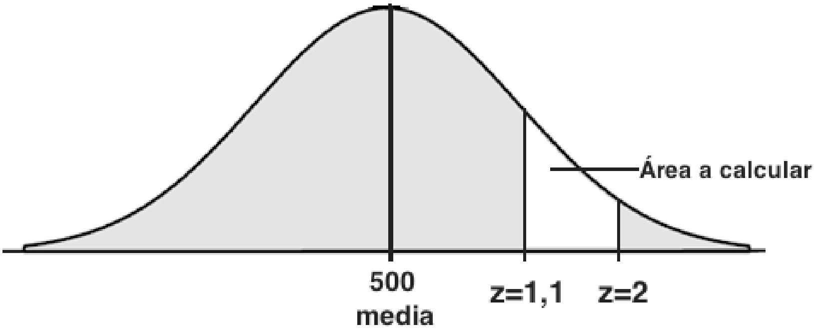

Figura 15. Áreas entre z=1,1 y z=2. *Fuente*: XXXX.

Lo que buscamos se encuentra entre los valores de Z = 1,10 y Z = 2.

Para cada valor de Z, buscamos el porcentaje que el corresponde en la tabla de distribución normal en el apéndice B3 al final del texto básico.

El porcentaje que corresponde a Z = 1,10 es 36,43% (en la tabla consta el valor 0,3643), esto corresponde al porcentaje desde la media hasta z = 1,1.

El porcentaje que corresponde a Z = 2 es 47,72% (en la tabla consta el valor 0,4772), esto corresponde al porcentaje desde la media hasta z = 2.

Lo que debemos hacer es restar: 47,72% - 36,43% = 11,29%

Por tanto, el área bajo la curva que se comprende entre los valores de 555 y 600 es el 11,29%.

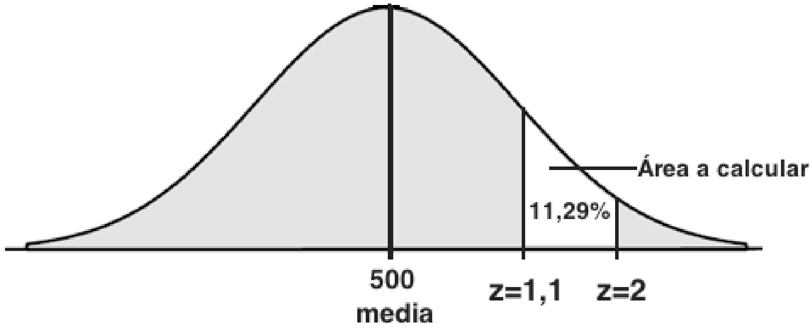

Figura 16. Áreas entre z=1,1 y z=2. *Fuente*: XXXX.

Apreciado estudiante, hasta aquí el desarrollo de la unidad, le recuerdo que en caso de tener dudas respecto a lo que abarca esta guía me puede escribir y consultar.

Ir al contenido

Documento 7.

Muestreo

En esta unidad usted aprenderá a determinar una muestra, su tamaño y un método para obtenerla, comprenderá también los conceptos necesarios que señalan de qué manera una muestra representa fielmente a una población. El objetivo es resultado de aprendizaje “Entender en qué situación se aplica un tipo de muestreo y cuál es el tamaño de la muestra”. Le invito a que revise cuidadosamente el contenido y vaya tomando nota de los aspectos importantes que encuentre. En caso de tener dudas le recuerdo que puede contar con su tutor a través de una consulta en el Entorno Virtual de Aprendizaje.

Anteriormente se había señalado que la estadística inferencial obtiene conclusiones para una población en base a los datos de una muestra. En esta unidad nos vamos a centrar en la selección de esa muestra para que los resultados puedan generalizarse y tener la certeza de la veracidad de las conclusiones.

7.1. Razones para muestrear

> En el apartado “Métodos de muestreo” del texto básico se señalan cinco razones de peso que le solicito lea, es muy importante que tenga claro por qué es más conveniente levantar información de un grupo de elementos que de todo el universo.

Las razones para muestrear se resumen a continuación:

- Costos
- Tiempo
- Algunas poblaciones son infinitas.
- Pruebas de naturaleza destructiva.
- Resultados adecuados.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

7.2. Tipos de muestreo

En el texto básico se señalan cuatro tipos de muestreo, lea detenidamente cada uno de los tipos y proponga dos ejemplos de cada uno; en los ejemplos asegúrese de que exista una razón que justifique el uso de ese tipo de muestreo.

Los tipos de muestreo son:

- Aleatorio simple.
- Aleatorio sistemático.
- Aleatorio estratificado.
- Muestreo por conglomerados.

Ejercicio

Ahora analice el siguiente caso.

En la ciudad de Quito se va a elegir alcalde y se desea hacer un estudio que determine cómo van a votar los ciudadanos. Si el 65% de los electores pertenecen al área urbana y el 35% al área rural, ¿qué tipo de muestreo se debe utilizar? Justifique de forma clara la respuesta.

En este ejercicio se requiere que la muestra sea representativa; es decir, que represente a toda la población (urbana y rural), esto significa que las proporciones de encuestados deberían mantenerse en el área urbana y en el área rural. Esto significa que, si se va a tomar una muestra de 1.000 personas, entonces 650 deben ser del área urbana y 350 del área rural, de esa forma en la muestra se representan los habitantes de los dos sectores.

7.3. Error de muestreo

Para empezar este apartado le recomiendo dar lectura a la sección "Error de muestreo" del texto básico.

Una vez realizada la lectura analicemos lo siguiente.

En una ciudad muy pequeña con 2.000 habitantes mayores de 18 años, se mide el peso de un grupo de 400 personas (una muestra). El peso promedio de las 400 personas es 69 kilos; días después se decide medir el peso de todas las personas de la ciudad que sean mayores de 18 años (2.000 personas) y el peso promedio resultante es de 70 kilos.

