UTPL

La Universidad Católica de Loja

Vicerrectorado de Modalidad Abierta y a Distancia

Estadística para las Ingenierías y la Arquitectura

Guía didáctica

UTPL

Facultad Ingenierías y Arquitectura

Estadística para las Ingenierías y la Arquitectura

Guía didáctica

Carrera	PAO Nivel
Tecnologías de la información	IV

Autor:

Juan Carlos Torres Díaz

ESTA_2029

Universidad Técnica Particular de Loja

Estadística para las Ingenierías y la Arquitectura

Guía didáctica

Juan Carlos Torres Díaz

Diagramación y diseño digital

Ediloja Cía. Ltda.
Marcelino Champagnat s/n y París
edilojacialtda@ediloja.com.ec
www.ediloja.com.ec

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

Año de edición: 29 de septiembre de 2020

Edición: primera edición reestructurada en septiembre 2024 (con un cambio del 5%)

Loja-Ecuador

Índice

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

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.

A través del siguiente resultado de aprendizaje, el estudiante comprenderá el significado y la utilidad de la estadística en la resolución de problemas. A través de la introducción a la estadística, se les proporcionará una base sólida para comprender cómo esta disciplina desempeña un papel fundamental en la recopilación, organización y análisis de datos, contribuyendo así a abordar una variedad de problemas en diversos contextos.

Contenidos, recursos y actividades de aprendizaje recomendadas

Recuerde revisar de manera paralela los contenidos con las actividades de aprendizaje recomendadas y actividades de aprendizaje evaluadas.

Semana 1

1.1 Definiciones de estadística

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.

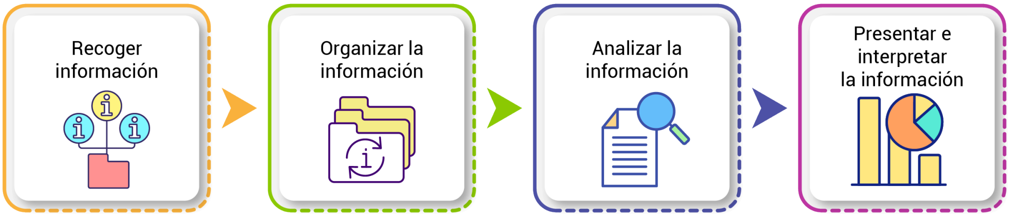
Antes de definir el término estadística es recomendable que usted lea las secciones "introducción" y "¿por qué se debe estudiar estadística?", del **texto básico**, en estas secciones el autor nos hace una introducción muy interesante que sirve como punto de partida. Seguidamente, es necesario leer la 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.

Estimado estudiante, observe el video denominado Estadística unidad 1: conceptos básicos y tome nota de las definiciones de estadística.

Nota: por favor, complete la actividad en un cuaderno de apuntes o documento Word.

Figura 1

Definición de estadística



Nota. Adaptado de *Estadística aplicada a los negocios y la economía* (p.2), por Lind, D., Marchall, W., Waten, S., 2015, CC BY 2.0.

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 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 considerables.

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

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 puede 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.

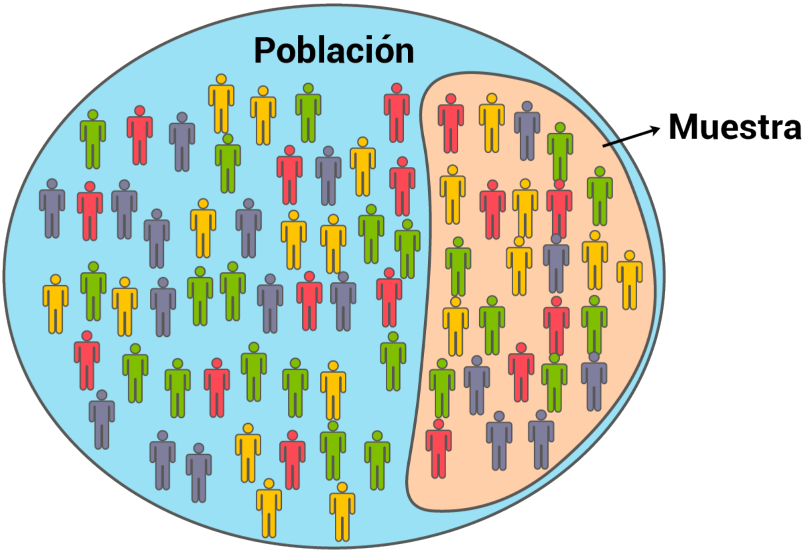
Hay una imagen en la página 8 del **texto básico** que sin necesidad de palabras nos va a aclarar qué es una población y qué es una muestra; sin embargo, es conveniente que tenga claro un concepto más formal.

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.



Nota. Tomado de *Muestreo* [Ilustración], por Requena, B., 2014, Universo de fórmulas, CC BY 2.0.

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.

Estimado estudiante, le invito a observar el video denominado Estadística unidad 1: conceptos básicos y, tome nota de las diferencias entre estadística descriptiva e inferencial.

Nota: por favor, complete la actividad en un cuaderno de apuntes o documento Word.

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, etc.

Estimado estudiante, observe el video denominado Estadística unidad 1: conceptos básicos en donde se dan ejemplos y explicaciones al respecto.

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.

Para una mayor comprensión del tema, lo invito a revisar un resumen de estos niveles en la siguiente infografía:

Niveles de medición

Estimado estudiante, observe el video denominado Estadística unidad 1: conceptos básicos y tome nota de los ejemplos dados para los distintos niveles de medición.

Nota: por favor, complete la actividad en un cuaderno de apuntes o documento Word.

Actividades de aprendizaje recomendadas

Para reforzar el contenido de una manera práctica le recomiendo desarrollar las siguientes actividades:

1. Busque en *Internet* información sobre las encuestas de elección del 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?

   Nota: por favor, complete la actividad en un cuaderno o documento Word.

2. Le invito cordialmente a llevar a cabo la práctica que se detalla en el anexo 1. Practica 1.
3. Estimado estudiante, lo invito a desarrollar la autoevaluación, la cual le permitirá determinar el nivel de aprovechamiento del tema estudiado.

Autoevaluación 1

Con base en la siguiente encuesta conteste las preguntas:

Edad
Sexo M ( ) F( )
Tiene teléfono móvil SÍ ( ) NO( )
Cuánto gasta al mes en telefonía móvil
¿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

Resultado de aprendizaje 2:

Representa datos utilizando gráficas.

A través del siguiente resultado de aprendizaje, el estudiante adquirirá habilidades para representar datos de manera efectiva utilizando gráficas, comprendiendo las distribuciones de frecuencia y siendo capaces de describir y comunicar de manera efectiva la información derivada de estas representaciones visuales.

Contenidos, recursos y actividades de aprendizaje recomendadas

Recuerde revisar de manera paralela los contenidos con las actividades de aprendizaje recomendadas y actividades de aprendizaje evaluadas.

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.

- Ingrese en Tableau Public para poder descargar el *software*.
- Para poder descargar R deberá ingresar en R for Windows y para descargar RStudio deberá ingresar en RStudio Desktop.

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

En esta sección ponemos especial énfasis en el indicador de aprendizaje "construye una distribución de frecuencias para presentar un conjunto de datos" planificado 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, el cual lo invito a revisar en el anexo 2. Ejemplo de organización de conjuntos de datos.

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 apartado de 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.

Estimado estudiante, lo invito a observar el video Estadística unidad 2: organización de datos en el que se indica paso a paso el proceso de construir distribuciones de frecuencia.

En el **texto básico** se presentan ejemplos detallados, en esta guía vamos a complementar esto con los ejercicios desarrollados en el anexo 3. Ejercicios de distribución de frecuencias.

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 $2k \geq n$.
3. Calcular el tamaño de clase, la fórmula utilizada es: Tamaño de clase= Rango/k.

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

4. Elaboración de la tabla de frecuencias.

Utilizando *Tableau* construya una tabla de frecuencias para la variable “marca” del teléfono 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.

Vamos a obtener las frecuencias relativas de los ejemplos planteados anteriormente (ver anexo 4).

Una explicación detallada de cómo obtener una frecuencia relativa y de cómo realizar la respectiva gráfica se halla en el siguiente video Estadística unidad 2: organización de datos.

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 que se detalla a continuación, en donde se construye una tabla de frecuencias de las marcas de teléfono, frecuencias y el respectivo gráfico. Tome nota de los aspectos que considere importantes.

Tabla 1
Tabla de frecuencias de la marca de teléfono

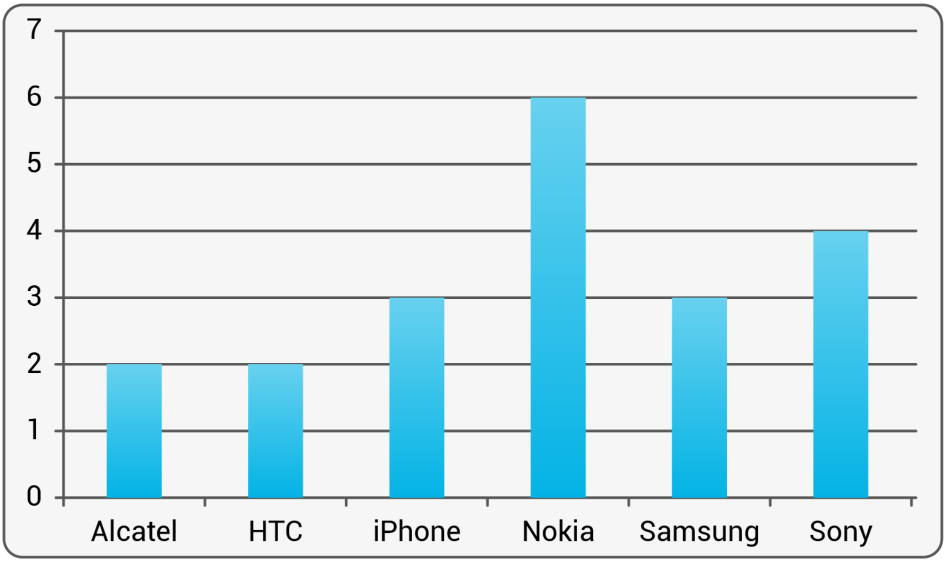
Marca	Frecuencia
Alcatel	2
HTC	2
iPhone	3
Nokia	6
Samsung	3
Sony	4

Nota. Torres, J., 2023.

Esta tabla se representa en la siguiente figura:

Figura 3

Diagrama de frecuencias



Nota. Torres, J., 2023.

Note los siguientes aspectos:

a. Las barras están separadas, esto es 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.

Observe también el video Estadística unidad 2: organización de datos, 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.

Actividades de aprendizaje recomendadas

Continuemos con el aprendizaje mediante su participación en las actividades que se describen a continuación:

1. Estimado estudiante, ingrese a Tableau y observe los siguientes videos:
    - Introducción.
    - La interfaz de *Tableau*.
    - Conexión a los datos.

- Análisis visual.

2. Utilizando *Tableau* construya un histograma para la variable “edad” del archivo de datos “encuesta”.
3. Corra línea a línea el código en R del archivo “histograma” y observe las gráficas resultantes.
4. Lo animo cordialmente a realizar la práctica, cuyos detalles se encuentran en el anexo 5. Práctica 2.
5. Realice la autoevaluación para comprobar sus conocimientos.

