



UTPL
La Universidad Católica de Loja

Modalidad Abierta y a Distancia

Estadística

Guía didáctica

Índice

Primer
bimestre

Segundo
bimestre

Solucionario

Referencias
bibliográficas



Departamento de Economía

Sección departamental de Métodos Cuantitativos

Estadística

Guía didáctica

Autora:

Luz María Castro Quezada



ESTA_1044

Asesoría virtual
www.utpl.edu.ec

Índice

Primer
bimestre

Segundo
bimestre

Solucionario

Referencias
bibliográficas

Índice

Primer
bimestre

Segundo
bimestre

Solucionario


Referencias
bibliográficas

Estadística

Guía didáctica

Luz María Castro Quezada

Universidad Técnica Particular de Loja

 4.0, CC BY-NY-SA

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 - 978-9942-25-610-2



La versión digital ha sido acreditada bajo la licencia Creative Commons 4.0, CC BY-NY-SA: Reconocimiento-No comercial-Compartir igual; la cual permite: copiar, distribuir y comunicar públicamente la obra, mientras se reconozca la autoría original, no se utilice con fines comerciales y se permiten obras derivadas, siempre que mantenga la misma licencia al ser divulgada. <https://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>

14 abril, 2020

Índice

1. Datos de información.....	7
1.1. Presentación de la asignatura	7
1.2. Competencias genéricas de la UTPL.....	7
1.3. Competencias específicas de la carrera	8
1.4. Problemática que aborda la asignatura	8
2. Metodología de aprendizaje.....	9
3. Orientaciones didácticas por resultados de aprendizaje	10
Primer bimestre.....	10
Resultado de aprendizaje 1	10
Contenidos, recursos y actividades de aprendizaje.....	10
Semana 1	10
Unidad 1. Fundamentos de la Teoría Estadística	11
1.1. Definiciones.....	11
1.2. Tipos de variables.....	12
1.3. Niveles de medición.....	13
Actividades de aprendizaje recomendadas	14
Autoevaluación 1	15
Semana 2	18
Unidad 2. Recolección, organización y presentación de información estadística	18
2.1. Elaboración de tablas de distribución de frecuencia de datos.....	18
2.2. Distribución de frecuencias de variables cuantitativas..	20
2.3. Representación gráfica de variables cuantitativas	20
Semana 3	22

Unidad 3. Medidas de tendencia central.....	22
3.1. Medidas de tendencia central en datos no agrupados ..	22
Actividades de aprendizaje recomendadas	23
Autoevaluación 2	24
Semana 4	27
3.2. Medidas de tendencia central en datos agrupados	27
Semana 5	28
Unidad 4. Medidas de dispersión, de posición y de forma	28
4.1. Medidas de dispersión.....	28
Semana 6	30
4.2. Medidas de ubicación.....	30
Semana 7	31
4.3. Medidas de forma	32
Actividades finales del bimestre	33
Semana 8	33
Segundo bimestre	34
Resultado de aprendizaje 2	34
Contenidos, recursos y actividades de aprendizaje.....	34
Semana 9	34
Unidad 5. Distribuciones de probabilidades	34
5.1. Conceptos generales	35
5.2. Reglas de conteo	36

Índice

Primer
bimestre

Segundo
bimestre

Solucionario

Referencias
bibliográficas

Semana 10	38
5.3. Distribución binomial: Variable aleatoria discreta	38
Semana 11	39
5.4. Distribución de probabilidad normal: Variables continuas	40
5.5. Aproximación de la distribución normal a la binomial ...	42
Actividades de aprendizaje recomendadas	44
Autoevaluación 3	45
Semana 12	48
Unidad 6. Prueba de hipótesis	48
6.1. Prueba de hipótesis	48
Actividades de aprendizaje recomendadas	49
Autoevaluación 4	50
Semana 13	54
Unidad 7. Muestreo	54
7.1. Tipos de muestreo	54
Semana 14	56
7.2. Distribuciones muestrales	56
Semana 15	60
7.3. Estimación por intervalo.....	61
Actividades finales del bimestre	65
Semana 16	65
4. Solucionario	66
5. Referencias bibliográficas	70

Índice

Primer
bimestre

Segundo
bimestre

Solucionario

Referencias
bibliográficas

[Índice](#)[Primer bimestre](#)[Segundo bimestre](#)[Solucionario](#)[Referencias bibliográficas](#)

1. Datos de información

1.1. Presentación de la asignatura



1.2. Competencias genéricas de la UTPL

- Vivencia de los valores universales del humanismo de Cristo
- Comunicación oral y escrita
- Orientación a la innovación y a la investigación
- Pensamiento crítico y reflexivo
- Trabajo en equipo
- Comunicación en inglés
- Compromiso e implicación social
- Comportamiento ético
- Organización y planificación del tiempo

1.3. Competencias específicas de la carrera

Desarrolla las capacidades científicas, teóricas, y técnico – instrumentales en los campos de gestión del territorio, gestión turística, patrimonio natural y cultural, con la finalidad de responder a las necesidades y requerimientos del sector turístico, fortaleciendo y recuperando los saberes ancestrales enmarcados en las costumbres, tradiciones, historia, folklor y gastronomía.

1.4. Problemática que aborda la asignatura

Partiendo del objeto de estudio del turismo que se fundamenta en la actividad turística, la cual pretende cambiar de una estructura de turismo convencional a un turismo sostenible, en el que se gestiona la articulación pública, privada y la intervención importante de la sociedad para brindar verdaderos destinos sostenibles, permitirá un desarrollo espacial armonizado del territorio, utilizando los recursos existentes en el país, por medio de mecanismos participativos y democráticos para lograr un beneficio común entre los actores involucrados.



2. Metodología de aprendizaje

La metodología que vamos a utilizar para alcanzar este resultado de aprendizaje es el Aprendizaje por descubrimiento, el cual es un tipo de aprendizaje activo, en el que la persona en lugar de aprender los contenidos de forma pasiva, toma una actitud activa, esto implica que descubre, relaciona y reordena los conceptos para adaptarlos a su esquema cognitivo. En este tipo de aprendizaje, son los profesores o educadores quienes proponen un tema o problema y los alumnos deciden cómo abordarlo. Uno de los beneficios es que estimula a los alumnos para pensar por sí mismos, plantear hipótesis y tratar de confirmarlas de una forma sistemática.

[Índice](#)[Primer bimestre](#)[Segundo bimestre](#)[Solucionario](#)[Referencias bibliográficas](#)



3. Orientaciones didácticas por resultados de aprendizaje



Primer bimestre

Resultado de aprendizaje 1

Identifica la importancia del uso de técnicas estadísticas en el tratamiento de la información

Contenidos, recursos y actividades de aprendizaje

Con este resultado usted tendrá una visión amplia de como coleccionar información, organizarla y presentarla mediante la aplicación de varias técnicas estadísticas, por lo tanto, tendrá la capacidad de discriminar la información para aplicarla posteriormente a sus investigaciones en el área turística.



Semana 1

[Índice](#)[Primer bimestre](#)[Segundo bimestre](#)[Solucionario](#)[Referencias bibliográficas](#)

[Índice](#)[Primer bimestre](#)[Segundo bimestre](#)[Solucionario](#)[Referencias bibliográficas](#)

Unidad 1. Fundamentos de la Teoría Estadística

En esta semana vamos a analizar los fundamentos básicos de la Estadística. La teoría que abordaremos en esta semana les permitirá comprender los tipos de variables con los que se trabaja en el campo de la estadística aplicada a las ciencias sociales y administrativas y como se puede utilizar para analizar problemas relacionados a su campo de formación profesional. Para facilitar su comprensión les recomiendo hacer uso de los siguientes recursos.

1.1. Definiciones

Las definiciones más importantes para avanzar en el estudio de la asignatura son las siguientes:

Estadística es la ciencia que recoge, organiza, presenta, analiza e interpreta datos con el fin de propiciar una toma de decisiones más eficaz. Existen dos tipos de estadística: descriptiva e inferencial.

La **estadística descriptiva** comprende métodos para describir las características importantes de un conjunto de datos. Dentro de este segmento de técnicas se incluyen los procedimientos para conocer la distribución de frecuencias y su representación gráfica y el cálculo de las medidas de tendencia central, de ubicación, dispersión y simetría. La **estadística inferencial** consiste en métodos que se emplean para determinar una propiedad de una población con base en la información obtenida a través de una muestra.

Población es el conjunto de individuos u objetos que tienen propiedades comunes. Les propongo un ejemplo de población y muestra. Como se visualiza en la Figura 1, la población está conformada por individuos de una región identificados por sus tipos de sombreros. Vemos que aquellos individuos con características diferentes (sombreros rojos) son excluidos de la población elegible, pues no cumplen las características requeridas. La **muestra** consiste de una porción o subconjunto de la población que se estudia y sobre la cual se colectará y procesará información que luego será presentada en estudios de investigación o de mercado.

1.2. Tipos de variables

Una variable estadística es una característica que puede fluctuar y cuya variación es susceptible de adoptar diferentes valores, los cuales pueden medirse u observarse.

Tipos de variables:

- **Variable cualitativa o atributo:** Analiza características descriptivas de los objetos o individuos de estudio como la procedencia, nivel de estudios, género, estado civil filiación religiosa, características étnicas, situación de pobreza, etc.
- **Variable cuantitativa:** describe características de tipo numérico, pueden ser discretas o continuas.
 - **Variable discreta:** son aquellas que se refieren a elementos que se pueden contar, como el número de personas en una fila del cine o de autos en un peaje. Asumen el valor de números enteros ya que no pueden tomar valores fraccionarios.