¿Por qué existe una diferencia?

¿Cómo se llama esa diferencia?

7.4. Tamaño de una muestra

El tamaño de la muestra es la cantidad de elementos a los que se debe estudiar para tener una estimación adecuada de una población. Ejemplo:

Durante la campaña para la elección de alcalde de la ciudad de Quito se necesita saber qué candidato tiene más opción de ganar. Para averiguar esto hay que hacer un estudio aplicando una encuesta, a más de las preguntas que se deben hacer constar en la encuesta, es necesario saber a cuántas personas se les va a preguntar; es decir, necesitamos saber cuál es el tamaño de la muestra.

En el mismo capítulo del texto básico, se encuentra el apartado "Elección del tamaño adecuado de una muestra". Lea el mismo y note que hay dos tipos de cálculos.

7.4.1. Tamaño de la muestra para calcular una media poblacional

En este caso se busca determinar un valor numérico que estime una medida de la población; es decir, se busca variables como la edad, el peso, el salario, etc., de una población.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

¿A cuántos jefes de familia hay que preguntar para conocer el salario promedio de la ciudad de Loja? Para responder a esta pregunta analice el siguiente ejemplo.

Ejercicio

En una ciudad pequeña se desea determinar el salario medio de los trabajadores. El error máximo que se permitirá es de 100 dólares y el nivel de confianza de 95%. Se conoce que en estudios anteriores aplicados en la ciudad la desviación estándar fue 120. ¿Cuántas personas debe abarcar la muestra?

La fórmula señala:

$n = ((z.\delta)/E)^2$

Z es el nivel de confianza y equivale a un valor de 1,96

- es la desviación estándar de la población, en este caso 120

E es el error que se va a permitir que alcanza 40 dólares

$n = ((z.\delta)/E)^2$

$n = ((1{,}96 * 150)/40)^2$

$n = (294/40)^2$

$n = (7{,}35)^2$

n = 54. El número de personas que conforman la muestra es 54

En el texto se señala otro ejercicio que ya se encuentra resuelto, le recomiendo revisarlo paso a paso e ir tomando nota del procedimiento.

7.4.2. Tamaño de la muestra para calcular una proporción de una población

Siempre que se hace levantamiento de opinión pública (¿qué opina de esto?, o ¿cuál es su candidato favorito?, etc.) es necesario determinar una proporción de la población que se utilice como muestra con la que se trabaje. La fórmula a aplicar es:

n = p.q(z/E)2

p y q. Generalmente a p y q se les asigna 0,5, esto maximiza el tamaño de la muestra; se utiliza estos valores cuando no se dispone de información previa.

Z. representa el nivel de confianza, 1,96 equivale al 95%.

Ejercicio

Suponga que el presidente de Estados Unidos desea un cálculo de la proporción de la población que apoya su actual política relacionada con las revisiones del sistema de seguridad social. El presidente quiere que el cálculo se encuentre a menos de 0.04 de la proporción real. Suponga un nivel de confianza de 95%. Los asesores políticos del presidente calculan que la proporción que apoya la actual política es de 0,60.

a. ¿De qué tamaño debe ser la muestra que se requiere?
b. ¿De qué tamaño debe ser una muestra si no hubiera disponible ningún estimador de la proporción que apoya la actual política?

n = p.q(z/E)2

n = 0,6 * 0,4 (1,96/0,04)2

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

n = 0,6 * 0,4 (2401) → 0,6 equivale a la proporción de personas que apoyan la actual política

n = 576,24

El tamaño de la muestra debe ser de 577 encuestas.

Actividad recomendada

Resolver los ejercicios 19 a 24 que se plantean al final de la sección “Tamaño de la muestra para calcular la proporción de una población” del texto básico.

7.5. Distribución de probabilidad

Aquí nos vamos a centrar en definir y entender qué es, cómo se obtiene y qué nos indica una distribución de probabilidad. Para ello es necesario remitirnos al capítulo 6 del texto básico y dar lectura a la introducción. Tome nota del concepto de distribución de probabilidad y de las características de una distribución de probabilidad.

Para complementar su lectura vamos a continuación a señalar lo siguiente.

Una distribución de probabilidad es una lista de todos los posibles resultados que se obtienen en un experimento y la probabilidad de cada uno.

Supongamos el ejemplo de un dado: Este tiene seis lados.

La probabilidad de que caiga 1 es de 1/6.

La probabilidad de que caiga 2 es de 1/6.

La probabilidad de que caiga 3 es de 1/6.

La probabilidad de que caiga 4 es de 1/6.

La probabilidad de que caiga 5 es de 1/6.

La probabilidad de que caiga 6 es de 1/6.

Lo que acabamos de señalar constituye una especie de tabla de distribución de probabilidad, lo único que hace falta es expresarla de manera formal:

Tabla 4. Probabilidad en lanzamiento de un dado

Resultado	Probabilidad
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Fuente: XXXX.

Al iniciar el capítulo 6 del texto básico se encuentra descrito un ejemplo en el que se busca determinar el número de caras que aparecen en tres lanzamientos de una moneda. Si los resultados posibles son: cero caras, una cara, dos caras y tres caras, ¿cuál sería la distribución de probabilidad de este experimento?

La resolución del problema inicialmente señala todas las posibles combinaciones que se pueden dar con tres lanzamientos (ver la tabla, p. 156), hay que recordar que lo que se busca es determinar cuántas caras caen con tres lanzamientos de la moneda. Si se observa con atención, en la última columna de la tabla se pueden observar el número total de caras que se han obtenido con los tres lanzamientos de la moneda. En el mismo ejemplo se encuentra la tabla 6.1, note que esta tabla contiene la distribución de probabilidad del experimento; en la primer columna se tiene el número de caras que pueden caer al lanzar tres veces la moneda; y, en la segunda columna se tiene la probabilidad de que caiga ese número de caras.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

La gráfica 6.1 del texto básico no es otra cosa que la representación de la tabla 6.1. Aquí lo importante es observar la forma que adopta la distribución. Como se puede observar esta forma se asemeja a la de la distribución normal.