Autoevaluación 2

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 fórmula 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

Resultado de aprendizaje 3:

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

A través del siguiente resultado de aprendizaje, el estudiante comprenderá la estructura de un conjunto de datos mediante la aplicación de medidas de tendencia central y dispersión. Se abordarán las medidas de tendencia central y medidas de dispersión. A la vez, se explorará cómo estas herramientas estadísticas proporcionan información crucial sobre la distribución y variabilidad de los datos.

Contenidos, recursos y actividades de aprendizaje recomendadas

Recuerde revisar de manera paralela los contenidos con las actividades de aprendizaje recomendadas y actividades de aprendizaje evaluadas.

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

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 1000 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 en cuanto a razón es la edad. Se puede obtener la edad promedio de los habitantes de una ciudad. La edad es una variable que se mide con respecto a razón.

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.

Estimado estudiante, lo invito a observar y tomar nota del video unidad 3. Medidas de tendencia central y medidas de dispersión.

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

17	18	18	19	21	22	23	60	70	77
				Valor central					

Nota. Torres, J., 2023.

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

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.

Estimado estudiante, lo invito a observar y tomar nota del video unidad 3. Medidas de tendencia central y medidas de dispersión.

3.3 Moda

La moda representa el 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**.

Actividades de aprendizaje recomendadas

Reforcemos el aprendizaje resolviendo las siguientes actividades:

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

   Nota: por favor, complete las actividades en un cuaderno o documento Word.

3. Le insto amablemente a llevar a cabo la actividad práctica cuyos pormenores están descritos en el anexo 6. Práctica 3.
4. Estimado estudiante, lo invito a desarrollar la autoevaluación, la cual le permitirá determinar el nivel de aprovechamiento del tema estudiado.

Autoevaluación 3

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 se repite 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

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 recomendadas

Semana 4

Unidad 3. Medidas de tendencia central y medidas de dispersión

3.4 Medidas de dispersión

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

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

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 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 2
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

Nota. Torres, J., 2023.

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 3

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	1260,25
77	77–34,5	42,5	1806,25
Total			**5278,5**
Varianza = 5278,5 / 10			**527,85**

Nota. Torres, J., 2023.

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

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 dar lectura, ha dicho teorema ubicado en la sección

“Interpretación y usos de la desviación estándar”. Seguidamente de lectura y, observe el gráfico correspondiente a la Regla empírica, esta regla 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 percibidos; 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.

Estimado estudiante lo invito a observar y tomar nota del video unidad 3. medidas de tendencia central y medidas de dispersión.

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

Es hora de reforzar los conocimientos adquiridos resolviendo las siguientes actividades:

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

   Nota: por favor, complete las actividades en un cuaderno o documento Word.

3. Es momento de aplicar su conocimiento, para ello le invito a resolver el anexo 7. Práctica 4.
4. Estimado estudiante, lo invito a desarrollar la autoevaluación, la cual le permitirá determinar el nivel de aprovechamiento del tema estudiado.

Autoevaluación 4

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 qué 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 no 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

Resultado de aprendizaje 2:

Representa datos utilizando gráficas.

A través del siguiente resultado de aprendizaje, el estudiante adquirirá habilidades para representar datos de manera efectiva utilizando gráficas, comprendiendo las distribuciones de frecuencia y siendo capaces de describir y comunicar de manera efectiva la información derivada de estas representaciones visuales.

Contenidos, recursos y actividades de aprendizaje recomendadas

Recuerde revisar de manera paralela los contenidos con las actividades de aprendizaje recomendadas y actividades de aprendizaje evaluadas.

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" planificado 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 llámeme, 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 leer la sección "Diagramas de puntos" del **texto básico** y observar el diagrama del ejemplo que se desarrolla.

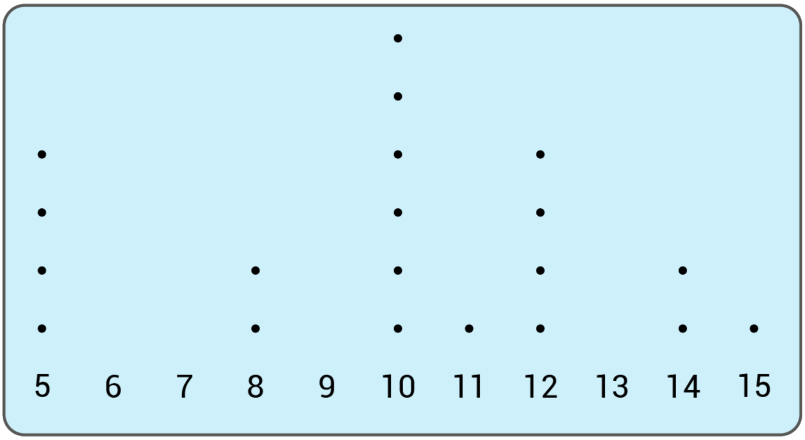
Para complementar vamos a resolver el siguiente ejemplo: dado el siguiente conjunto de datos, que corresponde al valor mensual que gastan en telefonía un grupo de 20 personas.

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



Nota. Torres, J., 2023.

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:

Tabla 4

Valores menores que 10

Tallo	Hojas					
0	5	5	5	5	8	8

Nota. Torres, J., 2023.

Para los valores mayores o iguales a 10:

Tabla 5
Valores mayores o iguales a 10

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

Nota. Torres, J., 2023.

Ahora unimos los dos

Tabla 6
Valores mayores o iguales a 10

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

Nota. Torres, J., 2023.

Y ya tenemos nuestro 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 el 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 7

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
300	450	540	590	640	710	810	900	1200	1600

Nota. Torres, J., 2023.

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 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.

Estimado estudiante, observe el siguiente video denominado unidad 4. Descripción de datos 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.

- 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 decil del conjunto de datos.

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.

Con los cuartiles se pueden 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.

Estimado estudiante, observe el siguiente video denominado unidad 4. Descripción de datos 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.

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 8

Peso y edad de un grupo de 18 niños

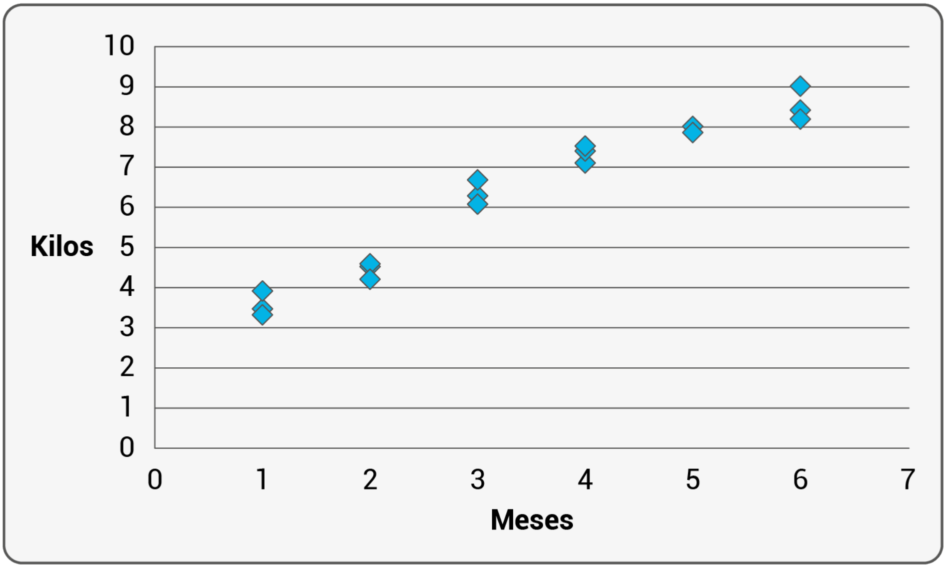
Edad en meses	Peso en libras
1	3,47
1	3,3
1	3,9
2	4,5
2	4,6
2	4,2
3	6,26
3	6,1
3	6,7
4	7,1
4	7,4
4	7,5
5	7,9
5	8
5	7,85
6	8,2
6	8,4
6	9

Nota. Torres, J., 2023.

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 Xx vamos a colocar la edad y en el eje de las y vamos a colocar el peso.

Figura 6

Relación edad peso en niños



Nota. Torres, J., 2023.

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.

Resumen

1. 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.
2. Un diagrama de caja permite apreciar la dispersión de los datos.
3. Tanto los deciles, como los cuartiles y los percentiles, son medidas de dispersión de los datos.

Actividades de aprendizaje recomendadas

Es momento de aplicar su conocimiento a través de las actividades que se han planteado a continuación:

1. Resuelva los ejercicios del 11 al 14 del **texto básico**, estos se encuentran al final de la sección que trata “Otras medidas de dispersión”.
2. Con los cuartiles se pueden 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.
3. 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”.

    Nota: por favor, complete las actividades en un cuaderno o documento Word.

4. Ahora es el instante propicio para poner en práctica sus conocimientos; por eso, lo invito a abordar la práctica detallada en anexo 8. Práctica 5.
5. Estimado estudiante, lo invito a desarrollar la autoevaluación, la cual le permitirá determinar el nivel de aprovechamiento del tema estudiado.

Autoevaluación 5

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 se 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 realiza 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

Resultados de aprendizaje 4 y 5:

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

A través de los resultados de aprendizaje, el estudiante comprenderá cómo calcular probabilidades de sucesos y expectativas de variables aleatorias, y distinguir entre sucesos dependientes e independientes. Se abordará las probabilidades y distribución de probabilidad discreta. Además, se explorarán conceptos clave que permiten evaluar la incertidumbre en eventos y entender la relación entre sucesos.

Contenidos, recursos y actividades de aprendizaje recomendadas

Recuerde revisar de manera paralela los contenidos con las actividades de aprendizaje recomendadas y actividades de aprendizaje evaluadas.

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 y expectativas 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.

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 asociada a cuán posible es que ocurra o no ocurra, en esta unidad vamos a estudiar las probabilidades y cómo se distribuyen.

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.

Aquí diremos qué 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.

Estimado estudiante lo invito a observar el video UTPL Aspectos básicos en el estudio de la probabilidad [(Área administrativa) (Estadística I)] y, tome nota de los conceptos.

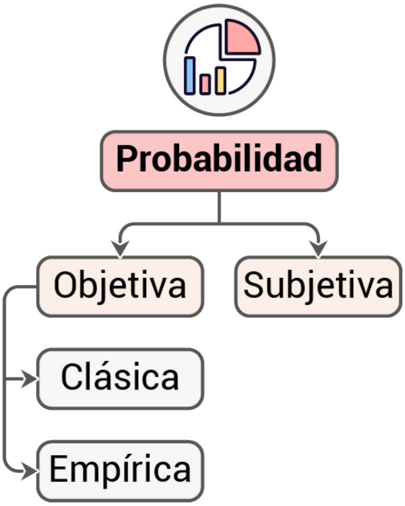
Nota: Por favor, complete la actividad en un cuaderno de apuntes o documento Word.

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



Nota. Torres, J., 2023.

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 toda 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 eso este asigna un valor de probabilidad.

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 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

Revisar los ejercicios y explicaciones dadas en el anexo 9. Probabilidades y distribución de probabilidad discreta.

- **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 9

Matriculados en asignaturas

Matriculados	Cantidad	Probabilidad
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**

Nota. Torres, J., 2023.

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 probabilidad 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 del 75 %, la probabilidad de que no apruebe es del 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 las asignaturas es del 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)

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, saque 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 referente al número de autos vendidos por la agencia John Ragsdale y determine cómo 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, se va a comprender sin mayores dificultades.