- **Variable Continua:** son aquellas que se expresan en cualquier valor dentro de un intervalo, o sea pueden asumir valores decimales. Ejemplos de variables continuas son el peso de las personas, la distancia entre dos objetos, ingresos mensuales, ingesta calórica diaria, etc.

1.3. Niveles de medición

Una vez que se han introducido los tipos de variables más usados en el análisis estadístico, es importante conocer los tipos de medición existentes y en qué casos aplican. Existen cuatro niveles de medición:

- Nominal
- Ordinal
- Intervalo
- Razón

Generalmente, los niveles nominal y ordinal aplican a las variables de tipo cualitativo. El nivel nominal cumple la función de describir la variable sin establecer ningún tipo de comparación entre los elementos; por ejemplo, si nos referimos a la religión, no existe una jerarquía que permita calificar si una religión es superior o inferior a otra. En el caso del nivel de medición ordinal, sí se considera un orden de las características descritas por la variable de acuerdo a una escala de valor. Un ejemplo sería la promoción de un curso con una valoración deficiente, aceptable o sobresaliente.

En cuanto a las variables cuantitativas, estas pueden adoptar niveles de intervalo y razón. Vale la pena señalar que la mayor parte de variables continuas tienen una escala de medición de razón. Para ampliar los conceptos abordados en esta guía didáctica, les recomiendo revisar el Capítulo 1: ¿Qué es la Estadística? del texto

básico de Lind et al. (2012) referentes a introducción a la estadística. En el libro se citan otros ejemplos que les permitirán comprender a cabalidad las definiciones estudiadas.

Recursos de aprendizaje

Para profundizar sobre este tema y afianzar sus conocimientos es necesario, además, realizar la lectura de la Unidad I de la *Guía Estadística* (Castro 2019), donde se enuncian las nociones básicas que usted debería conocer y dominar sobre esta asignatura.



Actividades de aprendizaje recomendadas

Estimados estudiantes, teniendo más claros los conceptos arriba descritos, les invito a desarrollar las siguientes actividades para afianzar sus conocimientos.

- **Actividad de aprendizaje:** *Clasificación de variables*

Procedimiento: *Organice un listado de variables de tipo cualitativo, de este listado identifique aquellas variables de tipo nominal y ordinal. Luego elabore un listado de variables cuantitativas, identifique si son de tipo discreto o continuo. Revise la Unidad I de la Guía Estadística (Castro 2019).*

- **Actividad de aprendizaje** *Autoevaluación 1*

Procedimiento: *En esta actividad debe contestar las preguntas al final de la Unidad 1, esta tarea tiene la finalidad de comprobar que usted está teniendo sintonía con los temas tratados.*



Autoevaluación 1

1. Estadística es:

- a. El sistema o método usado en la recolección, organización, análisis y descripción numérica de la información
- b. Comportamiento de hechos o fenómenos de grupo
- c. Reglas o principios para el análisis de fenómenos

2. La estadística descriptiva consiste en:

- a. La descripción de un conjunto de datos sin llegar a conclusiones con respecto a un grupo mayor.
- b. Interpretación de resultados.
- c. El análisis para llegar a conclusiones sobre una población.

3. Se entiende por población a:

- a. Un recuento de las unidades que tienen una característica común.
- b. Un recuento de unidades.
- c. Un conjunto de datos

4. Un ejemplo de población finita es:

- a. Células del cuerpo.
- b. Conjunto de números reales.
- c. Estudiantes del componente académico Estadística.

Índice

Primer
bimestre

Segundo
bimestre

Solucionario

Referencias
bibliográficas

5. **Un ejemplo de característica cualitativa puede ser:**
- Salario.
 - Género.
 - Distancia.
6. **Un ejemplo de variable discreta es:**
- Gastos en alimentación.
 - Número de turistas que visitan en una tienda.
 - Peso de una lata en un proceso de fabricación
7. **¿Cuál de las siguientes opciones describe una característica cuantitativa?**
- Nivel de formación.
 - Género.
 - Salario.
8. **¿Cuál de las siguientes opciones describe una variable nominal?**
- Profesión de los entrevistados.
 - Posición de llegada en una competencia.
 - Velocidad a la que circula un vehículo
9. **Escoja cuál de las siguientes opciones describe una característica cuantitativa de razón:**
- Procedencia de los turistas.
 - Temperatura en grados centígrados.
 - Velocidad a la que viaja un vehículo.

- 10. Escoja cuál de las siguientes opciones describe una característica cuantitativa de intervalo:**
- a. Temperatura en grados centígrados.
 - b. Género de las personas.
 - c. Número de vehículos que ingresan a un complejo turístico.

[Ir al solucionario](#)

[Índice](#)

[Primer
bimestre](#)

[Segundo
bimestre](#)

[Solucionario](#)

[Referencias
bibliográficas](#)



Semana 2



Unidad 2. Recolección, organización y presentación de información estadística

Durante esta semana vamos a conocer y aplicar algunas técnicas para la recolección, organización y presentación de datos. La organización de datos depende del tipo de variables que estamos utilizando para nuestro estudio. Como habíamos visto anteriormente las variables pueden ser cualitativas o cuantitativas. En el caso de las variables cualitativas se contabilizan las veces que la variable se presenta. En el caso de las variables cuantitativas tenemos las de tipo discreto y continuo, las variables discretas se organizan en forma similar a las cualitativas, en el caso de las variables continuas el procedimiento recomendable es organizar los datos en clases o intervalos.

2.1. Elaboración de tablas de distribución de frecuencia de datos

Estimados profesionales en formación, la organización y procesamiento de datos es un paso muy importante para los posteriores análisis estadísticos. La organización depende del tipo

[Índice](#)[Primer bimestre](#)[Segundo bimestre](#)[Solucionario](#)[Referencias bibliográficas](#)

de variables que se utiliza, tal como se había explicado en la unidad anterior. En esta unidad analizaremos como recoger, organizar y presentar los datos recogidos en una investigación.

Las fuentes de información pueden ser **primarias y secundarias**. Las fuentes primarias son aquellas que generan datos a partir del levantamiento de información a través de medios como encuestas y entrevistas. La información secundaria es aquella que encontramos en bases de datos o fuentes de consulta y que ha sido generada por otra persona o grupos de personas.

Una vez que hemos levantado la información, requerimos procesarla. Primeramente, analizaremos el tratamiento que se da a datos cualitativos. En el ejemplo que les propongo, los estudiantes de Gastronomía levantaron una encuesta para conocer las preferencias de restaurantes en la ciudad de Loja. En este caso, la organización de los datos consiste en agrupar la información de acuerdo con las categorías propuestas y contabilizar la frecuencia con la que se presentaron.

Tabla 1. Tipos de restaurantes preferidos en una encuesta a 100 personas en la ciudad de Loja

Tipo de restaurante	# de personas
Italiano	15
Mexicano	12
Asaderos	29
Comida rápida	30
Comida típica	14

Fuente: Encuesta realizada por estudiantes de Estadística II de Gastronomía-UTPL (2017)

La Tabla 1 nos presenta en forma organizada la información, así corresponden a la categoría comida rápida el mayor número de frecuencias, en este caso 30, por lo que conclusión es que es el

tipo de restaurante preferido en la ciudad. En cambio, el tipo de restaurante que la gente prefiere en menor medida es el de comida mexicana con 12 individuos.

2.2. Distribución de frecuencias de variables cuantitativas

Cuando se trata de variables cuantitativas, los datos se deben organizar en tablas de distribución de frecuencias. En esta sección se muestra el procedimiento que deben seguir para tabular y procesar sus datos. Para ello tomaremos el siguiente ejemplo como referencia en la Tabla 2 de la guía didáctica de Estadística, en la que se indica el número de instituciones educativas en 25 provincias del Ecuador. Para conocer el procedimiento completo deben ir a la *Guía de Estadística* (Castro et al. 2019), en donde se indica paso a paso el procedimiento a seguir para elaborar una tabla de distribución de frecuencias.

2.3. Representación gráfica de variables cuantitativas

a. Histogramas de frecuencias

Las variables cuantitativas se representan mediante histogramas. En el histograma de frecuencia se representa la marca de clase en el eje de las abscisas y la frecuencia absoluta en el eje de las ordenadas. Como se puede observar, en el gráfico no existe separación entre clases, esto se realiza con la finalidad de representar variables de tipo continuo, esto lo diferencia de las variables de tipo cualitativo, que se representaban en barras.

Recursos de aprendizaje

Para facilitar su comprensión les recomiendo hacer uso de los siguientes recursos. Vamos a empezar con un video que explica

cómo organizar datos en tablas de frecuencias. En el primer video se muestran la explicación cuando trabajamos con variables cualitativas, en el segundo para variables cuantitativas de tipo discreto y en el tercero para variables cuantitativas continuas.