Ir al contenido

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Documento 8.

Distribución muestral y teorema central del límite

En esta sección vamos a trabajar en comprender en qué consiste la distribución de la media muestral y el teorema central del límite, tener claro este tema permite comprender por qué el muestreo es representativo de una población y por qué a mayor tamaño de la muestra mayor tendencia a una forma normal para cualquier tipo de población. Se trata de un tema interesante que le animo a desarrollar, como siempre, en caso de tener dudas cuente con su tutor para resolverlas.

8.1. Distribución muestral de la media

> Ahora vamos a analizar un ejercicio, este se encuentra en el capítulo 8 del texto básico en la sección "Distribución muestral de la media". Inicialmente se recomienda leer todo el ejercicio y posteriormente complementamos con los siguientes comentarios y aclaraciones.

Para empezar, hay que tener claro que en este ejemplo vamos a observar la forma que toma la distribución de todas las medias de las muestras de tamaño 2. La población la constituyen un total de siete empleados; la tabla 8.2 muestra el listado de los empleados y el salario que estos tienen por horas.

En la siguiente tabla (tomada del texto básico) se presentan todas las muestras de tamaño 2 y la media de cada muestra.

Muestra	Empleados	Ingresos por hora	Suma	Media	Muestra	Empleados	Ingresos por hora	Suma	Media
1	Joe, Sam	$7, $7	$14	$7.00	12	Sue, Bob	$8, $8	$16	$8.00
2	Joe, Sue	7, 8	15	7.50	13	Sue, Jan	8, 7	15	7.50
3	Joe, Bob	7, 8	15	7.50	14	Sue, Art	8, 8	16	8.00
4	Joe, Jan	7, 7	14	7.00	15	Sue, Ted	8, 9	17	8.50
5	Joe, Art	7, 8	15	7.50	16	Bob, Jan	8, 7	15	7.50
6	Joe, Ted	7, 9	16	8.00	17	Bob, Art	8, 8	16	8.00
7	Sam, Sue	7, 8	15	7.50	18	Bob, Ted	8, 9	17	8.50
8	Sam, Bob	7, 8	15	7.50	19	Jan, Art	7, 8	15	7.50
9	Sam, Jan	7, 7	14	7.00	20	Jan, Ted	7, 9	16	8.00
10	Sam, Art	7, 8	15	7.50	21	Art, Ted	8, 9	17	8.50
11	Sam, Ted	7, 9	16	8.00					

Figura 17. Muestras del tamaño 2 del ejercicio del texto básico. Recuperado de Lind, D., Marchal, W., y Wathen, S. (2015). Estadística aplicada a los negocios y economía. Loja Ecuador: McGrawHill-UTPL

Es necesario recalcar que la distribución de la media muestral se da en base a todas las muestras de determinado tamaño, en este caso tenemos todas las posibles muestras de tamaño 2. Para cada muestra se obtiene la media del salario por hora.

El siguiente paso consiste en agrupar en una tabla todas las medias y sus respectivas frecuencias, esto se puede observar en la tabla 8.4 del texto básico:

En la fila 1 se tiene la media 7, esta se repite tres veces y el porcentaje respecto al total es 0,1429.

En la fila 2 se tiene la media 7,50, esta se repite nueve veces y el porcentaje respecto al total es 0,4285.

En la fila 3 se tiene la media 8, esta se repite seis veces y el porcentaje respecto al total es 0,2857.

En la fila 4 se tiene la media 8,50, esta se repite tres veces y el porcentaje respecto al total es 0,1429.

En la parte final del ejemplo se pueden observar dos gráficas, en la primera de ellas es necesario explicar lo siguiente.

La primera tabla corresponde al listado de empleados y su salario.

La segunda tabla corresponde a los salarios que se ganan y la frecuencia.

La gráfica de la derecha corresponde la tabla de frecuencias.

Note la forma que tiene la gráfica, aunque tiende a ser normal los dos primeros valores son iguales.

Empleado	Salario
Joe	7
Sam	7
Sue	8
Bob	8
Jan	7
Art	8
Ted	9

Media	Frecuencia
7	3
8	3
9	1

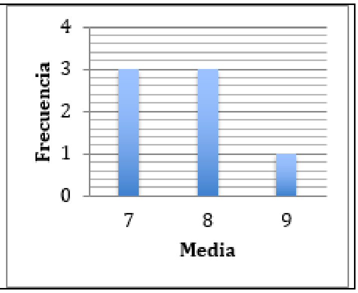

Figura 18. Frecuencia de los salarios. *Fuente*: XXXX.

En el siguiente caso, al hacer la distribución de la media muestral la forma de la gráfica cambia respecto a la anterior, adquiere una forma más parecida a la normal.

Media muestral	Frecuencia
7	3
7,5	9
8	6
8.5	3

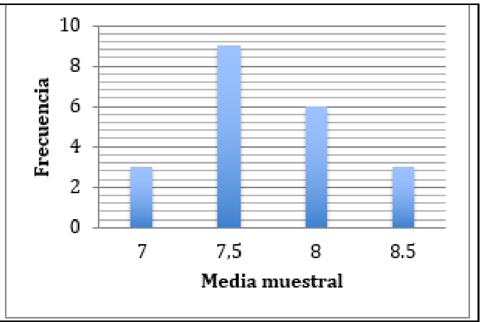

Figura 19. Distribución de la media muestral. *Fuente*: XXXX.