Actividades de aprendizaje recomendadas

Continuemos con el aprendizaje mediante su participación en las actividades que se describen a continuación:

1. 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”.

   Nota: por favor, complete la actividad en un cuaderno o documento Word.

2. Lo animo cordialmente a realizar la práctica, cuyos detalles se encuentran en el anexo 10. Práctica 6.
3. Estimado estudiante, lo invito a desarrollar la autoevaluación, la cual le permitirá determinar el nivel de aprovechamiento del tema estudiado.

Autoevaluación 6

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 simultánea.
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

Resultados de aprendizaje 1 a 5:

- Entiende el alcance y utilidad de la estadística en la resolución de problemas.
- Representa datos utilizando gráficas.
- Describe la estructura de un conjunto de datos a través de sus medidas de tendencia central y de dispersión.
- Calcula probabilidades de sucesos y expectativas de variables aleatoria.
- Diferencia entre sucesos dependientes e independientes.

Contenidos, recursos y actividades de aprendizaje recomendadas

Semana 7

Actividades finales del bimestre

Estimado estudiante, para esta semana lo invito a hacer un repaso de los contenidos del primer bimestre, específicamente de las siguientes unidades:

- Unidad 1. Introducción a la estadística.
- Unidad 2. Distribuciones de frecuencia y representaciones gráficas.
- Unidad 3. Medidas de tendencia central y medidas de dispersión.

Resultados de aprendizaje 1 a 5:

- Entiende el alcance y utilidad de la estadística en la resolución de problemas.
- Representa datos utilizando gráficas.
- Describe la estructura de un conjunto de datos a través de sus medidas de tendencia central y de dispersión.
- Calcula probabilidades de sucesos y expectativas de variables aleatoria.
- Diferencia entre sucesos dependientes e independientes.

Contenidos, recursos y actividades de aprendizaje recomendadas

Semana 8

Actividades finales del bimestre

Estimado estudiante, esta semana le animo a revisar los contenidos del primer bimestre, centrándose especialmente en las siguientes unidades:

- Unidad 4. Descripción de datos.
- Unidad 5. Probabilidades y distribución de probabilidad discreta.

¡Felicitaciones por completar el primer bimestre!

Continúe con el mismo entusiasmo y dedicación en sus estudios durante el próximo 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.

Este resultado de aprendizaje se centra en proporcionar a los estudiantes una comprensión teórica de la distribución de probabilidad normal. Además, tiene como objetivo equipar a los estudiantes con las habilidades prácticas necesarias para aplicar este conocimiento en la interpretación y análisis de conjuntos de datos, utilizando herramientas como intervalos de confianza y áreas bajo la curva normal.

Contenidos, recursos y actividades de aprendizaje recomendadas

Recuerde revisar de manera paralela los contenidos con las actividades de aprendizaje recomendadas y actividades de aprendizaje evaluadas.

Semana 9

Unidad 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 básico** 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.

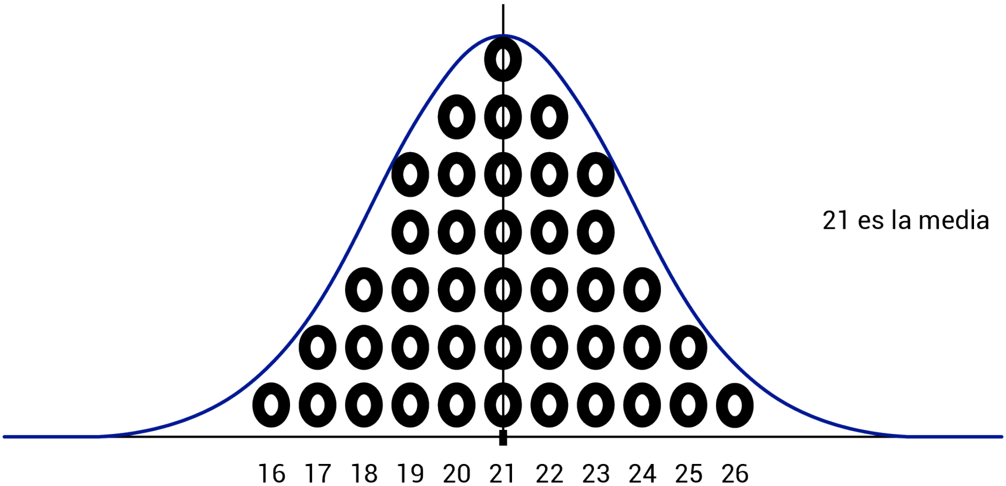
Inicialmente, es necesario entender qué es distribución normal y algunas de sus características, le recomiendo, por tanto, observar el video UTPL curva normal [(gestión ambiental) (bioestadística)] y tomar nota de los aspectos más importantes.

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



Nota. Torres, J., 2023.

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
- 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 qué es una distribución normal, vamos a dar lectura a la sección "Familia de distribuciones de la probabilidad normal" y determine 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, etc.
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*.

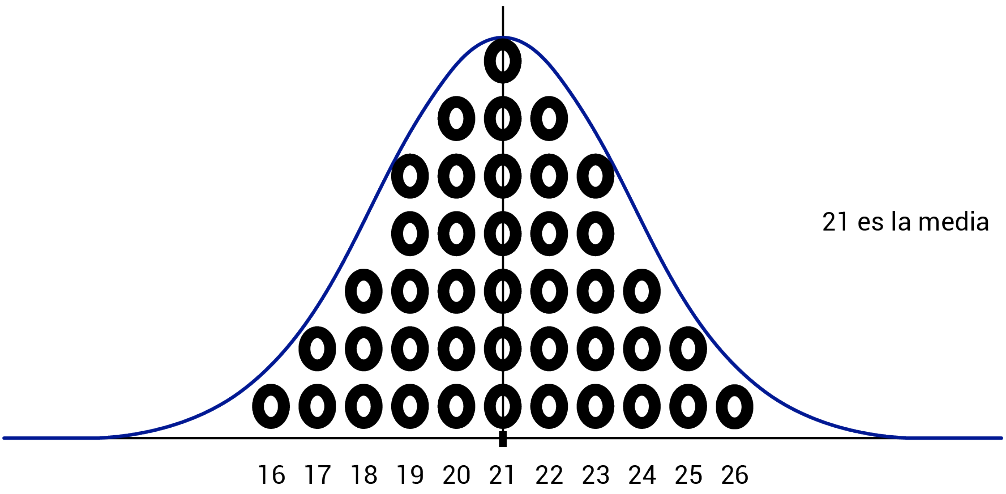
Ahora lo invito a dar lectura a la sección "Distribución de probabilidad normal *estándar*", en esta sección va a encontrar:

- 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



Nota. Torres, J., 2023.

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:

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

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.

Ahora encontremos el valor Z para la edad de 24 años:

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

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

Ahora encontremos el valor Z para la edad 19.

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

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

Luego de que revisó detenidamente los conceptos y ejemplos; usted está listo para 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, si tiene dudas en el video Cálculo de áreas y en el video UTPL distribución de probabilidad normal [(área administrativa)(estadística I)] puede verificar cada uno de los pasos.

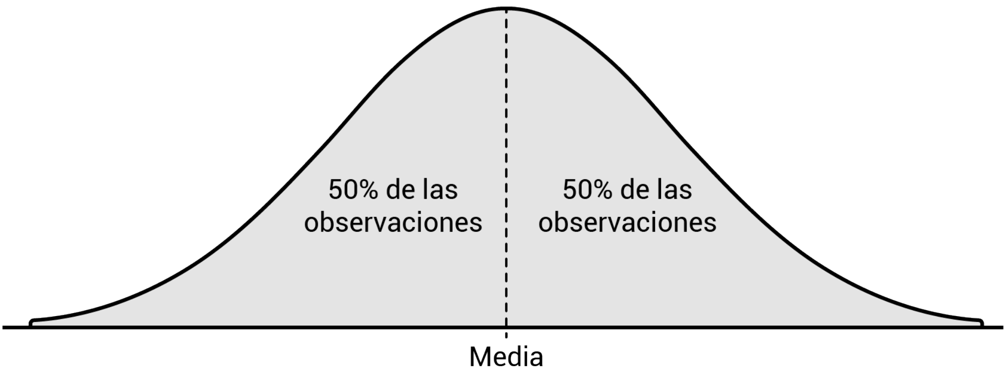
Más adelante, se presentan varios ejercicios de distinto tipo, le recomiendo que los revise todos, ponga especial atención a la gráfica en donde se señala el área que se está calculando, comprender esto es vital para entender las operaciones de sumas o restas de áreas.

Antes de sumergirnos en la lectura del **texto básico**, revisaremos 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



Nota. Torres, J., 2023.

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.

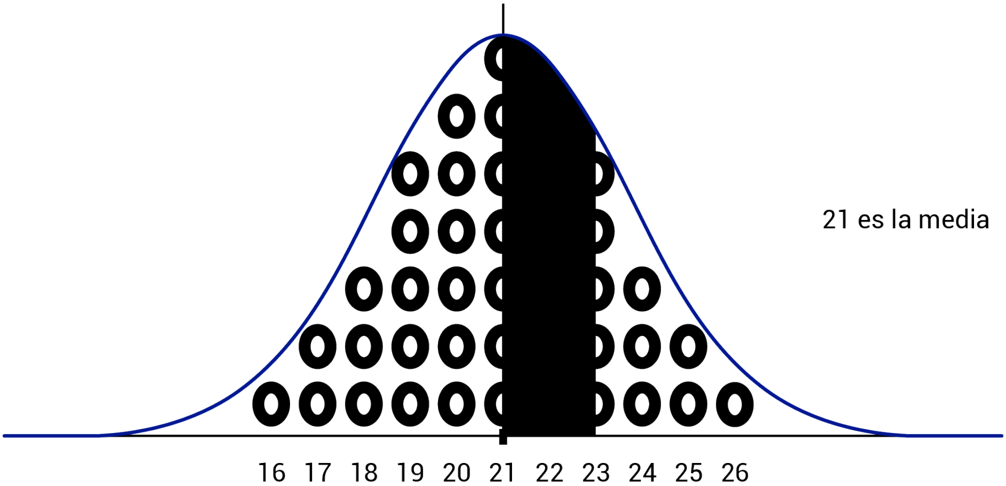
$$Z = \frac{X-\mu}{\delta} \quad Z = \frac{23-21}{2} \quad Z = \frac{2}{2} \quad Z = 1$$

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



Nota. Torres, J., 2023.

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 el 50 % del valor de la parte sombreada. Esto es:

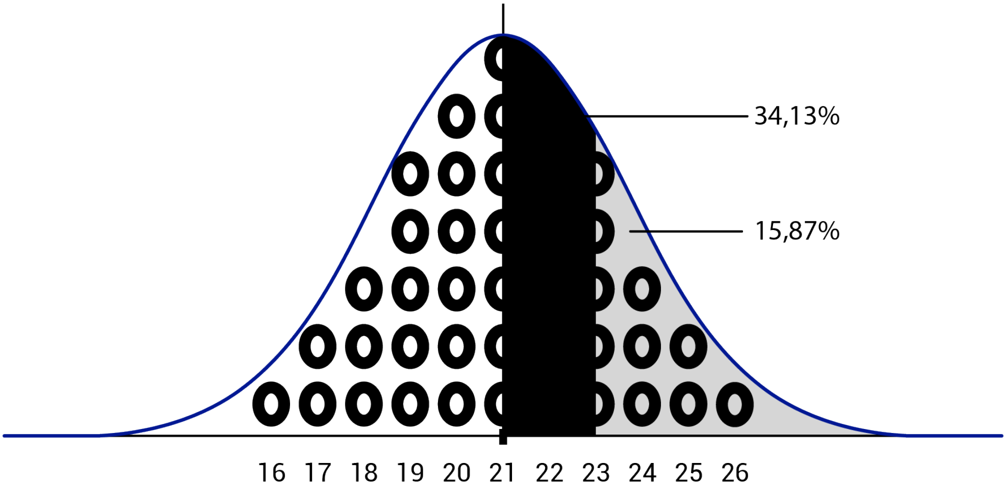
Proporción = 0,5 – 0,3413

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

El resultado se puede apreciar en la siguiente figura:

Figura 12

Distribución de áreas antes y después de 23



Nota. Torres, J., 2023.