VIDEO: Tuto Mate (2016) Elaboración de una tabla de frecuencias (variable discreta) (video). España. Recuperado de [Elaboración de una tabla de frecuencias \(variable discreta\)](#)

VIDEO: Tuto Mate (2016) Elaboración de una tabla de frecuencias (variable continua) (video). España. Recuperado de [Elaboración de una tabla de frecuencias \(variable continua\)](#)

Luego de mirar los recursos recomendados, usted estará en capacidad de identificar tipos de variables, organizarlas de acuerdo a sus características. La mayoría de estudiantes tiene problemas al diferenciar las variables discretas y las continuas, ya que las dos son numéricas, sin embargo, es importante considerar que las discretas son aquellas cuyos elementos se pueden enumerar, como el número turistas que visitan Loja durante el festival de las artes vivas o el número de obras de teatro que se exhiben (por ejemplo 10,000 turistas). Las variables continuas en cambio son aquellas que pueden adquirir valores en un rango, como la cantidad de dinero que gasta un turista en alimentación o en movilidad (por ejemplo, doscientos dólares americanos en alimentación USD\$200). Si necesita ampliar la información sobre la organización y presentación de datos puede revisar la *guía de Estadística* (Castro 2019), en la Unidad 2 se aborda esta temática.



Semana 3



Unidad 3. Medidas de tendencia central

Cuando se trabaja en estadística con datos colectados en un experimento o estudio, generalmente disponemos de una gran cantidad de información. Con la finalidad de que la información sea sintetizada se puede organizar la información en tablas de frecuencias, como se había analizado la semana anterior, o se pueden usar todos los datos para calcular parámetros que nos permitan en forma resumida describir la información.

Durante esta semana vamos a analizar las medidas de tendencia central para datos que no estén organizados en tablas de frecuencias. Las medidas de tendencia central más importantes son media, mediana y moda. Estas son muy útiles para tener una visión general de los datos con los que estamos trabajando.

3.1. Medidas de tendencia central en datos no agrupados

Existen dos situaciones con las que tendremos que analizar cómo proceder cuando tenemos un conjunto de datos, que estos se encuentren organizados en tablas de distribución de frecuencias o que se encuentren no agrupados. En este apartado conoceremos

[Índice](#)[Primer bimestre](#)[Segundo bimestre](#)[Solucionario](#)[Referencias bibliográficas](#)

como calcular las medidas de tendencia central cuando contamos con un conjunto de datos sin agrupar.

Las medidas más importantes a considerar son:

- a. Media poblacional
- b. Media Muestral
- c. Mediana
- d. Moda

Para conocer las definiciones, las fórmulas que se aplican y los procedimientos a seguir, les invito a que revisen la *guía didáctica de Estadística* (Castro 2019), en la Unidad 3 se aborda esta temática y se ilustra con ejemplos que ahí se muestran.



Actividades de aprendizaje recomendadas

- **Actividad de aprendizaje:** *Autoevaluación 2*
- **Procedimiento:** En esta actividad debe contestar las preguntas al final de la Unidad 3, esta tarea tiene la finalidad de comprobar que usted está teniendo sintonía con los temas tratados.



Autoevaluación 2

1. Las medidas que nos permiten determinar la posición de un valor con respecto al total de las observaciones se denomina medidas de:
 - a. Dispersión.
 - b. Tendencia central.
 - c. Forma
2. La media aritmética es la medida de posición más utilizada debido a que,
 - a. Es extraordinariamente estable en el muestreo.
 - b. No es sensible a los cambios en las observaciones.
 - c. Es el valor más común en la distribución de datos
3. ¿Cuál es la media aritmética del siguiente conjunto de datos 16, 17, 18, 19, 20?
 - a. 16
 - b. 18
 - c. 21
4. ¿Cuál el peso promedio del siguiente grupo de individuos? Carlos 76 kg, Rosario 62 kg, Esteban 72 kg, Mónica 44 kg
 - a. 72,1 kg
 - b. 63,5 kg
 - c. 68,0 kg

Índice

Primer
bimestre

Segundo
bimestre

Solucionario

Referencias
bibliográficas

5. Considerando la siguiente información, determinar el promedio ponderado del número de hijos por familia. Número de hijos (frecuencia): 5(3), 4(4), 6(2), 3(7).
- 3
 - 4
 - 5
6. ¿Cuáles son las desviaciones con respecto al promedio, de los siguientes ingresos mensuales: 480, 460, 340, 340, 800?
- 10; -5; 6, 40; 200
 - 58; -32; 0; 56; 180
 - 4; -24; -144; -144; 316
7. Indique, ¿cuál de las siguientes opciones es una característica de la media?
- La suma de las desviaciones respecto a la media, siempre serán iguales a cero.
 - La media aritmética de una variable por una constante, es igual a la constante.
 - La media aritmética de una constante es cero
8. ¿Cuáles son las desventajas del uso de la mediana?
- No es necesario ordenar los datos para calcularla.
 - Cuando se trabajan con diferentes grupos no se puede obtener una mediana para el conjunto.
 - Corresponde al valor más frecuente de la distribución.

9. El número de niños atendidos en los principales hospitales de un territorio es 50, 100, 74, 61, 112, 100 y 85. La mediana de esta variable es,
- a. 85
 - b. 100
 - c. 61
10. Cuando una distribución de datos n es par, la mediana corresponde a:
- a. El valor ubicado en el centro de la distribución
 - b. El promedio de los dos valores ubicados en el centro de la distribución
 - c. El valor máximo de la distribución.

[Ir al solucionario](#)

[Índice](#)

[Primer
bimestre](#)

[Segundo
bimestre](#)

[Solucionario](#)

[Referencias
bibliográficas](#)



Semana 4

En esta semana vamos a analizar las medidas de tendencia central para datos organizados en tablas de frecuencias. Algunos elementos que deben recordar de la Unidad II, son el cálculo de la frecuencia absoluta y la marca de clase, ya que serán muy útiles en este apartado para el cálculo de la media.

3.2. Medidas de tendencia central en datos agrupados

Cuando los datos han sido organizados en tablas de distribución de frecuencias como se estudió en la Unidad 2, también es posible calcular las medidas de tendencia central. De igual forma se cuenta con la media, mediana y moda. Para conocer las fórmulas que se aplican y los procedimientos a seguir, les invito a que revisen la *guía didáctica de Estadística* (Castro 2019), en la Unidad 3 se aborda esta temática y se ilustra con ejemplos que ahí se muestran.

[Índice](#)[Primer bimestre](#)[Segundo bimestre](#)[Solucionario](#)[Referencias bibliográficas](#)



Semana 5

[Índice](#)[Primer bimestre](#)[Segundo bimestre](#)[Solucionario](#)[Referencias bibliográficas](#)

Unidad 4. Medidas de dispersión, de posición y de forma

Estimados estudiantes, en esta sección vamos a conocer las medidas de dispersión de una distribución de datos, como calcularlas e interpretarlas. En esta semana vamos a aplicar herramientas para analizar datos. Las medidas de dispersión que constan dentro de este paquete nos permiten analizar el grado de variabilidad que presentan los datos, vamos a abordar el cálculo en datos agrupados y no agrupados. Las medidas más ampliamente usadas son la varianza, la desviación estándar y el coeficiente de variabilidad o variación. Para que se facilite la comprensión del mismo les recomiendo la revisión de la literatura básica de Lind et al. (2015) en el Capítulo 3 Descripción de datos: medidas numéricas; y la complementaria de Anderson et al. (2016) Capítulo 3 Estadística descriptiva: medidas numéricas.

4.1. Medidas de dispersión

a. Rango

Es la medida más simple de dispersión, representa la diferencia entre los valores máximo y mínimo de un conjunto de datos.

b. Varianza

La varianza es una medida de dispersión definida como la esperanza del cuadrado de la desviación de dicha variable respecto a su media. O en pocas palabras, es la media de los residuos al cuadrado. Su unidad de medida corresponde al cuadrado de la unidad de medida de la variable: por ejemplo, si la variable mide una distancia en metros, la varianza se expresa en metros al cuadrado.

c. Desviación estándar de la población

La desviación estándar es la raíz cuadrada de la varianza de la población.

Para conocer las fórmulas y los procedimientos para calcular las medidas de dispersión en datos agrupados y no agrupados pueden revisar la *guía didáctica Estadística* (Castro 2016), en la unidad 4 se detalla cómo llevar a cabo estas metodologías.

Recursos de aprendizaje

Para ilustrar como calcular las medidas antes descritas, vamos a observar el siguiente video:

VIDEO: Montero Espinoza (2018) Medidas de dispersión con intervalos (video). Recuperado de [MEDIDAS DE DISPERSIÓN CON INTERVALOS](#)

Los videos anteriores muestran como calcular la varianza y la desviación estándar, considerando dos escenarios, que nuestros datos estén agrupados en tablas de frecuencias o cuando tenemos datos que no hayan sido organizados. Si desea ampliar la información sobre el cálculo de medidas de dispersión en datos agrupados y no agrupados puede revisar la *guía didáctica de Estadística* (Castro et al. 2019), en la Unidad 4 se aborda esta temática y se muestran ejemplos.



Semana 6

En esta semana vamos a aplicar herramientas estadísticas para el análisis de datos. Las medidas de posición permiten dividir una distribución de datos en partes iguales, para análisis posteriores. Las medidas más ampliamente usadas son percentiles, deciles y cuartiles.

4.2. Medidas de ubicación

Las medidas de localización dividen la distribución en partes iguales, sirven para clasificar a un individuo o elemento dentro de una determinada población o muestra. Las más conocidas son los percentiles, deciles y cuartiles

a. Percentiles

Aunque los cuartiles son generalmente los elementos más usados para ubicar la posición de un valor de un conjunto de datos, en ocasiones los investigadores requieren conocer otras posiciones. Para este propósito se puede usar la fórmula general de los percentiles, los mismos que dividen un conjunto de observaciones en 100 partes iguales.