Ahora es necesario analizar las diferencias y entender por qué se dan. En el primer caso tenemos las frecuencias de los salarios de los 7 empleados; en el segundo caso tenemos la distribución de las medias de todas las muestras de tamaño 2. ***Se puede concluir que la distribución de la media muestral tiende a adoptar una forma normal.***

También hay que considerar algunos aspectos que se señalan en el texto en la página 232, estos se refieren a que la media de las medias muestrales es igual a la media poblacional, esto se pudo demostrar con el ejemplo anterior.

La dispersión que se obtiene de la distribución muestral es menor que la dispersión de la población y finalmente, lo que ya señalamos, la distribución muestral de la media tiene forma de campana, se aproxima a la distribución normal.

Una vez analizados los ejemplos resueltos del texto, vamos a resolver un ejercicio.

- Una población consta de los siguientes 4 valores: 11, 12, 14 y 16.
- Enumerar todas las muestras de tamaño 2 y calcular la media de cada muestra.

Calcular la media de la distribución muestral de la media y la media de la población; comparar luego los dos valores.

En vista de que hay valores repetidos y para diferenciarlos vamos a ingresarlos en una tabla de la siguiente forma:

a 12, b 12, c 14, d 16

Inicialmente hay que determinar todas las muestras de tamaño 2.

Tabla 1. Muestras de tamaño 2

Combinaciones	Valor 1	Valor 2	Media de la muestra
a, b	12	12	12
a, c	12	14	13
a, d	12	16	14
b, c	12	14	13
b, d	12	16	14
c, d	14	16	15

Fuente: XXXX.

Ahora construimos la tabla de frecuencias de las medias muestrales.

Tabla 2. Probabilidad de la media muestral

Media muestral	Frecuencia	Probabilidad
12	1	0,1666
13	2	0,333
14	2	0,333
15	1	0,1666

Fuente: XXXX.

Ahora vamos a responder los requerimientos del ejercicio:

a. Todas las combinaciones se pueden ver en la tabla anterior
b. La media de la población es: U= (12+12+14+16)/4 = 13,5.

La media de las medias muestrales es: (12 +13 + 14 + 13 + 14 +15)/6 = 13,5 En este caso las dos medias son iguales.

Ejercicios de este tipo se encuentran en el texto básico al final del apartado "distribución muestral de la media".

Vamos a resolver otro de los ejercicios.

Una población consta de los siguientes cinco valores: 2, 2, 4, 4 y 8.

a. Enumere todas las muestras de tamaño 2 y calcule la media de cada muestra
b. Calcular la media de la distribución muestral de la media y la media de la población; comparar luego los dos valores.

En vista de que hay valores repetidos y para evitar confundirlos vamos a identificarlos mediante la siguiente tabla:

a 2, b 2, c 4, d 4, e 8

Luego procedemos a hacer todas las combinaciones de tamaño 2 (todas las muestras posibles de tamaño 2).

Tabla 3. Muestras de tamaño 2

Combinaciones	Valor 1	Valor 2	Media de la muestra
a, b	2	2	2
a, c	2	4	3
a, d	2	4	3
a, e	2	8	5
b, c	2	4	3
b, d	2	4	3
b, e	2	8	5
c, d	4	4	4
c, e	4	8	6
d, e	4	8	6

Fuente: XXXX.

Para calcular la media de la distribución muestral hay que calcular la media de todos los valores de la última columna.

(2+3+3+5+3+3+5+4+6+6) / 10 = 4

Ahora calculamos la media de la población.

(2+2+4+4+8)/5 = 4

Los valores son iguales.

8.2. Teorema central del límite

Este teorema señala que si de una población se toman todas las muestras posibles de un determinado tamaño, la distribución muestral de la media aritmética presenta una tendencia a asemejarse a una distribución normal. Esta semejanza es mayor mientras mayor sea el tamaño de la muestra.

Para una mejor explicación vamos a plantear un ejemplo.

Una empresa tiene 10 empleados. Las edades son las siguientes:

29, 25, 26, 29, 31, 34, 33, 38, 45, 49

Si se toman todas las muestras posibles de tamaño 2, la distribución de la media muestral va a tener una apariencia de distribución normal.

Si se toman todas las muestras posibles de tamaño 3, la distribución de la media muestral va a tener una apariencia muestral y esta apariencia será mayor que la que obtiene si el tamaño de la muestra es 2.

Si se toman todas las muestras posibles de tamaño 4, la distribución de la media muestral va a tener una apariencia muestral y esta apariencia será mayor que las anteriores de tamaño 2 y 3.

Observe con atención la gráfica 8.3 del texto básico.

Para entenderla hay que revisarla de forma vertical y se puede observar 4 columnas. La primera fila contiene el tipo de distribución que tiene la población; ahora observe a partir de la segunda fila, si se tiene muestras de tamaño 2, las distribuciones muestrales de la media tratan de parecerse a una distribución normal.

Observe ahora la fila 3 y note que el tamaño de las muestras ha aumentado a 6 y la figura que se obtiene es ya de tipo normal.

Observe ahora la fila 4 y tome en cuenta que el tamaño de la muestra se incrementa a 30 y, por supuesto, la forma de la distribución es normal.

Se puede concluir que mientras mayor es el tamaño de la muestra, la forma que adquiere la distribución se parece más a una distribución normal.

Vamos a desarrollar ejercicios que se encuentran en la página 239 del texto básico, la idea es resolverlos y dar las pistas necesarias para que usted pueda trabajar con los ejercicios restantes.

Scraper Elevator Company tiene 20 representantes de ventas, que distribuyen su producto en EE. UU. y Canadá. La cantidad de unidades que el mes pasado vendió cada representante se incluye a continuación. Suponga que estas cifras representan los valores de la población.