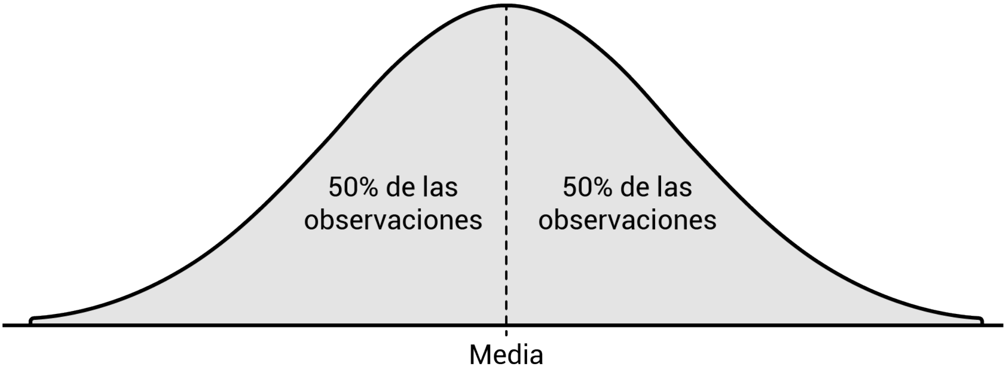
Ahora vamos a trabajar con el **texto básico**, dando lectura a la sección "Determinación de áreas bajo la curva normal" y tomando nota de los hallazgos importantes. Una recomendación para la lectura es: volver a resolver los ejercicios que ya han sido resueltos en el **texto básico** e ir comparando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

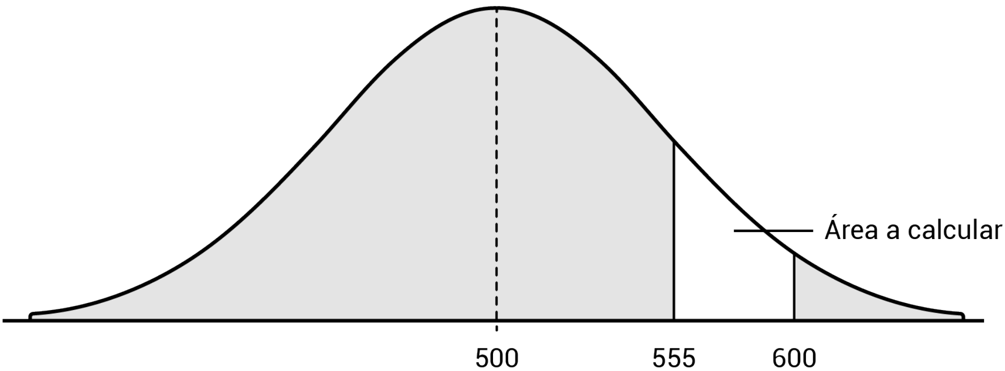


Nota. Torres, J., 2023.

Como paso 2, observe la figura 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



Nota. Torres, J., 2023.

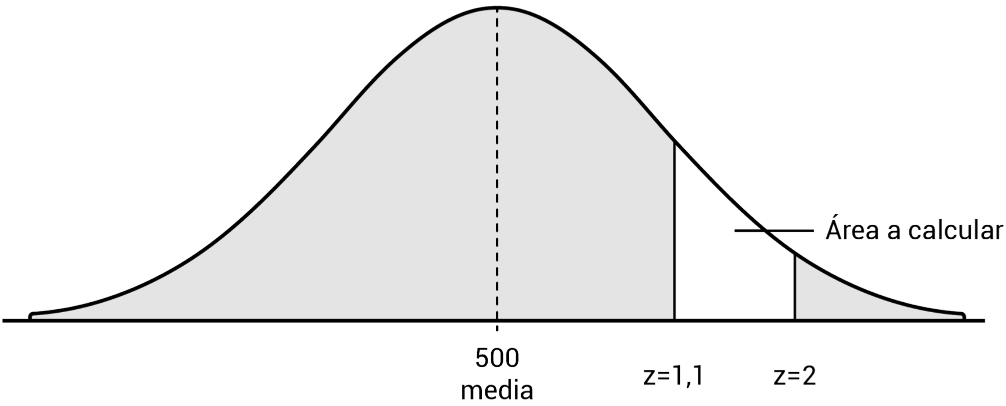
El valor de área que buscamos se encuentra en la tabla de distribución normal. Pero para poder determinarla necesitamos conocer los valores de Z para 555 y para 600.

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

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

Figura 15

Áreas entre z=1,1 y z=2



Nota. Torres, J., 2023.

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

Para cada valor de Z, buscamos el porcentaje que le 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 del 36,43 % (en la tabla consta el valor 0,3643), esto corresponde al porcentaje desde la media hasta Z=1,1.

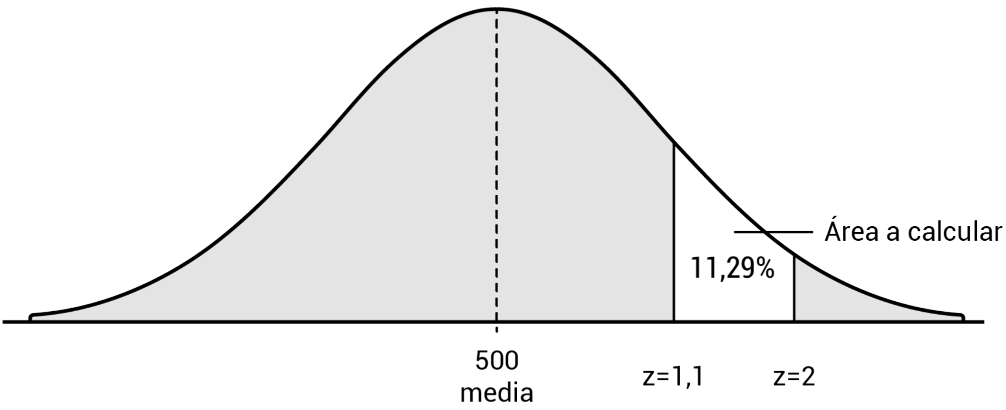
El porcentaje que corresponde a Z = 2 es del 47,72 % (en la tabla consta el valor 0,4772), esto corresponde al porcentaje desde la media hasta Z = 2.

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



Nota. Torres, J., 2023.

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.

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

Es momento de aplicar su conocimiento a través de las actividades que se han planteado a continuación:

1. 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*".
2. Resuelva los ejercicios del 23 al 29 del **texto básico**, estos se encuentran al final de la sección "Determinación de áreas bajo la curva".

   Nota: por favor, complete las actividades en un cuaderno o documento Word.

3. Le invito cordialmente a llevar a cabo la práctica que se detalla en el anexo 11. Práctica 7.
4. Estimado estudiante, lo invito a desarrollar la autoevaluación, la cual le permitirá determinar el nivel de aprovechamiento del tema estudiado.

Autoevaluación 7

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.
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

Resultado de aprendizaje 4:

Calcula probabilidades de sucesos y expectativas de variables aleatorias.

A través del resultado de aprendizaje 4, el estudiante aprenderá a calcular probabilidades y expectativas de variables aleatorias, fundamentales para el análisis estadístico. Este proceso abarca el estudio del muestreo, la distribución muestral y el teorema central del límite, además de los conceptos de correlación y regresión lineal. Este enfoque integral proporciona una base sólida para la aplicación de técnicas estadísticas en diversos contextos.

Contenidos, recursos y actividades de aprendizaje recomendadas

Recuerde revisar de manera paralela los contenidos con las actividades de aprendizaje recomendadas y actividades de aprendizaje evaluadas.

Semana 10

Unidad 7. Muestreo, distribución muestral y teorema central del límite

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 de muestreo”.

Además, 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.

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 con base en 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.

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.

Nota: por favor, complete la actividad en un cuaderno de apuntes o documento Word.

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 mil 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 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 10
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

Nota. Torres, J., 2023.

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, página 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 primera 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.

La figura 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.

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

Es hora de reforzar los conocimientos adquiridos resolviendo las siguientes actividades:

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

   Nota: por favor, complete las actividades en un cuaderno o documento Word.

3. Le insto amablemente a llevar a cabo la actividad práctica cuyos pormenores están descritos en el anexo 12. Práctica 8.

4. Estimado estudiante, lo invito a desarrollar la autoevaluación, la cual le permitirá determinar el nivel de aprovechamiento del tema estudiado.

Autoevaluación 8

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

Resultado de aprendizaje 4:

Calcula probabilidades de sucesos y expectativas de variables aleatorias.

Contenidos, recursos y actividades de aprendizaje recomendadas

Semana 11

Unidad 7. Muestreo, distribución muestral y teorema central del límite

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.

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, además 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”. Dé lectura al mismo y note que hay dos tipos de cálculos:

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.

¿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 va a permitir es de 100 dólares y el nivel de confianza del 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.
- δ es la desviación *estándar* de la población, en este caso 120.
- E es el error que se va a permitir que alcance 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.

7.5.2 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. La fórmula para 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 del 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?
    - $n = p.\,q(z/E)^2$
    - $n = 0,6 * 0,4(1,96/0,04)^2$
    - $n = 0,6 * 0,4(2401), = 0,6$ equivale a la proporción de personas que apoyan la actual política.
    - $n = 576,24$

- El tamaño de la muestra debe ser de 577 encuestas.

7.6 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.

Tabla 11

Muestras del tamaño 2 del ejercicio del texto básico

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	Sa, Ted	7,9	16	8,00					

Nota. Tomado de *Estadística aplicada a los negocios y economía*, por Lind et al., 2015, McGrawHill.

Es necesario recalcar que la distribución de la media muestral se da con base en 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.

En la siguiente figura, la gráfica de la derecha corresponde a la tabla de frecuencias.

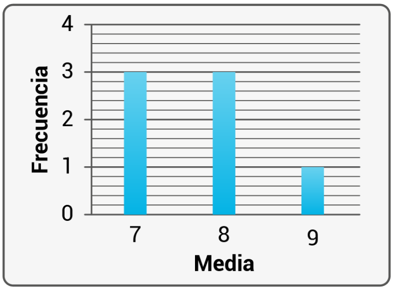
Note la forma que tiene la gráfica, aunque tiende a ser normal, los dos primeros valores son iguales.