La fórmula para calcular cualquier percentil es la siguiente:

$$P_n = \frac{np}{100}$$

[Índice](#)[Primer bimestre](#)[Segundo bimestre](#)[Solucionario](#)[Referencias bibliográficas](#)

Donde

P : Es el percentil que buscamos y se reemplaza en la fórmula.

n : número de observaciones

b. Cuartiles

Dividen a un conjunto de observaciones en cuatro partes iguales. Un primer paso consiste en ordenar los datos de menor a mayor. Luego procedemos a ubicar las posiciones que representan a los cuartiles 1, 2 y 3.

Si desea ampliar la información sobre el cálculo de medidas de posición en datos agrupados y no agrupados puede revisar la *guía didáctica de Estadística* (Castro et al. 2019), en la Unidad 4 se aborda esta temática.



Semana 7

Durante la presente semana vamos a analizar las medidas de forma, siendo la más importante la Simetría. Este tipo de medidas nos permiten analizar si la distribución de datos con la que trabajamos tiene una forma simétrica, esto ocurre cuando los datos se distribuyen uniformemente en las distintas clases y tiene forma de Campana de Gauss; o si tiene una forma asimétrica, en la cual la mayor parte de los datos se ubican en las clases menores o mayores.

4.3. Medidas de forma

Las medidas de forma son indicadores estadísticos que permiten identificar si una distribución de frecuencia presenta uniformidad. Son necesarias para determinar el comportamiento de los datos y así, poder adaptar herramientas para el análisis probabilístico.

a. Sesgo

Muestra la falta de simetría en un grupo de observaciones, se refiere a la forma de un conjunto de datos y se distinguen tres formas:

- Simétrica: la media y la mediana son iguales, y los valores de datos se dispersan uniformemente en torno a estos valores.
- Sesgo positivo o sesgo a la derecha: son mucho más comunes; existe un solo pico y los valores se extienden mucho a la derecha que a la izquierda de este. En este caso la media es más grande que la mediana.
- Sesgo negativo: existe un solo pico y las observaciones se extienden más a la izquierda, en dirección negativa.

b. Regla empírica

En cualquier distribución de frecuencias simétrica con forma de campana, aproximadamente 68% de las observaciones se encontrarán entre más y menos una desviación estándar de la media; cerca de 95% de las observaciones se encontrarán entre más y menos dos desviaciones estándares de la media y, aproximadamente todas las observaciones (99.7%), estarán entre más y menos tres desviaciones estándares de la media.

Recursos de aprendizaje

Para que conozcan más sobre este tipo de medidas pueden visitar el siguiente sitio:

VIDEO: Montero Espinoza (2018) Coeficiente de simetría (video).
Recuperado de [COEFICIENTE DE SIMETRÍA](#)

El video muestra como calcular el coeficiente de simetría y la forma de interpretarlo. Además, se observa cómo se ven los datos en forma gráfica, como campana de Gauss si es simétrica o con asimetría izquierda o derecha. Si desea ampliar la información sobre el cálculo de medidas de posición en datos agrupados y no agrupados puede revisar la guía Estadística (Castro 2019), en la Unidad 4 se aborda esta temática.



Actividades finales del bimestre



Semana 8

¡Estamos finalizando el bimestre!

Durante esta última semana les recomiendo realizar una revisión general de lo estudiado en el bimestre y revisar las autoevaluaciones una vez más. Esta información será de gran utilidad para preparar su evaluación presencial.

¡Éxitos!



Segundo bimestre

Resultado de aprendizaje 2

Analiza las características de un conjunto de datos a través del cálculo de sus indicadores.

Contenidos, recursos y actividades de aprendizaje



Semana 9



Unidad 5. Distribuciones de probabilidades

Estimados estudiantes, en este bimestre vamos a abordar los contenidos referentes a estadística inferencial. Los temas a tratar presentan mayor complejidad que los que abordamos en el primer bimestre. Sin embargo, estoy convencida que con su trabajo organizado y la colaboración que tendremos durante el bimestre, vamos a tener éxito en el proceso enseñanza-aprendizaje. ¡Vamos a iniciar positivamente estas unidades!

[Índice](#)[Primer bimestre](#)[Segundo bimestre](#)[Solucionario](#)[Referencias bibliográficas](#)

5.1. Conceptos generales

En esta sección abordaremos los conceptos más importantes que deben manejar en el campo de las distribuciones de probabilidad.

Probabilidad: valor entre cero y uno que describe la posibilidad relativa que ocurra un evento. La fórmula clásica para calcular la probabilidad de un evento es la siguiente:

$$Probabilidad = \frac{\text{Número de resultados favorables}}{\text{Número total de resultados posibles}}$$

Vamos a ver un ejemplo ¿Cuál es la probabilidad de que al lanzar un dado salga el número 2?

Los resultados posibles del lanzamiento de un dado son 1, 2, 3, 4, 5, y 6. El caso favorable es tan sólo uno (que salga el dos), mientras que los casos posibles (n) son seis.

Entonces se tendría que:

$$Probabilidad = \frac{1}{6} = 0,16$$

La probabilidad de obtener un 2 en el lanzamiento de un dado es 0,16 o del 16% si multiplicamos el valor obtenido por 100.

Distribución de probabilidad: Son todos los valores posibles que resultan de un experimento aleatorio, junto con la probabilidad asignada a cada valor.

Es importante también conocer el concepto de variable aleatoria, que pueden ser de tipo discreta y continua.

Variable aleatoria discreta: se considera así cuando los valores que asume se pueden contar y pueden organizarse en una secuencia al igual que los números enteros positivos; por tanto, solo podrá asumir un valor finito de valores. Ejemplo: número de turistas que visitan la playa en temporada.

Variable aleatoria continua: se considera así cuando puede asumir cualquier valor dentro de un intervalo. Se pueden considerar el resultado de mediciones de longitud, tiempo ya que estas variables admiten fracciones. Ejemplo: valor que gastan los turistas en alimentación (dólares)

Espacio muestral: son todos los posibles resultados que pueden producirse de un experimento.

Por ejemplo, si lanzamos un dado por una ocasión los resultados posibles serían:

$$E = \{1, 2, 3, 4, 5, 6\}$$

Si lo lanzamos en dos ocasiones, o en pasos múltiples, los resultados posibles serían:

$$E = \left\{ \begin{array}{l} 1, 1 \quad 1, 2 \quad 1, 3 \quad 1, 4 \quad 1, 5 \quad 1, 6 \\ 2, 1 \quad 2, 2 \quad 2, 3 \quad 2, 4 \quad 2, 5 \quad 2, 6 \\ 3, 1 \quad 3, 2 \quad 3, 3 \quad 3, 4 \quad 3, 5 \quad 3, 6 \\ 4, 1 \quad 4, 2 \quad 4, 3 \quad 4, 4 \quad 4, 5 \quad 4, 6 \\ 5, 1 \quad 5, 2 \quad 5, 3 \quad 5, 4 \quad 5, 5 \quad 5, 6 \\ 6, 1 \quad 6, 2 \quad 6, 3 \quad 6, 4 \quad 6, 5 \quad 6, 6 \end{array} \right\}$$

5.2. Reglas de conteo

La regla de conteo para experimentos de varias etapas permite determinar el número de resultados experimentales sin alistarlos.

a. Experimentos de pasos múltiples

La regla de conteo para experimentos de pasos múltiples permite determinar el número de resultados experimentales sin tener que enumerarlos.

Para conocer el número de elementos que componen un espacio muestral cuando se cuenta con experimentos con pasos múltiples k , se recomienda considerar el número de resultados posibles en el primer paso n_1 , resultados posibles en el segundo paso n_2 y así en lo sucesivo, entonces el número total de resultados experimentales serán $(n_1) \times (n_2) \dots \times (n_k)$.

En el caso del lanzamiento del dado en el primer lanzamiento se tienen 6 resultados posibles, y en el segundo lanzamiento también se tienen 6 resultados posibles, para determinar el número de elementos que integran el espacio muestral se recomienda multiplicar los resultados posibles de n_1 por el número de resultados posibles de n_2 . Así tendríamos $6 \times 6 = 36$, lo que implica que 36 elementos conformarían el espacio muestral. Para visualizar la conformación del espacio muestral en forma gráfica, vamos a revisar el Capítulo 5 Estudio de los conceptos de la probabilidad del texto básico de Lind et al. (2015).

b. Combinaciones

La regla de conteo para combinaciones permite contar el número de resultados experimentales cuando el experimento consiste en seleccionar n objetos de un conjunto (usualmente mayor) de N objetos. Para conocer el procedimiento para calcular combinaciones ir a la *Guía de Estadística* (Castro 2019), donde se detalla la fórmula correspondiente y se ilustra con ejemplos.

c. Permutaciones

Esta regla permite calcular el número de resultados experimentales cuando se seleccionan n objetos de un conjunto de N objetos

y el orden de selección es relevante. Los mismos objetos n seleccionados en orden diferente se consideran un resultado experimental diferente. De forma similar para conocer el procedimiento para calcular permutaciones ir a la *Guía de Estadística* (Castro 2019), donde se detalla la fórmula correspondiente y se ilustra con ejemplos.