2, 3, 2, 3, 3, 4, 2, 4, 3, 2, 2, 7, 3, 4, 5, 3, 3, 3, 3, 5

a. Trace una gráfica que muestre la distribución de la población

b. Calcule la media de la población

c. Seleccione cinco muestras de tamaño 5 cada una. Calcule la media de cada muestra. Utilice los métodos descritos en el capítulo y en el apéndice B.4 para determinar los elementos que deben incluirse en la muestra.

d. Compare la media de la distribución muestral de medias con la poblacional. ¿Esperaría que los valores fueran aproximadamente iguales?

e. Trace un histograma de las medias muestrales. ¿Nota alguna diferencia en la forma de la distribución muestral de las medias en comparación con la forma de distribución de la población?

a. Desarrollando el primer literal es necesario construir primero una tabla de frecuencias:

Tabla 4. Tabla de frecuencias

Valor	Frecuencia	Frecuencia relativa
2	5	0.25
3	9	0.45
4	3	0.15
5	2	0.1
7	1	0.05
	20	1

Fuente: XXXX.

Y con la tabla se puede construir la gráfica que muestra la distribución de la población.

Figura 20. Distribución de la población. *Fuente*: XXXX.

b. Ahora calculamos la media de la población.

U = (2+3+…3+5)/20 = 3.3

c. Ahora vamos a seleccionar 5 medias muestrales de forma aleatoria.

Tabla 5. Muestras de tamaño 5 con media muestral

Valor1	Valor2	Valor3	Valor4	Valor5	Media muestral
2	2	3	2	3	2.4
3	3	4	2	4	3.2
2	2	2	5	3	2.8
3	3	4	3	3	3.2
3	2	3	2	3	2.6

Fuente: XXXX.

d. Ahora obtenemos la media de las muestras:

(2.4 + 3.2 + 2.8 + 3.2 + 2.6)/5 = 2.84

Al comparar la media muestral con la media poblacional los valores son aproximadamente iguales, existe una diferencia mínima, esto se debe a que solo se han tomado 5 muestras y lo ideal es tomar todas las muestras de tamaño 5 que se puedan obtener.

e. Ahora comparamos las gráficas.

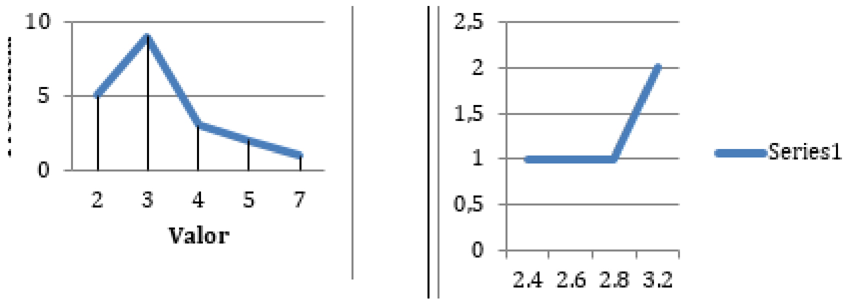

Figura 21. Comparación de gráficas. *Fuente*: XXXX.

Las diferencias se presentan debido a que no se ha trabajado con todas las muestras de tamaño 5, se trabajó únicamente con 5.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Vamos a resolver otro ejercicio (Tomado del texto básico)

El apéndice B4 es una tabla de números aleatorios uniformemente distribuidos. De ahí que cada dígito tenga la misma probabilidad de presentarse.

a. Trace una gráfica que muestre la distribución de la población. ¿Cuál es la media de la población?

b. A continuación, se registran los 10 primeros renglones de cinco dígitos del apéndice B4; suponga que se trata de 10 muestras aleatorias de cinco valores cada una. Determine la media de cada muestra y trace una gráfica (histograma). Compare la media de la distribución muestral de las medias con la media de la población.

Tabla 6. Diez muestras de tamaño 5

0	2	7	1	1
9	4	8	7	3
5	4	9	2	1
7	7	6	4	0
6	1	5	4	5
1	7	1	4	7
1	3	7	4	8
8	7	4	5	5
0	8	9	9	9
7	8	8	0	4

Fuente: XXXX.

Para resolver el primer literal determinamos la media de la población y realizamos la gráfica.

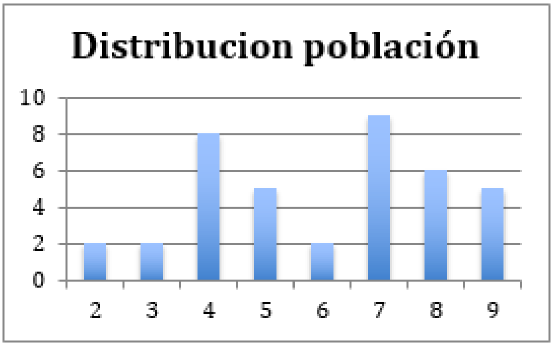

Figura 22. Distribución de la población. *Fuente*: XXXX.

La media de todos los valores (media poblacional) es: 4,84

En cuanto al segundo literal tenemos la tabla siguiente cuya última columna está la media de cada muestra (media muestral), en esa tabla se puede observar en la segunda columna utilizada para dibujar el histograma. Note que en el histograma se puede apreciar ya una forma normal.

0	2	7	1	1	2.2
9	4	8	7	3	6.2
5	4	9	2	1	4.2
7	7	6	4	0	4.8
6	1	5	4	5	4.2
1	7	1	4	7	4
1	3	7	4	8	4.6
8	7	4	5	5	5.8
0	8	9	9	9	7
7	8	8	0	4	5.4

Media muestral	Frecuencia
2.2	1
4	1
4.2	2
4.6	1
4.8	1
5.4	1
5.8	1
6.2	1
7	1

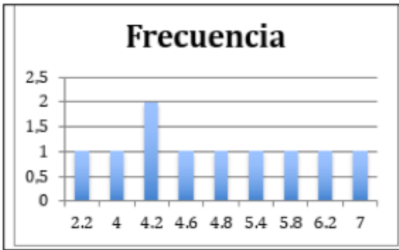

Figura 23. Frecuencia de las medias muestrales. *Fuente*: XXXX.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

La media de todas las muestras es 4,84; es decir, coincide con la media de la población que es 4,84. Esto confirma los conceptos hasta ahora revisados.