Figura 17

Frecuencia de los salarios

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

Media	Frecuencia
7	3
8	3
9	1



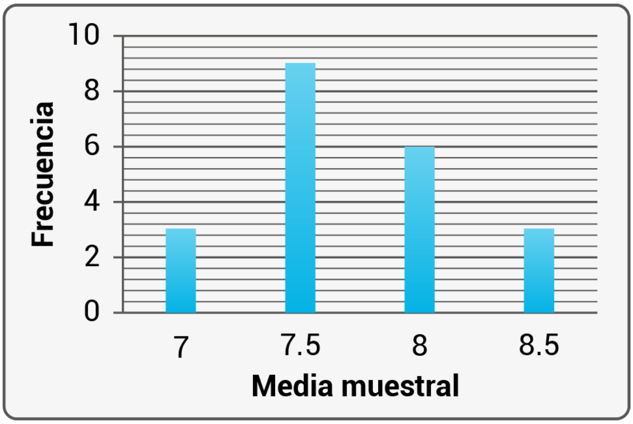
Nota. Torres, J., 2023.

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.

Figura 18

Distribución de la media muestral

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



Nota. Torres, J., 2023.

Ahora es necesario analizar las diferencias y entender el porqué 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: 12, 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 12
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

Nota. Torres, J., 2023.

Ahora construimos la tabla de frecuencias de las medias muestrales.

Tabla 13
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

Nota. Torres, J., 2023.

Ahora vamos a responder los requerimientos del ejercicio:

- Todas las combinaciones se pueden ver en la tabla anterior.
- 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 de la siguiente manera:

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

Luego hacemos todas las combinaciones de tamaño 2 (todas las muestras posibles de tamaño 2).

Tabla 14
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

Nota. Torres, J., 2023.

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.

7.7 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 revisar 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.

En el anexo 13. Ejercicios de teorema central del límite, 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.

Resumen

1. La distribución muestral de la media consiste en todas las posibles medias de todas las muestras de determinado tamaño.
2. La media de las medias muestrales es igual a la media poblacional.
3. La distribución muestral de la media tiene forma de campana y se perfecciona conforme aumenta el tamaño de la muestra.
4. 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.
5. 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

Es momento de aplicar su conocimiento a través de las actividades que se han planteado a continuación:

1. 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**.

2. Es momento de aplicar su conocimiento, para ello le invito a resolver el anexo 14. Práctica 9.
3. Resolver los ejercicios que se encuentran en el **texto básico** en las páginas 232 y 233.
4. 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.

   Nota: por favor, complete las actividades en un cuaderno o documento Word.

5. Ahora es el momento propicio para poner en práctica lo que ha aprendido. También le extendemos la invitación a abordar la práctica 10, la cual se encuentra descrita en detalle en el anexo 15.
6. Estimado estudiante, lo invito a desarrollar la autoevaluación, la cual le permitirá determinar el nivel de aprovechamiento del tema estudiado.

Autoevaluación 9

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 un 95 %. Se conoce que en estudios anteriores aplicados en la ciudad la desviación estándar fue 100. El tamaño adecuado para esa muestra es 700.
11. ( ) A mayor tamaño de la muestra menor nivel de error muestral.
12. ( ) La distribución de la media muestral tiende a mostrarse como una campana de forma normal.
13. ( ) 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.
14. ( ) 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.
15. ( ) 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.

16. ( ) 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.
17. ( ) Mientras más grande es la muestra, tiende menos a una distribución normal.
18. ( ) 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.
19. ( ) Si no se toman todas las muestras de tamaño n, la media de la distribución muestral de medias va a tener una gran diferencia con la media poblacional.
20. ( ) 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

Resultado de aprendizaje 7:

Realiza modelos de regresión con la ayuda de paquetes computacionales.

A través del resultado de aprendizaje, el estudiante adquirirá habilidades para construir y analizar modelos de regresión empleando herramientas computacionales. Mediante el estudio de la correlación y la regresión lineal, se podrá identificar y evaluar relaciones entre variables, optimizando así la toma de decisiones fundamentadas en datos cuantitativos.

Contenidos, recursos y actividades de aprendizaje recomendadas

Recuerde revisar de manera paralela los contenidos con las actividades de aprendizaje recomendadas y actividades de aprendizaje evaluadas.

Semana 12

Unidad 8. Correlación y regresión lineal

8.1 Correlación

El análisis de correlación es el estudio de la relación entre variables. También se define como un grupo de técnicas para medir el nivel de asociación entre dos variables.

Suponga el siguiente ejemplo:

Tabla 15

Registros del archivo de llamadas telefónica

Vendedor	Llamadas	Copiadoras vendidas
V1	20	30
V2	40	60
V3	20	40
V4	30	60
V5	10	30
V6	10	40
V7	20	40
V8	20	50
V9	20	30
V10	30	70

Nota. Torres, J., 2023.

En el archivo se presentan los 10 primeros registros de las llamadas que hacen los vendedores de una fábrica de copiadoras para vender sus productos. Para ejecutar este código utilice en Rstudio el archivo copiadoras.csv que se encuentra en Canvas. Cargue el archivo de datos:

>mis_datos <- read.csv("copiadoras.csv")

Luego ejecute el comando cor (obtiene la correlación)

> cor(mis_datos$Llamadas, mis_datos$CopiadorasVendidas)

El resultado que se obtiene es: 0.7179361

El Coeficiente de correlación (r) es una medida de la fuerza de la relación entre dos variables. Requiere datos a nivel de intervalo o escala. El índice r puede variar desde -1.00 a 1.00.

Cuando r toma los valores de -1.00 o 1.00 tenemos una correlación perfecta y fuerte.

Cuando los valores son cercanos a 0.0 tenemos una correlación débil. En nuestro ejemplo el valor de la correlación es de 0.7179361, se trata de una relación fuerte, se acerca a 1, es positivo lo que nos indica que conforme se incrementa una variable, se incrementa también la otra.

Los valores negativos indican una relación inversa, es decir que cuando la una variable se decrementa la otra también lo hace.

Los valores positivos indican una relación directa, es decir cuando se incrementa la variable independiente, se incrementa también la dependiente.

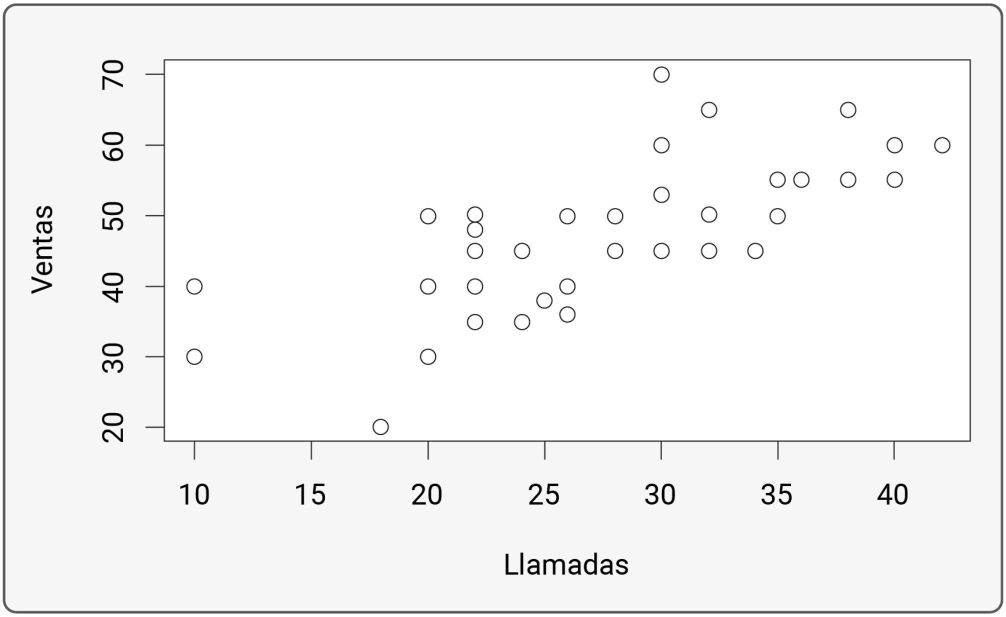
La relación entre dos variables se puede graficar, se utiliza un diagrama de dispersión que se dibuja con el siguiente comando:

plot(mis_datos$Llamadas, mis_datos$CopiadorasVendidas, xlab='Llamadas', lab='Ventas')

El resultado es el siguiente:

Figura 19

Diagrama de dispersión de las llamadas telefónicas y las ventas



Nota. Torres, J., 2023.

En la figura se nota una tendencia, conforme se incrementa el número de llamadas, se incrementa también la cantidad de copiadoras vendidas. El diagrama de dispersión es un gráfico que muestra la relación entre dos variables. Es el primer paso habitual en el análisis de correlaciones.

8.2 Regresión lineal

La regresión lineal es una técnica estadística que se utiliza para analizar las relaciones entre dos variables. En su forma más simple, implica dos variables:

- una variable independiente X (también llamada variable predictora).
- una variable dependiente Y (también llamada variable de respuesta).

El objetivo es encontrar la mejor línea recta que describe cómo cambia Y en función de X.

Esta línea recta se define por la ecuación:

$$Y = a + bX$$

- La regresión lineal mide la relación entre una variable dependiente y una o más variables independientes.
- La variable dependiente es la variable que se predice o estima.
- La variable independiente proporciona la base para la estimación. Es la variable predictora.

Se utiliza una ecuación para expresar la relación entre variables, lo que nos permite estimar una variable sobre la base de otra.

A manera de ejemplo diremos:

La UTPL gasta en publicidad cada semestre una determinada cantidad de dinero, esa cantidad tiene incidencia en el número de estudiantes matriculados. Si la cantidad de dinero se incrementa, se incrementarán los estudiantes matriculados; si esa cantidad disminuye, disminuirán los estudiantes matriculados. La regresión lineal busca determinar con detalle cuál es la afectación de una variable sobre la otra, la afectación que produce la variable independiente sobre la dependiente.

Existe una relación entre la cantidad de horas que un estudiante dedica a estudiar y la calificación obtenida en el examen de una materia.

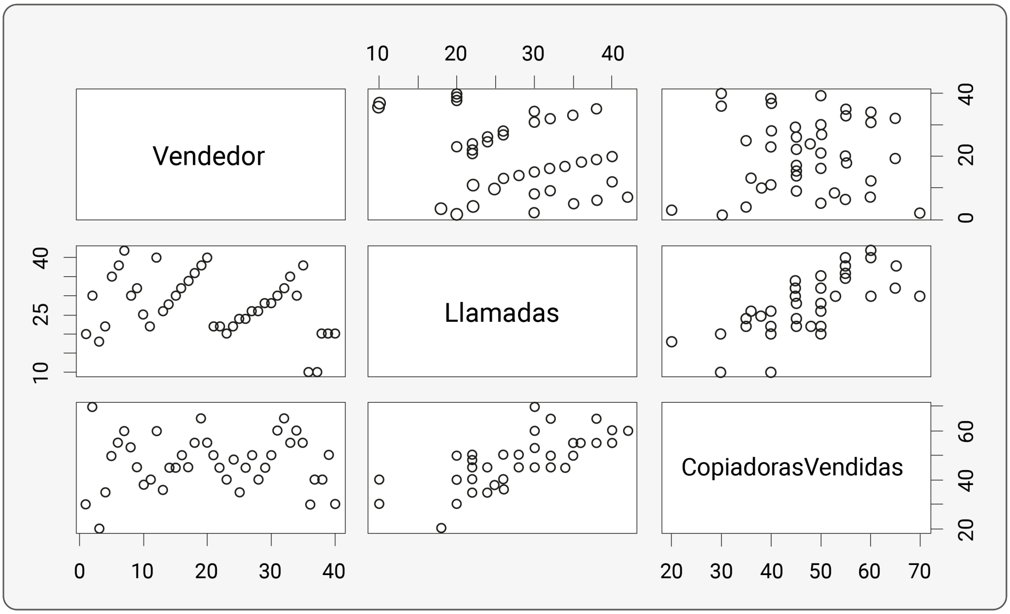
- La variable dependiente Y es la calificación obtenida.
- La variable independiente X es la cantidad de horas que la persona dedica a estudiar.