Semana 10

El estudio de la distribución de probabilidad binomial se aplica a variables cuantitativas discretas. Recuerde que en la primera semana revisamos a detalle este tipo de variables. Para que conozca a detalle los procedimientos a seguir para el cálculo de este tipo de probabilidad puede revisar los siguientes videos que he seleccionado.

5.3. Distribución binomial: Variable aleatoria discreta

La distribución de probabilidad binomial es una distribución de probabilidad que tiene muchas aplicaciones. Está relacionada con un experimento de pasos múltiples al que se le llama experimento binomial. Un experimento binomial tiene las cuatro propiedades siguientes.

1. El experimento consiste en una serie de n ensayos idénticos.
2. En cada ensayo hay dos resultados posibles. A uno de estos resultados se le llama éxito y al otro se le llama fracaso.
3. La probabilidad de éxito, que se denota p , no cambia de un ensayo a otro. Por ende, la probabilidad de fracaso, que se denota $q = (1 - p)$, tampoco cambia de un ensayo a otro.

[Índice](#)[Primer bimestre](#)[Segundo bimestre](#)[Solucionario](#)[Referencias bibliográficas](#)

4. *Los ensayos son independientes.*

En un experimento binomial lo que interesa es el número de éxitos en n ensayos. Si x denota el número de éxitos en n ensayos, es claro que x tomará los valores 0, 1, 2, 3, ... n . Dado que el número de estos valores es finito, x es una variable aleatoria discreta. A la distribución de probabilidad correspondiente a esta variable aleatoria se le llama distribución de probabilidad binomial. Para conocer el procedimiento a seguir revise la *Guía de Estadística* Castro (2019).

Recursos de aprendizaje

Para apoyar su comprensión he seleccionado algunos videos en donde se muestra en forma didáctica los fundamentos de la distribución de probabilidad binomial:

VIDEO: Física y Mates (2015) Distribución Binomial (video).

Recuperado de: [Distribución Binomial | Explicación y ejercicio resuelto](#).

VIDEO: Física y Mates (2015) Distribución Binomial: media y

varianza (video). Recuperado de: [Distribución Binomial | Media y Varianza | Ejercicio resuelto](#)

Si desea ampliar la información sobre el cálculo probabilidad binomial puede revisar la *guía didáctica de Estadística*, en la Unidad 5 se aborda esta temática.



Semana 11

Para culminar los contenidos relacionados al estudio de probabilidades, vamos a abordar la distribución de probabilidad normal. Este tipo de probabilidad se aplica a variables continuas,

cuya distribución tiene una forma simétrica. Así vemos que, en este tema, es necesario que los datos cumplan con características estudiadas anteriormente como que la media, mediana y moda tengan el mismo valor. Recuerde revisar los conceptos para afianzar sus conocimientos y aplicarlos a esta unidad.

5.4. Distribución de probabilidad normal: Variables continuas

Una diferencia fundamental entre las variables aleatorias discretas y las variables aleatorias continuas es cómo se calculan las probabilidades. En las variables discretas se calcula la probabilidad de que la variable tome un valor determinado. En cambio, cuando se calculan probabilidades de variables aleatorias continuas, se calcula la probabilidad de que la variable aleatoria tome alguno de los valores dentro de un intervalo que se calcula como el área bajo la curva normal.

La distribución normal tiene gran cantidad de aplicaciones prácticas, en las cuales la variable aleatoria puede ser el peso o la estatura de las personas, puntuaciones de exámenes, resultados de mediciones científicas, precipitación pluvial u otras cantidades similares.

La distribución normal es simétrica, por tanto, su sesgo es cero. Las colas de la curva normal se extienden al infinito en ambas direcciones y en teoría jamás tocan el eje horizontal. La desviación estándar determina qué tan plana y ancha es la curva normal. Desviaciones estándar grandes corresponden a curvas más planas y más anchas, lo cual indica mayor variabilidad en los datos.

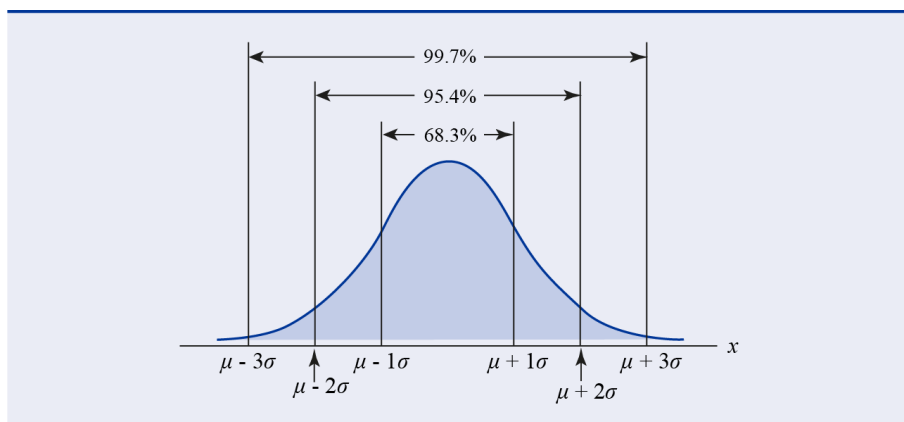


Figura 1. Curva de distribución normal

Tomado de: Anderson et al. (2016)

La figura anterior presenta el área bajo la curva normal, como se observa en la gráfica aproximadamente el 68% de los datos se ubican a 1 desviación estándar del promedio, aproximadamente el 95% se ubican a dos desviaciones estándar del promedio y aproximadamente el 99% de las observaciones a 3 desviaciones estándar del promedio. Esto se conoce como teorema central del límite que habíamos revisado al final de la Unidad 4 de esta guía didáctica.

Como ocurre con otras variables aleatorias continuas, los cálculos de la probabilidad en cualquier distribución normal se realizan calculando el área bajo la gráfica de la función de densidad de probabilidad. Por tanto, para hallar la probabilidad de que una variable aleatoria normal esté dentro de un determinado intervalo, se tiene que calcular el área que se encuentra bajo la curva normal y sobre ese intervalo.

Para la distribución normal estándar ya se encuentran calculadas las áreas bajo la curva normal y se cuenta con tablas que dan estas

áreas y que se usan para calcular las probabilidades. Estas tablas generalmente se encuentran al final de los textos de estadística.

Es importante el cálculo del valor normal estándar z , y por esta razón se indica la fórmula a continuación:

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

Donde,

z: valor normal estándar

x: Cualquier observación i

μ : Media de la distribución

σ : Desviación estándar de la distribución

Los tres casos más comunes para calcular probabilidades de distribuciones normales son: (1) la probabilidad de que la variable aleatoria normal estándar z sea menor o igual que un valor dado; (2) la probabilidad de que z esté entre dos valores dados, y (3) la probabilidad de que z sea mayor o igual que un valor dado. Para mostrar el uso de las tablas de probabilidad acumulada de la distribución normal estándar en el cálculo de estos tres tipos de probabilidades, se consideran algunos ejemplos tomados de Martínez (2012, p. 251). Ejemplos de la aplicación de distribuciones de probabilidad normal se encuentran en la guía Estadística (Castro 2019).

5.5. Aproximación de la distribución normal a la binomial

Recuerde que un experimento binomial consiste en una serie de n ensayos idénticos e independientes, habiendo para cada ensayo dos resultados posibles, éxito o fracaso. La probabilidad de éxito en un

ensayo es la misma que en cualquier otro de los ensayos y se denota **p**. La variable aleatoria binomial es el número de éxitos en **n** ensayos y lo que se quiere saber es la probabilidad de **x** éxitos en **n** ensayos. Sin embargo, se dificulta cuando el número de ensayos es muy grande. En los casos en que **$np \geq 5$ y $nq \geq 5$** , la distribución normal proporciona una aproximación a las probabilidades binomiales que es fácil de usar. Cuando se usa la aproximación normal a la binomial, es importante recordar el valor esperado y la varianza presentados en la sección 5.3 de la guía didáctica (Castro 2019), por lo tanto, les sugiero revisar las fórmulas ahí recomendadas.

Recursos de aprendizaje

Para reforzar los contenidos, revise los siguientes videos en donde se muestra como realizar los cálculos de la distribución de probabilidad normal.

VIDEO: Píldoras matemáticas (2017) Función de distribución normal (video). Recuperado de: [Función de distribución normal](#).

VIDEO: Píldoras matemáticas (2017) Como usar una tabla de distribución normal (video). Recuperado de: [Cómo usar la tabla de distribución normal](#).

Si desea ampliar la información sobre el cálculo probabilidad binomial puede revisar el texto de Lind et al (2015), en el Capítulo 7 Distribución de probabilidad normal, se detalla información referente a este tipo de distribuciones de probabilidad.