Ir al contenido

Documento 9.

Hipótesis y pruebas de hipótesis

> En esta unidad vamos a aprender a verificar hipótesis siguiendo cada uno de los pasos. Para esto es necesario entender antes una serie de conceptos que complementan. Para iniciar es necesario que dé lectura al capítulo "Prueba de hipótesis de una muestra" del texto básico.

Al terminar la lectura asegúrese de tener claros los conceptos hipótesis, prueba de hipótesis, pasos para resolver una hipótesis, hipótesis nula, hipótesis alternativa, nivel de significancia y estadístico de prueba. Estos conceptos ya se encuentran desarrollados en el texto y es necesario conocerlos en detalle. Quiero motivarle a que en caso de tener dudas me escriba, mi trabajo es ayudarle a lograr sus objetivos de aprendizaje y, por lo tanto, estaré presto a responder sus dudas.

9.1. Pasos para verificar una hipótesis

Los pasos para verificar una hipótesis se señalan claramente en el texto básico y se puede observar un ejemplo que se va desarrollando conforme se van explicando los pasos. En el apartado "Pruebas de la media de una población: se conoce la desviación estándar poblacional" se encuentra un ejemplo que señala lo siguiente.

"Jamestown Steel Company fabrica y arma escritorios y otros muebles para oficina en diferentes plantas en el oeste del estado de Nueva York. La producción semanal del escritorio modelo A325 en la planta de Fredonia tiene una distribución normal, con una media de 200 y una desviación estándar de 16. Hace poco, con motivo de la expansión del mercado, se introdujeron nuevos métodos de producción y se contrató a más empleados. El vicepresidente

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

de fabricación pretende investigar si hubo algún cambio en la producción semanal del escritorio modelo A325. En otras palabras, ¿la cantidad media de escritorios que se produjeron en la planta de Fredonia es diferente de 200 escritorios semanales con un nivel de significancia de 0.01?”.

La tarea consiste en seguir paso a paso la resolución de este ejercicio e ir tomando nota de los pasos necesarios para resolverlo y a continuación se presentan algunas pistas que ayudan a comprender el proceso de resolución. Recuerde, es necesario que primero lea y siga paso a paso la resolución del ejercicio.

En el paso 1 es necesario plantear las hipótesis nula y alternativa.

Lo que se va a probar es la hipótesis nula. Para construir las hipótesis se recomienda seguir las siguientes indicaciones.

Hay que leer atentamente el enunciado, de él se obtienen los insumos para construir las hipótesis. Generalmente la pregunta del enunciado es la que nos resuelve la estructura de las hipótesis. En el ejemplo que estamos tratando se pide: “¿la cantidad media de escritorios que se produjeron en la planta de Fredonia es diferente de 200 escritorios semanales…”

Note que resalta la **palabra diferente (≠)** que es la que afecta directamente a la hipótesis alternativa, la misma que queda así: $H1: U \neq 200$.

Por lo tanto, en la hipótesis nula la comparación debe abarcar todas las posibilidades restantes, en este caso lo único que nos resta es la posibilidad de que sean iguales, por lo que la H0 queda así: $H0: U = 200$.

Organizando las Hipótesis tenemos:

$H0: U = 200$

$H1: U \neq 200$

Es necesario señalar que la hipótesis nula siempre afirma que no hay cambio en la media poblacional.

Una explicación adicional requiere el paso 4 que consiste en formular una regla de decisión. Hay que indicar cuando se trata de una prueba de dos colas y cuando de una cola.

Pruebas de dos colas: cuando en la hipótesis nula tenemos = y en la alternativa tenemos ≠. Esto aplica cuando en el enunciado se pregunta si hay cambio en la media de la población.

Cuando tenemos dos colas, el nivel de significancia, que es el error permitido, se divide para dos: 0,01 / 2 = 0,005.

El valor 0,005 es el error permitido a cada lado de la curva, se ubica a los extremos (Observar la gráfica 10.4 del texto básico)

El valor de 0,495 es el porcentaje desde la media hasta donde inicia el error, hay que buscar el valor Z que le corresponde en la tabla de distribución normal que equivale a 2,575. Este valor se ubica a cada lado de la curva (en el sitio en donde inicia el error) y se coloca el signo negativo al valor que va a la izquierda de la curva y positivo al valor que va a la derecha de la curva.

De este modo ya tenemos la regla de decisión lista: si el valor de Z que se calcula es menor que -2,576 o mayor a 2,576 significa que se rechaza la hipótesis nula, en caso de que se encuentre entre los valores -2,576 y 2,576 significa que se debe aceptar la hipótesis nula.

Prueba de una sola cola: cuando el nivel de significancia, que es el error permitido, se va a un solo extremo de la curva. El lado de la curva en el que se ubica el error depende de la pregunta que se plantea en el enunciado del problema.

Si se tiene una pregunta como: ¿Se puede concluir que el número de escritorios que ahora se fabrican es mayor? Hay que considerar que se está preguntando si la media de la población se ha incrementado.

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

Por lo tanto, se debe partir de la hipótesis alternativa señalando que la producción se ha incrementado:

Hipótesis alternativa H1: U > 200.

Por lo tanto, la hipótesis nula contemplara las opciones restantes, es decir ≤ (menor o igual).

H0: U ≤ 200

Ordenando las hipótesis tenemos:

H0: U ≤ 200

H1: U > 200

En este caso el símbolo > de la H1 nos indica que el error se ubica en la parte derecha de la curva normal.