Una fábrica de copiadoras tiene agentes que venden su producto a través de llamadas. El siguiente código R genera la regresión lineal para determinar como influyen las llamadas en el número de copiadoras vendidas.

```
#Lectura de datos desde el archivo csv
setwd("~/Google Drive/guia regresion lineal")
mis_datos <- read.csv("copiadoras.csv")

#pairs calcula la correlación de todas las variables del archivo
pairs(mis_datos)

#Cálculo del índice r de correlación
cor(mis_datos$Llamadas, mis_datos$CopiadorasVendidas)
```

Figura 20

Correlación de las variables

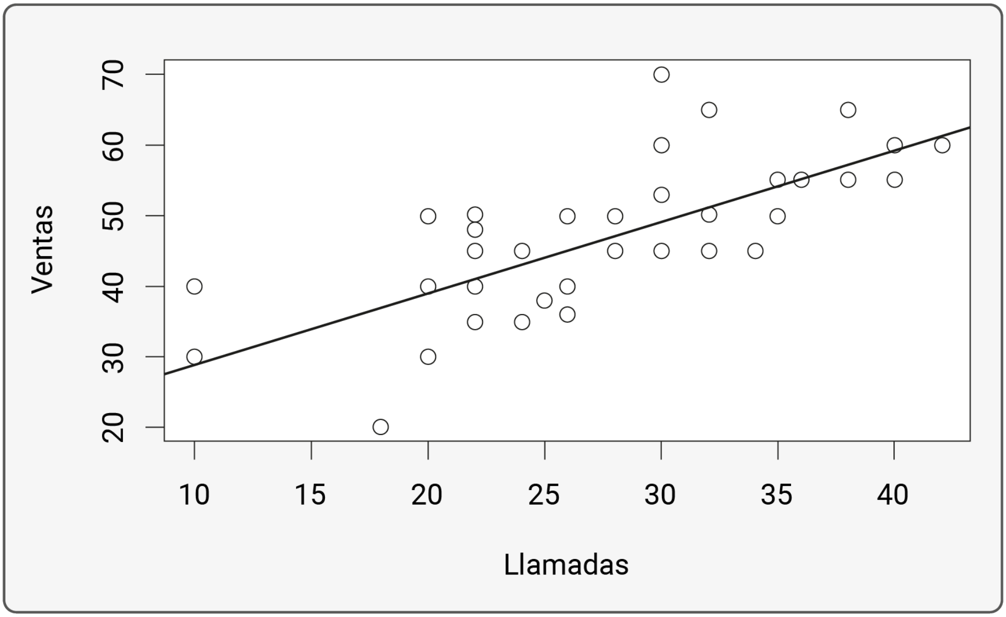


Nota. Torres, J., 2023.

```
#El comando lm de R permite calcular la regresión, en esta se especifica
#la variable dependiente: CopiadorasVendidas
#la variable independiente: Llamadas (número de llamdas telefónicas realizadas)

regresion <- lm(CopiadorasVendidas ~ Llamadas, data = mis_datos)
summary(regresion)

#Se grafica el resultado en un diagrama de puntos
plot(mis_datos$Llamadas, mis_datos$CopiadorasVendidas, xlab='Llamadas', ylab='Ventas')

#Se traza la recta de regresión
abline(regresion)
```

Figura 21

Recta de regresión



Nota. Torres, J., 2023.

La ecuación resultante se escribe de la siguiente manera:

$$Y = 19.2821 + 1.0079*X$$

esta ecuación representa la relación entre las dos variables, se puede utilizar para predecir el valor de Y si se va asignando valores a X. Ejemplo:

Para X=23 el valor que se proyecta para Y es:

$Y = 19.2821 + 1.0079*(23)$
$Y = 19.2821 + 23.1817$
$Y = 42.47$

Esto significa que si se hacen 23 llamadas el número de copiadoras que se espera vender es 42

Para X=41 el valor que se proyecta para Y es:

$Y = 19.2821 + 1.0079*(41)$
$Y = 19.2821 + 41.3239$
$Y = 60.60$

Esto significa que si se hacen 41 llamadas el número de copiadoras que se espera vender es 60.

Actividades de aprendizaje recomendadas

Es momento de aplicar su conocimiento a través de las actividades que se han planteado a continuación:

1. Vaya al capítulo correspondiente del **texto básico**, en la sección: ¿Qué es el análisis de correlación? Y determine:
    - ¿Qué es análisis de correlación?
    - ¿Coeficiente de correlación?
    - Plantee al menos tres ejemplos de correlación.
2. Con los datos del archivo estudiantes.csv, que contiene registros del número de horas que dedica al estudio una persona y la calificación

que obtiene, programe la regresión lineal y obtenga la ecuación de regresión correspondiente.

3. Una vez obtenida la ecuación de regresión, haga predicción de la calificación para los siguientes valores: 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10.

   El archivo estudiantes.csv fue descargado de: Students Score Dataset - Linear Regression.

   Nota: por favor, complete las actividades en un cuaderno de apuntes o documento Word.

4. Le invito a reforzar sus conocimientos, participando en la siguiente autoevaluación.

Autoevaluación 10

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

1. ( ) Correlación se denomina al nivel de relación que existe entre dos variables.
2. ( ) Cuando el coeficiente de correlación es cercano a cero, significa que se cuenta con una correlación fuerte entre las variables.
3. ( ) Un valor del coeficiente de correlación de -1 indica que se tiene una relación perfecta pero inversa.
4. ( ) Una relación directa entre dos variables indica que si se incrementa variable independiente, se incrementa también la dependiente.
5. ( ) La variable dependiente se denomina también predictor.
6. ( ) La variable independiente es la que incide en la dependiente.
7. ( ) El comando cor de R permite calcular el coeficiente de correlación.

8. (   ) El comando pairs de R permite calcular el coeficiente de correlación.
9. (   ) La recta de regresión cuenta con un valor llamado intercepto que indica el valor de Y cuando x toma el valor de cero.
10. (   ) La recta de regresión se puede utilizar para predecir cuando ya se ha determinado su ecuación.

Ir al solucionario

Resultado de aprendizaje 6:

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

Para alcanzar este resultado de aprendizaje se equipará a los estudiantes con las habilidades necesarias para llevar a cabo un proceso completo y riguroso de verificación de hipótesis, abordando tanto escenarios en los que se conoce como en los que no se conoce la desviación estándar de la población.

Contenidos, recursos y actividades de aprendizaje recomendadas

Recuerde revisar de manera paralela los contenidos con las actividades de aprendizaje recomendadas y actividades de aprendizaje evaluadas.

Semana 13

Unidad 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 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 ≠ 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 ≠ 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 cuándo se trata de una prueba de dos colas y cuándo 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 comienza 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. 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 contempla 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 16
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
6	Existe diferencia	≠	H1
7	No ha cambiado	=	H0

Nota. Tomado de *Estadística aplicada a los negocios y economía*, *presentación correspondiente*, por Lind, D. et al, 2015, Mc GrawHill-UTPL, Loja Ecuador.

Nota: Los operadores se utilizan en la hipótesis indicada según el enunciado del problema.

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 menciona 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 Mac Burger 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 Mac Burger 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 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, porque el operador de la H1 indica hacia el lado derecho).

Luego de finalizar la unidad, deberá estar en capacidad de plantear correctamente los operadores relacionales de las hipótesis.

Actividades de aprendizaje recomendadas

Es hora de reforzar los conocimientos adquiridos resolviendo las siguientes actividades:

1. Estimado estudiante, lo invito a observar el video UTPL verificación de hipótesis para resolver dudas que puedan existir, en este video se presentan más ejercicios y proceso para resolverlos.

2. Del **texto básico** lo invito a 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…”.

   Nota: por favor, complete la actividad en un cuaderno o documento Word.

3. Le invito cordialmente a llevar a cabo el ejercicio práctico que se detalla en el anexo 16. Práctica 11.
4. Estimado estudiante, lo invito a desarrollar la autoevaluación, la cual le permitirá determinar el nivel de aprovechamiento del tema estudiado.

Autoevaluación 11

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 una 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 calcula.
7. ( ) La prueba de una cola implica que el error se ubica en los dos lados de la curva.

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

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 recomendadas

Semana 14

Unidad 9. Hipótesis y pruebas de hipótesis

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 una comprensión más clara de esta verificación, les invito a explorar la siguiente infografía, en la cual se ilustra un ejemplo:

[Verificación de hipótesis]

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 el procedimiento es similar y solo se cambia los valores Z por valores t. Para más detalles revise el siguiente ejercicio:

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.

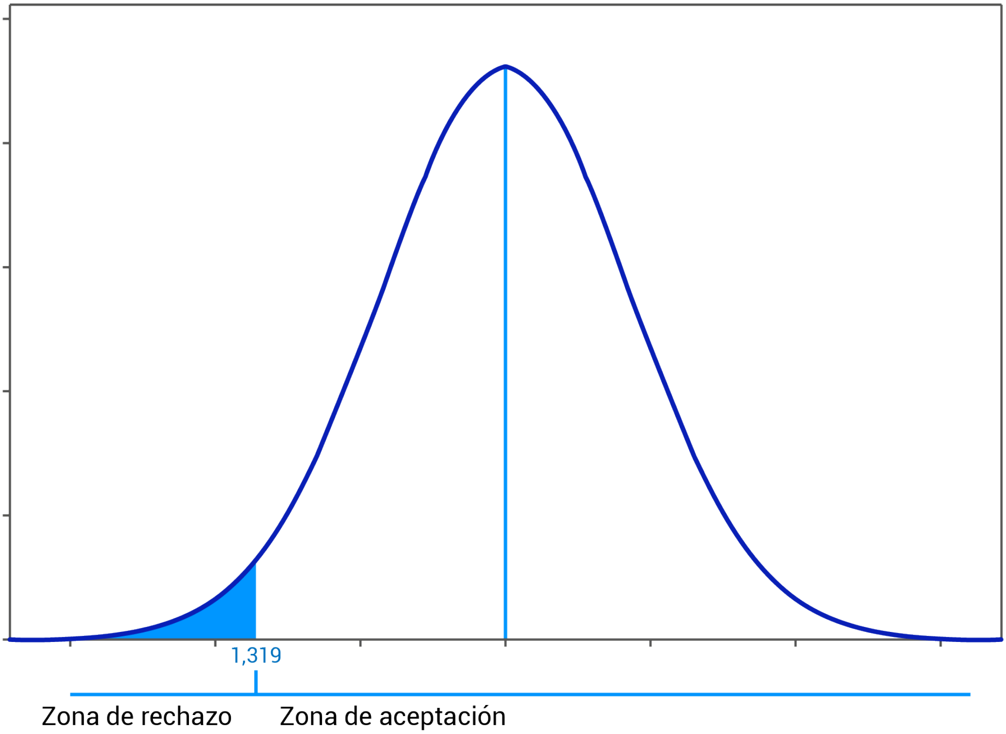
Paso 4: regla de decisión.