Actividades de aprendizaje recomendadas

- **Actividad de aprendizaje: Autoevaluación 3**

Procedimiento: *En esta actividad debe contestar las preguntas al final de la Unidad 5, esta tarea tiene la finalidad de comprobar que usted está teniendo sintonía con los temas tratados. Al final puede revisar el solucionario para conocer si sus respuestas fueron correctas.*

[Índice](#)[Primer bimestre](#)[Segundo bimestre](#)[Solucionario](#)[Referencias bibliográficas](#)



Autoevaluación 3

1. **Algunos modelos de probabilidad que utilizan de variables aleatorias continuas son:**
 - a. Binomial
 - b. De Poisson
 - c. Normal
2. **Identifique la función binomial en un experimento en el cual p es la probabilidad de éxito, y q la probabilidad de fracaso, entonces la probabilidad que se obtenga x en n ensayos.**
 - a. $(p)^{n+} (q)^{n-x}$
 - b. $(1) - (p)^n$
 - c. $(1) - (q)^n$
3. **La probabilidad de éxito de un acontecimiento en la distribución binomial es:**
 - a. Fija.
 - b. Variable.
 - c. Cero.
4. **Si en un ejercicio planteado como binomial el número de experimentos realizados es grande se recomienda usar la distribución:**
 - a. De Poisson.
 - b. Normal.
 - c. t de Student

Índice

Primer
bimestre

Segundo
bimestre

Solucionario

Referencias
bibliográficas

5. La probabilidad de la variable aleatoria continua se calcula considerando:
- Un evento específico x .
 - El área bajo la curva entre dos valores.
 - La ocurrencia de un evento binomial
6. La probabilidad de la variable aleatoria continua de un valor puntual $x=5$ es:
- Cero.
 - Positiva.
 - Negativa
7. El área bajo la curva normal es igual a:
- 3.
 - 1.
 - 0.
8. Una variable aleatoria x tiene una media de 5 y una desviación típica de +1. El valor z calculado para determinar la probabilidad de que la variable aleatoria sea inferior a 3, es:
- +2.
 - 2.
 - 0.
9. La probabilidad que la variable aleatoria del ejemplo anterior sea inferior a 3 es de:
- 0,4772
 - 0,0228
 - 0,5000

10. El promedio de las calificaciones en una asignatura es 12 con una desviación típica de 3. Para calcular la probabilidad de que un estudiante obtenga una calificación menor a 10 ($x < 10$), necesitamos calcular:

- a. El valor z, buscar el valor correspondiente en la tabla, y restar dicho valor a 0.5 que es el área bajo la mitad de la curva.
- b. El valor z, buscar el valor correspondiente en la tabla, esa es la probabilidad.
- c. El valor z, buscar el valor correspondiente en la tabla, y sumar dicho valor a 0.5 que es el área bajo la mitad de la curva

[Ir al solucionario](#)

[Índice](#)

[Primer bimestre](#)

[Segundo bimestre](#)

[Solucionario](#)

[Referencias bibliográficas](#)



Semana 12

El siguiente tema que vamos a revisar en esta asignatura es la prueba de hipótesis de investigación. Este tema es de gran interés en el campo de la investigación y por este motivo deben prestar mucha atención. Para reforzar los contenidos descritos en la Guía Didáctica, revise los siguientes videos.



Unidad 6. Prueba de hipótesis

6.1. Prueba de hipótesis

Las pruebas de hipótesis se basan en la diferencia entre el estadístico muestral y el parámetro poblacional. Los pasos básicos para probar una hipótesis son:

[Índice](#)[Primer bimestre](#)[Segundo bimestre](#)[Solucionario](#)[Referencias bibliográficas](#)

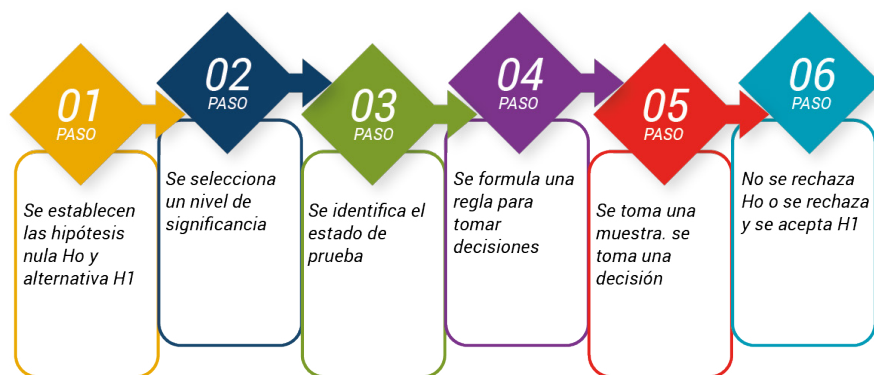


Figura 2. Pasos para la prueba de hipótesis.

Adaptado de: Lind et al. (2015)

Si desea ampliar la información sobre pruebas de hipótesis puede revisar la *guía didáctica de Estadística*, en la Unidad 7 se aborda esta temática.



Actividades de aprendizaje recomendadas

- **Actividad de aprendizaje:** Autoevaluación 4

Procedimiento: En esta actividad debe contestar las preguntas al final de la Unidad 6, esta tarea tiene la finalidad de comprobar que usted está teniendo sintonía con los temas tratados. Al final puede revisar el solucionario para conocer si sus respuestas fueron correctas.

Índice

Primer
bimestre

Segundo
bimestre

Solucionario

Referencias
bibliográficas



Autoevaluación 4

1. El primer paso en una prueba de hipótesis es:
 - a. Establecer la hipótesis nula y alternativa.
 - b. Definir el nivel de significancia.
 - c. Seleccionar el estadístico de prueba.
2. El valor de $z_{\alpha/2}$ para un nivel de confianza del 99% es
 - a. 1,96
 - b. 2,58
 - c. 1,65
3. En una muestra aleatoria simple de 60 artículos la media muestral fue 80. La desviación estándar poblacional es 15. Seleccione la fórmula con la que calcularía el intervalo de confianza.
 - a. $\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$
 - b. $\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$
 - c. $\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$

Índice

Primer
bimestre

Segundo
bimestre

Solucionario

Referencias
bibliográficas

4. El promedio de bonificaciones de fin de año por empleado en el sector público es de \$1200. Suponga que se desea tomar una muestra de los empleados para ver si la media de la bonificación es diferente de la media reportada para la población. La forma correcta de plantear la hipótesis sería:

a. $H_0: \mu \geq \mu_0$
 $H_a: \mu < \mu_0$

b. $H_0: \mu \leq \mu_0$
 $H_a: \mu > \mu_0$

c. $H_0: \mu = \mu_0$
 $H_a: \mu \neq \mu_0$

5. La definición de Error Tipo I en una prueba de hipótesis es la siguiente:

- a. Se rechaza la hipótesis nula cuando es verdadera.
- b. Se rechaza la hipótesis nula cuando la hipótesis alternativa es verdadera.
- c. Se acepta la hipótesis nula cuando es verdadera.

Índice

Primer
bimestre

Segundo
bimestre

Solucionario

Referencias
bibliográficas

6. De acuerdo al sindicato de comunicadores los ingresos siguen una distribución normal, con media de 800 y una desviación estándar poblacional de 150. En una muestra tomada a un grupo de 120 conductores su ingreso medio ascendió a 825. Si quiere determinar que el ingreso medio es mayor a la media poblacional, ¿de qué forma plantearía las hipótesis?
- a. $H_0: \mu \geq 800$
 $H_a: \mu < 800$
 - b. $H_0: \mu \leq 800$
 $H_a: \mu > 800$
 - c. $H_0: \mu = 800$
 $H_a: \mu \neq 80$
7. En el ejemplo anterior, el estadístico de prueba que elegiría debería seguir una distribución:
- a. Normal.
 - b. t de Student.
 - c. Chi cuadrado.
8. La definición de Error Tipo II en una prueba de hipótesis es la siguiente:
- a. Se rechaza la hipótesis nula cuando es verdadera.
 - b. Se rechaza la hipótesis nula cuando la hipótesis alternativa es verdadera.
 - c. Se acepta la hipótesis nula cuando es falsa.

9. **El método del valor crítico en una prueba de hipótesis de la cola superior en una variable que sigue una distribución normal, implica que, si el valor z calculado es:**
- a. Menor al valor z del nivel α seleccionado, se rechaza la hipótesis nula.
 - b. Mayor al valor z del nivel α seleccionado, se rechaza la hipótesis nula.
 - c. Mayor al valor z del nivel α seleccionado, no se rechaza la hipótesis nula.
10. **El método del valor- p en una prueba de hipótesis de la cola inferior en una variable que sigue una distribución normal, implica que, si el valor- p calculado es menor al α seleccionado:**
- a. Se rechaza la hipótesis nula.
 - b. Se rechaza la hipótesis nula.
 - c. No se rechaza la hipótesis nula

[Ir al solucionario](#)

[Índice](#)

[Primer
bimestre](#)

[Segundo
bimestre](#)

[Solucionario](#)

[Referencias
bibliográficas](#)



Semana 13

Durante las siguientes tres semanas vamos a analizar técnicas de muestreo probabilísticas y no probabilísticas. Para empezar, vamos a revisar las definiciones básicas de muestra y población, parámetros y estadísticos. Esta información será una primera aproximación al tema y les permitirá ampliar sus conocimientos sobre este tema.



Unidad 7. Muestreo

7.1. Tipos de muestreo

Existen diferentes tipos de muestreo que pueden ser utilizados para realizar investigaciones. El tipo de muestreo que se seleccione depende en gran medida de los objetivos de la investigación, del presupuesto y tiempo disponibles que tenga el investigador. A continuación, les presento un esquema de los diferentes tipos de muestreo que existen para su conocimiento y la descripción de cada uno de ellos.