9.2. Operadores de la Hipótesis Alternativa H1

A continuación, presento una tabla, en la que se indica el operador relacional que se debe utilizar dependiendo de la pregunta que se plantea en el enunciado del problema:

Tabla 7. Operadores a utilizar en las hipótesis

	Si la pregunta del enunciado es	utilizar este operador	en la Hipótesis nula o alternativa
1	Es mayor que	>	H1
2	Es menor que o Es más pequeña	<	H1
3	No es mayor No es más que	£	H0
4	Por lo menos	≥	H0
5	Se ha incrementado	>	H1

	Si la pregunta del enunciado es	utilizar este operador	en la Hipótesis nula o alternativa
6	Existe diferencia	≠	H1
7	No ha cambiado	=	H0

Nota. Los operadores se utilizan en la hipótesis indicada según el enunciado del problema. *Fuente:* Tomado de la presentación correspondiente al capítulo 10 del texto básico. Lind, D., Marchal, W., Wathen, S. (2015). *Pruebas de hipótesis de una muestra*[Slides]. McGrawHill-UTPL, Loja Ecuador.

Para poner en práctica los lineamientos de la tabla vamos a resolver el paso 1 (plantear las hipótesis) de los siguientes ejercicios del texto básico.

El fabricante de neumáticos radiales con cinturón de acero X-15 para camiones señala que el millaje medio que cada uno recorre antes de que se desgasten las cuerdas es de 60 000 millas. La desviación estándar del millaje es de 5 000 millas. La Crosset Truck Company compró 48 neumáticos y comprobó que el millaje medio para sus camiones es de 59 500 millas. ¿La experiencia de Crosset es diferente de lo que afirma el fabricante en el nivel de significancia de 0.05?

En este ejercicio la pregunta claramente pregunta si es diferente, por lo tanto, se aplica la fila 6 de la tabla y se debe colocar el símbolo ≠ en la hipótesis alternativa, por lo que las hipótesis quedarían así:

H0: U = 60.000

H1: U ≠ 60.000 (En este caso, el error se reparte a ambos lados de la curva)

La cadena de restaurantes MacBurger afirma que el tiempo de espera de los clientes es de 8 minutos con una desviación estándar poblacional de 1 minuto. El departamento de control de calidad halló en una muestra de 50 clientes en Warren Road MacBurger que el tiempo medio de espera era de 2.75 minutos. Con el nivel de significancia de 0.05, ¿puede concluir que el tiempo medio de espera sea menor a 3 minutos?

En este caso la pregunta señala la frase ***menor a 3 minutos***, por lo tanto, se debe aplicar la fila 2 y utilizar el símbolo < en la H1, las hipótesis quedarían:

H0: U ≥ 3.

H1: U < 3 (en este caso el error se ubica en la parte izquierda de la curva).

Una encuesta nacional reciente determinó que los estudiantes de secundaria veían en promedio (media) 6.8 películas en DVD al mes, con una desviación estándar poblacional de 0,5 horas. Una muestra aleatoria de 36 estudiantes universitarios reveló que la cantidad media de películas en DVD que vieron el mes pasado fue de 6,2. Con un nivel de significancia de 0,05, ¿puede concluir que los estudiantes universitarios ven menos películas en DVD que los estudiantes de secundaria?

En este ejemplo la pregunta señala ***menos que***, por lo tanto, se debe aplicar la fila 2 de la tabla y colocar el símbolo < en H1, las hipótesis quedarían así:

H0: U ≥ 6,8,

H1: U < 6,8 (en este caso el error se ubica en la parte izquierda de la curva).

Finalmente vamos a analizar este ejercicio.

En el momento en que fue contratada como mesera en el Grumney Family Restaurant, a Beth Brigden le dijeron: “Puedes ganar en

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

promedio más de $80 al día en propinas." Suponga que la desviación estándar de la distribución de población es de $3.24. Los primeros 35 días de trabajar en el restaurante, la suma media de sus propinas fue de $84.85. Con el nivel de significancia de 0.01, ¿la señorita Brigden puede concluir que gana un promedio de más de $80 en propinas?".

La palabra clave del enunciado es ***más que***, por lo tanto, hay que aplicar la fila 1 de la tabla y colocar en H1 el símbolo >. Las hipótesis quedarían así:

H0: U £ 80.

H1: U > 80 (en este caso el error se ubica en la parte derecha de la curva, porqué el operador de la H1 indica hacia el lado derecho).

9.3. Verificación de hipótesis cuando se conoce la desviación estándar de la población

Si se conoce la desviación estándar de la población se utiliza el estadístico Z.

Para ampliar las explicaciones vamos a resolver el último de los ejercicios anteriores.

En el momento en que fue contratada como mesera en el Grumney Family Restaurant, a Beth Brigden le dijeron: "Puedes ganar en promedio más de $80 al día en propinas." Suponga que la desviación estándar de la distribución de población es de $3.24. Los primeros 35 días de trabajar en el restaurante, la suma media de sus propinas fue de $84,85. Con el nivel de significancia de 0,01, ¿la señorita Brigden puede concluir que gana un promedio de más de $80 en propinas?".

Paso 1: Hipótesis nula y alternativa

La palabra clave del enunciado es más que, por lo tanto, hay que aplicar la fila 1 de la tabla y colocar en H1 el símbolo >. Las hipótesis quedarían así:

H0: U £ 80.
H1: U > 80 (el operador apunta al lado derecho).

Paso 2: Nivel de significancia (error permitido)

En este caso se trata de una prueba de una sola cola, para ellos hay que observar el operador relacional de H1, este apunta a la dirección de la curva en la que se acumula el error; es decir el error estará en el lado derecho y tiene un valor de 0,01.

Paso 3: Estadístico de prueba

En el ejercicio se conoce la desviación estándar de la población, por esa razón se debe utilizar el estadístico z.

Paso 4: Regla de decisión

Aquí es necesario establecer el valor crítico; es decir el valor de Z obtenido de la tabla. El nivel de significancia corresponde a 0,01 y la H1 indica que el error se localiza en la parte derecha de la curva. Lo que se debe hacer es buscar en la tabla de distribución normal el valor z que corresponde al 0,49(49%) y colocar ese valor en la gráfica como valor crítico. El valor Z es 2,33 y la regla de decisión sería la siguiente:

Si el valor de Z que se calcula es mayor a 2,33 la H0 se rechaza, en caso contrario no se rechaza.