- 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 22

Curva normal con valor t para 23 grados de libertad y 0.9 nivel de significancia



Nota. Torres, J., 2023.

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.

$$t = (40,6–42,3)/(2,7/\sqrt{24})$$
$$t = -1,7/(2,7/4,89)$$
$$t = -1,7/0,552$$
$$t = -3,079$$

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.

Actividades de aprendizaje recomendadas

Es momento de aplicar su conocimiento a través de las actividades que se han planteado a continuación:

1. Estimado estudiante, lo invito a revisar el video UTPL verificación de hipótesis, para resolver dudas que puedan existir, en este video se presentan más ejercicios y procesos para resolverlos.
2. Lea la teoría respectiva de la temática prueba de hipótesis de una muestra.
3. Resolver los ejercicios 1 al 4 de las páginas 293 y 294.
4. Resolver los ejercicios 9 al 14 de la página 298 y 299.

   Nota: por favor, complete las actividades en un cuaderno o documento Word.

5. Es momento de aplicar su conocimiento, para ello le invito a resolver el anexo 17. Práctica 12.
6. Estimado estudiante, lo invito a desarrollar la autoevaluación, la cual le permitirá determinar el nivel de aprovechamiento del tema estudiado.

Autoevaluación 12

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 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 cuándo 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

Resultados de aprendizaje 4, 6 y 7:

- Calcula probabilidades de sucesos y expectativas de variables aleatorias.
- Aplica la distribución normal a conjuntos de datos y calcula intervalos de confianza y áreas bajo la curva normal.
- Realiza modelos de regresión con la ayuda de paquetes computacionales.

Contenidos, recursos y actividades de aprendizaje recomendadas

Semanas 15 y 16

Actividades finales del bimestre

Apreciado estudiante, en esta semana lo invito a revisar las unidades correspondientes al segundo bimestre.

¡En hora buena por haber completado con éxito el estudio de los dos bimestres! Este logro refleja su dedicación y esfuerzo continuo en el aprendizaje.

Ahora, se encuentra en una posición sólida para abordar desafíos académicos futuros. ¡Siga así, con determinación y entusiasmo!

4. Solucionario

Autoevaluación 1

Pregunta	Respuesta	Retroalimentación
1	F	La edad es una variable medida a nivel de razón, ya que tiene un punto cero claro y se pueden realizar operaciones matemáticas.
2	F	La edad no se mide a nivel ordinal ni de intervalo, sino de razón.
3	V	El sexo es una variable nominal, ya que, no tiene un orden inherente.
4	F	La pregunta 3 se refiere a tener o no tener teléfono móvil, que es una variable nominal, no de razón.
5	V	El gasto mensual en telefonía móvil es una variable de razón, porque tiene un punto cero y se pueden realizar operaciones matemáticas.
6	V	La autoevaluación en términos de mensajes diarios escritos implica un orden cualitativo, por lo que es una variable ordinal.
7	V	La inferencia estadística implica hacer afirmaciones sobre una población basándose en muestras, y a menudo se utiliza un estadístico como estimación del parámetro poblacional.
8	V	Una muestra es un subconjunto representativo de la población, y se utiliza para hacer inferencias sobre la población.
9	V	Las variables continuas pueden tomar cualquier valor en un intervalo, incluyendo decimales.
10	F	El número de hijos de una familia generalmente se considera una variable discreta, ya que solo puede tomar valores enteros y no valores continuos.

Ir a la autoevaluación

Autoevaluación 2

Pregunta	Respuesta	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 cuántos 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 de forma progresiva.

Ir a la autoevaluación

Autoevaluación 3

Pregunta	Respuesta	Retroalimentación
1	F	La diferencia entre la media muestral y la media poblacional radica en que una se calcula a partir de una muestra y la otra de la población completa.
2	F	La media poblacional no es un estadístico; es un parámetro poblacional.
3	V	La mediana es menos sensible a valores extremos, por lo que puede ser preferible en conjuntos de datos muy dispersos.
4	V	La moda es el valor que ocurre con mayor frecuencia en un conjunto de datos porque representa el dato que se repite más veces
5	V	La mediana se encuentra en la posición central de un conjunto de datos ordenados.
6	V	La mediana es el valor que se encuentra en la posición central cuando los datos están ordenados.
7	F	La mediana y la media pueden tener valores diferentes, especialmente en conjuntos de datos sesgados.
8	V	La media aritmética es la medida de tendencia central más comúnmente utilizada.
9	F	La media y la moda pueden ser diferentes y proporcionar información diferente sobre la distribución de los datos.
10	V	El promedio de edad de los ecuatorianos sería un parámetro, ya que se refiere a la población completa y no a una muestra.

Ir a la autoevaluación

Autoevaluación 4

Pregunta	Respuesta	Retroalimentación
1	V	El rango representa la diferencia entre el valor máximo y el valor mínimo en un conjunto de datos.
2	V	La desviación estándar mide la dispersión de los datos con respecto a la media.
3	V	Datos dispersos indican que los valores se extienden a lo largo del rango. En lugar de concentrarse cerca de un valor central, los datos dispersos abarcan una amplia gama de valores.
4	F	Si la mayoría de los datos están cerca de la media, la dispersión es baja, ya que indica que los datos están agrupados alrededor de la media.
5	V	La varianza y la desviación estándar son medidas más precisas y robustas de la dispersión que el rango.
6	V	La dispersión se refiere a la extensión o distribución de los valores en un conjunto de datos.
7	V	La desviación estándar se calcula tomando la raíz cuadrada de la varianza, que a su vez se calcula a partir de las diferencias entre cada valor y la media.
8	V	Para calcular la desviación estándar, se deben seguir varios pasos: a. Calcular la media de los valores: (4+5+6)/3=15/3=5(4+5+6)/3= 15/3=5. b. Restar la media de cada valor y elevar al cuadrado: (4−5)2=1(4−5)2=1, (5−5)2=0(5−5)2=0, (6−5)2=1(6−5)2=1. c. Calcular el promedio de esos cuadrados: (1+0+1)/3=2/3(1+0+1)/3=2/3. d. Tomar la raíz cuadrada del promedio de los cuadrados: 2/32/3, que no es igual a 1. Por lo tanto, la desviación estándar no es 1 para este conjunto de valores.
9	V	El rango es la diferencia entre el valor máximo (78) y el valor mínimo (23), que es 55.
10	F	

Pregunta	Respuesta	Retroalimentación
		El rango no involucra elevar al cuadrado las desviaciones. La varianza y la desviación estándar son las medidas que implican elevar al cuadrado las diferencias respecto a la media.

Ir a la autoevaluación

Autoevaluación 5

Pregunta	Respuesta	Retroalimentación
1	V	Un diagrama de puntos puede proporcionar una visión general, pero la tabla de distribución de frecuencias ofrece detalles específicos.
2	V	El diagrama de tallo y hojas condensa los datos manteniendo su información esencial.
3	F	Los cuartiles dividen un conjunto de observaciones en cuatro partes iguales, no en 10.
4	V	Los deciles indican la posición relativa de un valor dentro de un conjunto de datos dividido en diez partes iguales.
5	V	Los percentiles representan la posición relativa de un valor en un conjunto de datos.
6	V	Los diagramas de caja muestran la dispersión, la mediana y los cuartiles en un conjunto de datos.
7	V	Un diagrama de caja abarca desde el primer cuartil (Q1) hasta el tercer cuartil (Q3), representando la dispersión central de los datos.
8	V	El primer cuartil (Q1) es el mismo que el percentil 25, ya que ambos representan el 25 % de los datos.
9	F	El decil 4 es el percentil 40, ya que divide los datos en 10 partes iguales, y el percentil 50 es la mediana.
10	V	El percentil 100 representa el 100 % de los datos y es equivalente al cuarto cuartil (Q4), que es el máximo.

Ir a la autoevaluación

Autoevaluación 6

Pregunta	Respuesta	Retroalimentación
1	V	La probabilidad cuantifica la posibilidad de que ocurra un evento.
2	F	La probabilidad se calcula dividiendo el número de resultados favorables entre el número total de resultados posibles.
3	V	La mutua exclusión significa que no pueden ocurrir simultáneamente.
4	V	El término “probabilidad relativa” generalmente se asocia con la probabilidad condicional.
5	V	Un evento puede ser un solo resultado o una combinación de resultados.
6	V	Estos enfoques representan diferentes formas de asignar probabilidades.
7	V	En la probabilidad clásica, se asume que todos los resultados son igualmente probables.
8	V	La regla de multiplicación se aplica para calcular la probabilidad conjunta de eventos en diferentes contextos.
9	F	La regla general de la adición se aplica tanto a eventos mutuamente excluyentes como a eventos que pueden ocurrir conjuntamente.
10	V	Esta regla es fundamental en la teoría de la probabilidad y proporciona una manera rápida de calcular la probabilidad de un evento complementario.

Ir a la autoevaluación

Autoevaluación 7

Pregunta	Respuesta	Retroalimentación
1	V	La distribución normal exhibe simetría en relación con su medida central, que en este caso es la mediana.
2	V	Esta forma característica se debe a que la mayoría de las observaciones se concentran alrededor de la media, y la probabilidad de observar valores disminuyen a medida que nos alejamos de la media en ambas direcciones.
3	V	La distribución normal es continua y puede tomar infinitos valores.
4	V	La distribución normal es adecuada para variables continuas.
5	F	La variable "marca de un vehículo" es una variable categórica nominal y no cuantitativa. La distribución normal se aplica a variables cuantitativas continuas.
6	V	La distribución normal estándar es un caso específico de la distribución normal.
7	F	"Z" representa la puntuación Z o valor Z indicando cuántas desviaciones estándar, un dato particular se aleja del promedio en una distribución normal.
8	F	La regla del 68-95-99.7 establece que el 68 % de los datos están dentro de una desviación estándar de la media, el 95 % dentro de dos desviaciones estándar y el 99.7 % dentro de tres desviaciones estándar.
9	F	Aproximadamente el 99.7 % de las observaciones se encuentran dentro de 3 desviaciones estándar antes y después de la media.
10	V	La tabla de distribución normal proporciona áreas bajo la curva normal, expresadas en porcentaje.

Ir a la autoevaluación

Autoevaluación 8

Pregunta	Respuesta	Retroalimentación
1	V	La distribución de probabilidad enumera todos los posibles resultados de un experimento junto con las probabilidades asociadas a cada resultado.
2	V	La probabilidad se calcula dividiendo el número de resultados favorables entre el total de resultados posibles.
3	V	Las encuestas de muestra son más prácticas y menos costosas que encuestar a toda la población.
4	V	El muestreo es un procedimiento que implica seleccionar una muestra de una población más grande para hacer inferencias sobre toda la población.
5	F	El error de muestreo es la variación natural que se espera al estimar parámetros de población basados en muestras.
6	F	La simpatía por un candidato generalmente se mediría a través de la proporción de personas que lo favorecen.
7	F	Para estimar el salario promedio, se utilizará una medida de tendencia central, no una proporción.
8	V	Una tabla de distribución de probabilidad lista todos los posibles resultados de un experimento junto con las probabilidades asociadas a cada resultado.
9	V	Una muestra representativa debe reflejar las características clave de la población.
10	F	La diferencia radica en el tipo de variable que se está midiendo (media para variables cuantitativas y proporción para variables categóricas).