[Índice](#)[Primer bimestre](#)[Segundo bimestre](#)[Solucionario](#)[Referencias bibliográficas](#)

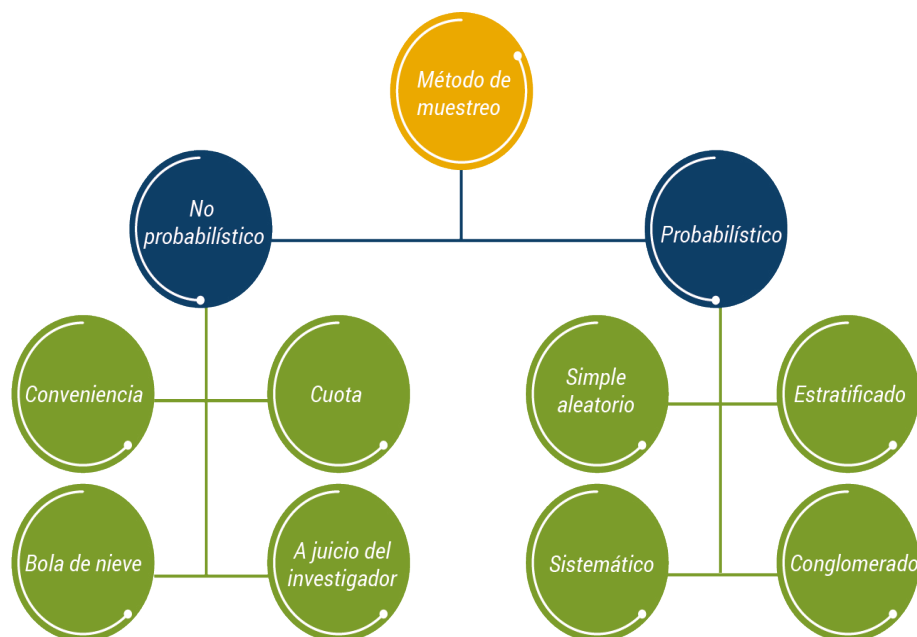


Figura 3. Tipos de muestreo

Tomado de: Lind et al. (2012)

a. Muestreo no probabilístico

Una muestra es no aleatoria cuando los elementos son elegidos por métodos dirigidos (no aleatorios). Una desventaja de este procedimiento es que como es de carácter subjetivo está sujeto a errores. La muestra dirigida está constituida por una parte de la población, que el investigador considera representativa del universo. Si se recogen características homogéneas, la representatividad de la muestra puede ser satisfactoria. Otro tipo de muestreo dirigido es el de bola de nieve, en la que el sujeto en estudio sugiere quien podría ser el próximo sujeto de muestreo.

b. Muestreo aleatorio

Una muestra es aleatoria cuando los elementos que constituyen la población o universo tienen la misma posibilidad de ser

seleccionados. Para escoger una muestra de una población hay diversos métodos que se detallan en la *guía de estadística* (Castro 2019), revíselos para que conozcan los casos en los que aplica.



Semana 14

En esta semana vamos a revisar como calcular las distribuciones muestrales de la media y la proporción. Para reforzar los contenidos, revise los siguientes videos.

7.2. Distribuciones muestrales

a. Distribución muestral de la media \bar{x}

Es la distribución de probabilidad de todos los valores de la media muestral \bar{x} . La distribución muestral de tiene un valor esperado, una desviación estándar y una forma característica.

El valor esperado de es igual a la media de la población de la que se tomó la muestra.

$$E(\bar{x}) = \mu$$

La desviación estándar de la distribución muestral de , es posible demostrar que usando el muestreo aleatorio simple, la desviación estándar depende de si la población es finita o infinita. Las dos fórmulas para la desviación estándar son las siguientes:

Población finita

$$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{n-1}} \left(\frac{\sigma}{\sqrt{n}} \right)$$

El factor $\sqrt{\frac{N-n}{n-1}}$ se requiere cuando la población es finita, pero no cuando es infinita. A este factor se le conoce como factor de corrección para una población finita.

Población infinita

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Esta fórmula también se recomienda en aquellos casos cuando la población es finita y el tamaño de la muestra sea menor o igual a 5% del tamaño de la población; es decir, $n/N \leq 0.05$, ya que en estos casos el factor de corrección será casi igual a 1.

Los resultados respecto al valor esperado y a la desviación estándar en la distribución muestral son aplicables a cualquier población. El paso final en la identificación de las características de la distribución muestral es determinar la forma de la distribución muestral. Se considerarán dos casos: a) la población tiene distribución normal y, b) la población no tiene distribución normal. Cuando la población tiene distribución normal, la distribución muestral de está distribuida normalmente, sea cual sea el tamaño de la muestra. Cuando la población de la que se tomó la muestra aleatoria simple no tiene distribución normal, el teorema del límite central ayuda a determinar la forma de la distribución muestral de . El enunciado del teorema del límite central aplicado a la distribución muestral de dice lo siguiente.

TEOREMA DEL LÍMITE CENTRAL

Cuando se seleccionan muestras aleatorias simples de tamaño n de una población, la distribución muestral de la media muestral puede aproximarse a una distribución normal a medida que el tamaño de la muestra se hace grande.

Se sugiere la lectura del Capítulo 8: Métodos de muestreo y teorema central del límite del texto básico de Lind et al. (2015) con la finalidad que ustedes puedan identificar diferentes tipos de formas de las distribuciones muestrales. De acuerdo con estos autores, la mayoría de los especialistas en estadística consideran que una muestra de 30 o mayor es lo bastante grande para aplicar el teorema del límite central.

b. Distribución muestral de \bar{p}

La proporción muestral es el estimador puntual de la proporción poblacional p . la proporción muestral es una variable aleatoria y su distribución de probabilidad se conoce como distribución muestral de \bar{p} .

La fórmula para calcular la proporción muestral es:

$$\bar{p} = \frac{x}{n}$$

Donde

x: número de elementos de la muestra que poseen la característica de interés

n: *tamaño de la muestra*

Para determinar qué tan cerca está la proporción muestral \bar{p} de la proporción poblacional p , se necesita entender las propiedades de la distribución muestral de \bar{p} : el valor esperado de \bar{p} , la desviación estándar de \bar{p} y la forma de la distribución muestral de \bar{p} .

El valor esperado de \bar{p} , la media de todos los posibles valores de \bar{p} , es igual a la proporción poblacional p .

$$E(\bar{p}) = p$$

Las dos fórmulas para calcular la desviación estándar de \bar{p} se presentan a continuación.

Población finita

$$\sigma_{\bar{p}} = \sqrt{\frac{N-n}{n-1}} \left(\frac{p(1-p)}{n} \right)$$

El factor $\sqrt{\frac{N-n}{n-1}}$ se requiere cuando la población es finita, pero no cuando es infinita. A este factor se le conoce como factor de corrección para una población finita.

Población infinita

$$\sigma_{\bar{p}} = \left(\frac{p(1-p)}{n} \right)$$

Ahora que se conoce la media y la desviación estándar de la distribución muestral de \bar{p} , el último paso es determinar la forma de la distribución muestral. La proporción muestral es $\bar{p} = \frac{x}{n}$. En una muestra aleatoria simple de una población grande, el valor de x es una variable aleatoria binomial que indica el número de los

elementos de la muestra que tienen la característica de interés. Como n es una constante, la probabilidad de x/n es la misma que la probabilidad de x , lo cual significa que la distribución muestral de p también es una distribución de probabilidad discreta.

Una distribución binomial se aproxima mediante una distribución normal siempre que el tamaño de la muestra sea lo suficientemente grande para satisfacer la siguiente condición.

$$np \geq 5$$

Entonces, la distribución de probabilidad de x en la proporción muestral, $\bar{p} = \frac{x}{n}$, puede aproximarse por medio de una distribución normal. Y como n es una constante, la distribución muestral también se aproxima mediante una distribución normal.

Si desea ampliar la información sobre el cálculo de distribuciones de la media y de la proporción poblacional puede revisar la *guía didáctica de Estadística*, en la Unidad 6 se aborda esta temática.



Semana 15

Durante la última semana que realizamos muestreo vamos a analizar el procedimiento para llevar a cabo estimaciones por intervalo. Para reforzar los contenidos, revise los siguientes videos.

7.3. Estimación por intervalo

a. De la media poblacional

El objetivo de la estimación por intervalo es aportar información de qué tan cerca se encuentra la estimación puntual, obtenida de la muestra, del valor del parámetro poblacional. En esta sección se muestra cómo obtener una estimación por intervalo para la media poblacional μ y para la proporción poblacional p . La fórmula general para obtener una estimación por intervalo para la media poblacional es

$$\bar{x} \pm \text{Margen de error}$$

En el caso de la media poblacional, la estimación por intervalos considera dos casos:

- **Desviación estándar poblacional es conocida**

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Donde $(1 - \alpha)$ es el coeficiente de confianza y $z_{\frac{\alpha}{2}}$ es el valor de z que proporciona un área $\alpha/2$ en la cola superior de la distribución de probabilidad normal estándar. Los niveles de confianza más recomendados son 90%, 95% y 99%. En la Tabla 11 se presenta los valores de z recomendados para estos niveles de confianza.

Tabla 2. Valores de $z_{\alpha/2}$ para los niveles de confianza más usados

Nivel de confianza	A	$\alpha/2$	$z_{\alpha/2}$
90%	0,10	0,05	1,64
95%	0,05	0,025	1,96
99%	0,01	0,005	2,58

▪ Desviación estándar poblacional es desconocida

Cuando se desconoce la desviación estándar de la población es necesario recurrir a la **distribución t**. La distribución **t** es una familia de distribuciones de probabilidad similares; cada distribución **t** depende de un parámetro conocido como grados de libertad.