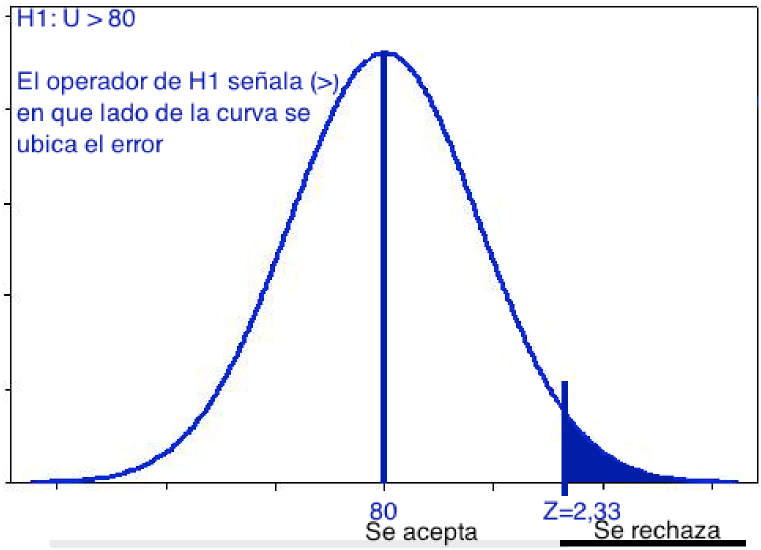

Figura 24. Curva normal con valor Z para 0.99 de nivel de significancia.
Fuente: XXXX.

Paso 5: Calcular Z y decidir

Z = (84,25 – 80) / (3,24 / √35)

Z = 4,25 / (3,24 / 5,91)

Z = 4,25 / 0,54

Z = 7,71

Como podemos ver el valor 7,71 está mucho más a la derecha de 2,33; esto significa que el valor cae en el área de rechazo, por lo tanto, se procede a rechazar la H0 y se concluye que la media de dinero de propinas diarios que se recibe no es ni menor ni igual a 80.

9.4. Verificación de hipótesis cuando NO se conoce la desviación estándar de la población

Vamos ahora a resolver un caso en el que ya no se utiliza el estadístico Z, sino el estadístico t. Esto se debe a que se desconoce la desviación estándar de la población.

La administración de White Industries analiza una nueva técnica para armar un carro de golf; la técnica actual requiere 42.3 minutos de trabajo en promedio. El tiempo medio de montaje de una muestra aleatoria de 24 carros, con la nueva técnica, fue de 40.6 minutos, y la desviación estándar, de 2.7 minutos. Con un nivel de significancia de 0.10, ¿puede concluir que el tiempo de montaje con la nueva técnica es más breve?

Paso 1: Planteamiento de hipótesis

H0: U ≥ 42,3

H1: U < 42,3

Paso 2: Nivel de significancia

0,10

Paso 3: Estadístico de prueba

No se conoce la desviación estándar de la población por ese motivo vamos a utilizar t.

Paso 4: Regla de decisión

Ubicamos en la gráfica el error y el valor crítico (que equivale a t de la tabla)

El valor t de la tabla se busca considerando que se tiene una sola cola con nivel de significancia 0,10 y 23 grados de libertad (GL = n-1).

El valor t de la tabla es 1,319

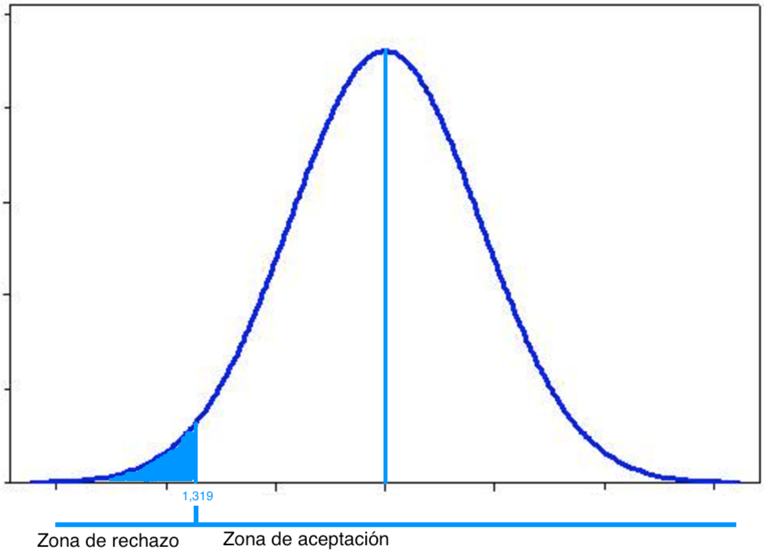

Figura 25. Curva normal con valor t para 23 grados de libertad y 0.9 nivel de significancia. *Fuente*: XXXX.

La regla de decisión por lo tanto nos señala que si el valor de Z que se calcule es menor (está más hacia la izquierda) que -1,319 la H0 se rechazará; y, por el contrario, si el valor de Z que se calcule, se encuentra a la derecha de -1,319, entonces se aceptará H0.

Paso 5: Calcular Z y decidir

t = (40,6 – 42,3) / (2,7 / √24)

t = -1,7 / (2,7 / 4,89)

t = -1,7 / 0,552

t = - 3,079

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos

El valor de t calculado es de -3,079 y este está ubicado en la zona de rechazo según la gráfica, por este motivo se rechaza H0 y se concluye que la media de la población no es mayor y tampoco es igual a 42,3.

Ir al contenido

Índice
Primer bimestre
Segundo bimestre
Solucionario
Referencias bibliográficas
Recursos