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Donde s es la desviación estándar muestral, $(1 - \alpha)$ es el coeficiente de confianza y $t_{\alpha/2}$ es el valor de **t** que proporciona un área de $\alpha/2$ en la cola superior de la distribución **t** para $n - 1$ grados de libertad.

Tamaño de la muestra

Como se mencionó anteriormente $z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ es el margen de error. Sea,

$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Despejando n se tiene,

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{E^2}$$

Para usar la ecuación anterior es necesario contar con el valor de la desviación estándar poblacional σ . Sin embargo, aun cuando este valor no se conozca, puede usarse la ecuación siempre que se tenga un valor preliminar. En la práctica, se suele usar alguno de los procedimientos siguientes para obtener este valor planeado de σ :

1. Usar como valor planeado de σ una estimación de la desviación estándar poblacional calculada a partir de datos de estudios anteriores.
2. Emplear un estudio piloto seleccionando una muestra preliminar. La desviación estándar muestral obtenida de la muestra preliminar puede usarse como valor planeado de σ .
3. Use su juicio para el valor de σ . Por ejemplo, se puede empezar por estimar el mayor y el menor valor en los datos de la población. Esta diferencia entre el mayor y el menor valor proporciona una estimación del rango de los datos. Por último, este valor dividido entre 4 suele considerarse como una aproximación burda a la desviación estándar y tomarse como un valor planeado aceptable de σ .

b. De la proporción poblacional

De manera similar, la fórmula general para obtener una estimación por intervalo para la proporción poblacional es:

$$\bar{p} \pm \text{Margen de error}$$

Siendo el margen de error:

$$E = z_{\alpha/2} \sqrt{\frac{(\bar{p})(1-\bar{p})}{n}}$$

Donde $(1 - \alpha)$ es el coeficiente de confianza, y $z_{\alpha/2}$ es el valor de z que deja un área $\alpha/2$ en la cola superior de la distribución normal estándar.

Tamaño de la muestra

Ahora vamos a revisar cómo calcular el tamaño de la muestra para obtener una estimación de la proporción poblacional con una precisión determinada. La función que tiene el tamaño de la muestra en la determinación de la estimación por intervalo de p es semejante a la que tiene en la estimación de la media poblacional, ya estudiada en la sección anterior.

Despejando n de la fórmula del margen de error, se obtiene la fórmula para calcular el tamaño de la muestra con el que se tendrá el margen de error deseado, E .

$$n = \frac{\left(z_{\alpha/2}\right)^2 (\bar{p})(1-\bar{p})}{E^2}$$

Como generalmente no se conoce el valor poblacional de la proporción se recomienda usar algunos de los métodos siguientes para realizar una aproximación.

1. Utilizar la proporción poblacional de una muestra previa de las mismas unidades o de unidades similares.
2. Utilizar un estudio piloto y elegir una muestra preliminar.
3. Si no aplica ninguna de las alternativas anteriores, emplear como valor planeado $p = 0.50$.

El margen de error deseado para calcular una proporción poblacional casi siempre es **0,10** o menos. En las encuestas de opinión pública un margen de error de **0,03** o **0,04** es común. Con dichos márgenes de error, la ecuación suministra un tamaño de la muestra que es suficiente para satisfacer los requerimientos de **$np > 5$ y $n(1 - p) > 5$** para usar una distribución normal como una aproximación de la distribución muestral de \bar{x} .

Si desea ampliar la información sobre el cálculo de intervalos de confianza puede revisar la *guía didáctica de Estadística*, en la Unidad 6 se aborda esta temática.



Actividades finales del bimestre



Semana 16

¡Estamos finalizando el bimestre!

Durante esta última semana les recomiendo realizar una revisión general de lo estudiado en el bimestre y revisar las autoevaluaciones una vez más. Esta información será de gran utilidad para preparar su evaluación presencial.

¡Éxitos!

Índice

Primer
bimestre

Segundo
bimestre

Solucionario

Referencias
bibliográficas



4. Solucionario

Autoevaluación 1		
Pregunta	Respuesta	Retroalimentación
1	a	La Estadística es el sistema o método usado en la recolección, organización, análisis y descripción numérica de la información.
2	a	La estadística descriptiva consiste en la descripción de un conjunto de datos sin llegar a conclusiones con respecto a un grupo mayor
3	a	Población es un recuento de las unidades que tienen una característica común
4	c	Un grupo de estudiantes es una población finita
5	b	El género es una cualidad que describe una característica o atributo de un individuo
6	b	El número de visitantes en una tienda es una variable de tipo discreto
7	c	El salario es una variable cuantitativa
8	a	Una variable nominal no realiza comparaciones
9	c	La velocidad es una variable cuantitativa de escala de razón
10	a	Temperatura es una variable cuantitativa en escala de intervalo

Ir a la
autoevaluación

Autoevaluación 2		
Pregunta	Respuesta	Retroalimentación
1	b	Este tipo de medidas comprenden la media, mediana y moda
2	a	En aquellos casos cuando se ha seleccionado bien la muestra, la media es el valor más estable.
3	b	Se calcula sumando los valores de cada observación y dividiendo para n
4	b	Se calcula sumando los valores de cada observación y dividiendo para n .
5	b	Se calcula multiplicando el valor de la observación por la frecuencia, luego se suman estos valores, y se divide para n .
6	c	Es la diferencia que existe entre cada observación y el promedio calculado.
7	a	En toda distribución de datos la suma de las desviaciones respecto a la media es cero.
8	b	Solamente se puede calcular un valor por grupo y no se permiten hacer generalizaciones para otras muestras.
9	a	Luego de haber organizado los valores de menor a mayor, es el valor que ocupa la cuarta posición, siendo $n=7$.
10	b	Ya que no existe un valor particular para la mediana, es necesario calcular el promedio entre los valores que ocupen la posición n y $n+1$.

Ir a la
autoevaluación

Autoevaluación 3		
Pregunta	Respuesta	Retroalimentación
1	c	Los modelos binomial y de Poisson corresponden de distribuciones de probabilidad de variables discretas.
2	a	Es la fórmula de la distribución binomial
3	a	Es una de las características de los experimentos binomiales, la probabilidad de éxito no puede cambiar de ensayo en ensayo.
4	b	Cuando $(n \cdot p)$ es mayor o igual a 5.
5	b	Siempre se requiere un área bajo la curva, hay tres casos $P(a < x < b)$; $P(a < x)$; $P(x > b)$.
6	a	La probabilidad de un valor específico x no se puede calcular bajo una curva, por lo tanto sería igual a cero.
7	b	La probabilidad de un valor específico x no se puede calcular bajo una curva, por lo tanto sería igual a cero.
8	b	El valor de la variable aleatoria es menor que el promedio el valor será negativo.
9	b	Es el valor que se obtiene del valor de la tabla para distribución normal, el valor se ubica a la izquierda del promedio.
10	a	El valor que provee la tabla es el área entre $x=10$ y el promedio, pero el área de interés es $x < 10$, por lo tanto es el área que queda por fuera del valor que nos da la tabla de probabilidad normal.

Ir a la
autoevaluación

Autoevaluación 4		
Pregunta	Respuesta	Retroalimentación
1	b	Lo primero es establecer lo que se quiere evaluar o probar por medio de una prueba de hipótesis.
2	b	Al buscar en la tabla de distribución normal el valor que corresponde a $Z_{\alpha/2}$ para un área bajo la curva de 0,495; el valor que corresponde es 2,58
3	a	Se trata de un ejemplo con desviación estándar poblacional conocida
4	c	Se trata de una prueba de hipótesis que mide si existen diferencias entre el promedio muestral y poblacional, es una prueba a dos colas
5	a	El error tipo I consiste en rechazar una hipótesis nula cuando es verdadera
6	b	Es una prueba de hipótesis de la cola superior
7	a	Se debe usar z, ya que se tiene información de la población.
8	c	El error tipo II implica que se acepta la hipótesis nula cuando es falsa.
9	b	Se rechaza la hipótesis si el valor crítico calculado es mayor que el valor z del alpha del nivel de significancia
10	a	Se rechaza la hipótesis si el valor-p calculado es menor que el alpha del nivel de significancia

Ir a la
autoevaluación



5. Referencias bibliográficas

Anderson, D.R., Sweeney D.J. y Williams, T.A. (2016). *Estadística para administración y economía*. México, Cengage Learning.

Física y Mates (2015) *Distribución Binomial* (video). Recuperado de [enlace web](#).

Física y Mates (2015) *Distribución Binomial: media y varianza* (video). Recuperado de [enlace web](#).

Lind, D., Marchal, W., Wathen, S., (2012). *Estadística aplicada a los negocios y la economía*. México, Mc Graw Hill.

Martínez, C., (2012) *Estadística y muestreo*. Colombia, Ecoe Ediciones.

Montero Espinoza (2018) *Coeficiente de simetría* (video). Recuperado de [enlace web](#).

Píldoras matemáticas (2017) *Como usar una tabla de distribución normal* (video). Recuperado de [enlace web](#).

Píldoras matemáticas (2017) *Función de distribución normal* (video). Recuperado de [enlace web](#).

Tuto Mate (2016) *Elaboración de una tabla de frecuencias (variable discreta)* (video). España. Recuperado de [enlace web](#).

Tuto Mate (2016) *Elaboración de una tabla de frecuencias (variable continua)* (video). España. Recuperado de [enlace web](#).