Ir a la autoevaluación

Autoevaluación 9

Pregunta	Respuesta	Retroalimentación
1	F	Aunque se prefiere que sea representativa, hay casos en los que puede no serlo.
2	V	Z representa el valor crítico asociado al nivel de confianza.
3	F	E representa el margen de error deseado, no está directamente relacionado con un límite del 5 %.
4	F	Para conocer las preferencias, se usaría la fórmula para la proporción, ya que estás midiendo una categoría.
5	V	Cuando no hay información previa, se asume el peor escenario (p = 0,5) para maximizar el tamaño.
6	V	El tamaño de la población influye en el cálculo del tamaño de la muestra.
7	F	No necesariamente. Una muestra representativa implica representar todas las características de la población, no solo del sector central.
8	V	Los valores de p y q de 0.5 maximizan el tamaño de la muestra.
9	F	En realidad, para el salario promedio se usaría la fórmula para la media, no la proporción.
10	F	Falta de información importante. En la práctica, para realizar estimaciones precisas, es crucial tener información completa sobre la población y seguir procedimientos estadísticos apropiados para determinar el tamaño de la muestra.
11	V	A mayor tamaño de muestra, el error muestral tiende a disminuir.
12	V	Esto es parte del teorema central del límite, que establece que, sin importar la forma de la población original, la distribución de la media muestral se aproxima a una distribución normal a medida que aumenta el tamaño de la muestra.
13	F	Aunque la distribución de la media muestral tiende a ser más normal a medida que aumenta el tamaño de la muestra, no adoptará una forma normal perfecta con cualquier conjunto de datos.
14	F	No necesariamente una muestra más pequeña lleva a una forma normal “mejor”. La convergencia a la normalidad es más pronunciada con un tamaño de muestra grande.

Pregunta	Respuesta	Retroalimentación
15	F	La elección del tamaño de la muestra depende de diversos factores, y no hay una regla general de que un tamaño específico sea mejor para obtener una distribución más normal.
16	V	Una muestra grande suele garantizar que la distribución de la media muestral se aproxime a una distribución normal, incluso si la población original no lo es.
17	F	A medida que aumenta el tamaño de la muestra, la distribución de la media muestral tiende a una distribución normal.
18	V	Esto es parte del teorema central del límite.
19	V	Incluso si no se toman todas las muestras posibles, la media de la distribución muestral de medias aún tiende a ser cercana a la media poblacional.
20	V	El Teorema del Límite Central es esencial en estadísticas al permitir inferencias sobre la media poblacional con métodos basados en normalidad, incluso si la población original no es normal.

Ir a la autoevaluación

Autoevaluación 10

Pregunta	Respuesta	Retroalimentación
1	V	La correlación mide el grado de asociación o relación entre dos variables, indicando cómo el cambio en una variable se relaciona con el cambio en otra.
2	F	Un coeficiente de correlación cercano a cero, indica una relación débil o nula entre las variables.
3	V	Es decir, cuando una variable aumenta, la otra disminuye en una proporción constante.
4	V	En una relación directa, un aumento en la variable independiente está asociado con un aumento en la variable dependiente, reflejado por un coeficiente de correlación positivo.
5	F	La variable dependiente, también conocida como variable respuesta, es la que se intenta predecir o explicar. La variable independiente, o predictor, es la que se utiliza para hacer la predicción.
6	V	La variable independiente, o explicativa, es la que se manipula o se utiliza para predecir o influir en la variable dependiente.
7	V	En R, el comando *cor* se utiliza para calcular el coeficiente de correlación entre dos variables, evaluando su asociación lineal.
8	F	El comando *pairs* en R crea una matriz de dispersión o gráficos de pares para visualizar las relaciones entre múltiples variables.
9	V	El intercepto de la recta de regresión es el valor de la variable dependiente (Y) cuando la variable independiente (X) es igual a cero.
10	V	Una vez que se ha determinado la ecuación de la recta de regresión, esta puede utilizarse para predecir valores de la variable dependiente en función de los valores de la variable independiente.

Ir a la autoevaluación

Autoevaluación 11

Pregunta	Respuesta	Retroalimentación
1	F	Las hipótesis pueden referirse tanto a estadísticas de la muestra como a parámetros de la población.
2	V	Verificar la hipótesis consiste en realizar pruebas estadísticas para evaluar la evidencia en contra de la hipótesis nula, no solo determinar si es razonable.
3	V	El nivel de significancia establece la probabilidad de cometer un error tipo I al rechazar incorrectamente la hipótesis nula.
4	V	Cuando la desviación estándar de la población es desconocida, se utiliza la estadística t en lugar de z.
5	V	La regla de decisión generalmente involucra comparar el valor calculado con los valores críticos, no un rango específico.
6	F	Tanto los valores z como los valores críticos pueden ser calculados o buscados en una tabla, dependiendo del contexto y de la información disponible.
7	F	Una prueba de una cola se enfoca en un solo extremo de la distribución.
8	V	En una prueba de dos colas, el nivel de significancia se divide entre las dos colas de la distribución.
9	F	En este caso, el operador debería ser <=, ya que la hipótesis nula generalmente establece la ausencia de cambio o incremento.
10	V	En este caso, la hipótesis nula afirmará que no hay diferencia, por lo que se utiliza el operador =.

Ir a la autoevaluación

Autoevaluación 12

Pregunta	Respuesta	Retroalimentación
1	V	La hipótesis establece afirmaciones sobre los parámetros de la población
2	V	En la verificación de hipótesis, se compara un estadístico calculado a partir de una muestra con un valor esperado bajo la hipótesis nula.
3	V	Cuando se conoce la desviación estándar de la población, se utiliza el estadístico Z en la verificación de hipótesis.
4	F	El nivel de significancia es la probabilidad de cometer un error tipo I, no necesariamente es igual al nivel de confianza.
5	V	Un error de tipo I está relacionado con el nivel de significancia, que es la probabilidad de rechazar la hipótesis nula cuando es verdadera.
6	F	El valor del estadístico de prueba se calcula a partir de la información de la muestra y la hipótesis nula.
7	F	Un error de tipo 2 ocurre cuando no se rechaza la hipótesis nula a pesar de que es falsa.
8	V	La regla de decisión establece los criterios para aceptar o rechazar la hipótesis nula basándose en el valor observado y los valores críticos.
9	F	El valor crítico se elige de tablas estadísticas o mediante métodos específicos según el enfoque utilizado.
10	V	La hipótesis alternativa (H1) refleja la dirección del cambio que se está buscando.

Ir a la autoevaluación

5. Referencias Bibliográficas

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

6. Anexos

Anexo 1. Práctica 1

Estadística para las ingenierías y arquitectura
Prof: Juan Carlos Torres
Práctica 1 (Semana 1)

Responda a cada una de las siguientes preguntas. El resolverlas le garantiza estar alcanzar los resultados de aprendizaje planteados para la primera semana de trabajo.

1. Revise el concepto de estadística que consta en el texto, luego, escriba al menos tres acciones que se incluyen en ese concepto de estadística	
Acción 1	
Acción 2	
Acción 3	

2. Escriba:	
3 variables ordinales	
3 variables nominales	
1 variable a nivel de intervalo	
3 variables a nivel de razón	

3. Escriba tres ejemplos de:	
Variables discretas	
Variables continuas	

4. Escriba un ejemplo de población y muestra.	
Población	
Muestra	

5. ¿En qué se diferencia la estadística descriptiva de la inferencial? Escriba al menos dos diferencias	
Diferencia 1	
Diferencia 2	

Anexo 2. Ejemplo de organización de conjuntos de datos

Se encuesta a los estudiantes de la modalidad a distancia de la UTPL (a 20 estudiantes) y se les hace las siguientes 4 preguntas:

1. Sexo: ____________________
2. Edad: ____________________
3. Gasto mensual en teléfono: ____________________
4. Marca del teléfono: ____________________

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
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

Nota. Torres, J., 2023.

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**

Nota. Torres, J., 2023.

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 organice 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
iPhone	3
Nokia	6
Samsung	3
Sony	4
Total	**20**

Nota. Torres, J., 2023.

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 cuántos teléfonos de cada marca hubo 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 vamos a estudiar a continuación.

Anexo 3. Ejercicios de distribución de frecuencias

Ejercicio 1

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

Tabla 1

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

Nota. Torres, J., 2023.

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:

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

$$Tamaño\ de\ clase = \frac{23}{6}$$

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 2

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**

Nota. Torres, J., 2023.

Importante: note que 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).

Ejercicio 2

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 formulan varias preguntas, una de las preguntas le consulta sobre el nivel de ingresos de la familia del 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 3

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
1560	1450	540	1480	582
278	1600	620	1510	593
450	1200	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
1200	720	375	840	498
1400	840	280	756	578
360	600	590	895	890
570	560	560	1000	1250
1100	760	840	467	1150
347	450	810	624	1080
180	580	814	1050	1400
150	990	640	300	1450

Nota. Torres, J., 2023.

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: 1600
   El rango o recorrido de la variable es = valor máximo – valor mínimo

El rango o recorrido de la variable es = 1600 – 150
El rango o recorrido de la variable es = 1450

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 Kapropiado es 7.

3. Tamaño de clase

La fórmula utilizada es:

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

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

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 4

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**

Nota. Torres, J., 2023.

En el texto básico existen muchos ejemplos del mismo tipo, mi recomendación es seguir los pasos que se han señalado aquí.

Anexo 4. Ejemplos de frecuencias relativas

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

Tabla 1

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**

Nota. Torres, J., 2023.

Vemos que la tabla tiene 1 clase, para obtener la frecuencia relativa de cada clase hacemos lo siguiente:

Tabla 2

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**

Nota. Torres, J., 2023.

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 3

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%**

Nota. Torres, J., 2023.

En la tabla anterior he 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 4

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**

Nota. Torres, J., 2023.

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

Tabla 5

Tabla de distribución de frecuencias y frecuencia relativa

Clase	Frecuencia	Frecuencia relativa
150 – 358	13	13 / 1000
359 – 567	22	22 / 1000
568 – 776	32	32 / 1000
777 – 985	14	14 / 100
986 – 1194	7	7 / 100
1195 – 1403	6	6 / 1000
1404–1602	6	6 / 100
Total	**100**	**1**

Nota. Torres, J., 2023.

El resultado es el siguiente:

Tabla 6

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**

Nota. Torres, J., 2023.

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 7

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			

Nota. Torres, J., 2023.

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 8

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		

Nota. Torres, J., 2023.

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 9

Tabla de frecuencias con frecuencia acumulada

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

Nota. Torres, J., 2023.

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 10

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**

Nota. Torres, J., 2023.

Anexo 5. Práctica 2

Estadística para las ingenierías y arquitectura

Prof.: Juan Carlos Torres

Práctica 2 (semana 2)

Responda a cada una de las siguientes preguntas. El resolverlas le garantiza estar alcanzar los resultados de aprendizaje planteados para la segunda semana de trabajo.

1. Dado el siguiente conjunto de datos: 12, 14, 16, 18, 20, 21, 22, 23.	
Determine el rango	
Determine K (número de clases)	
Ancho de clase	
Construya el histograma respectivo	

2. Dado el siguiente conjunto de datos: 5,6,7,9,10,15,17,18,6,9,11,10,19,16,18,16,19,10,11 Determine cuántas y cuáles son las clases se generan a partir del siguiente conjunto de datos:	
Enumere las clases resultantes. Ejemplo: Clase 1: 5 - 10 Clase 2: 10 - 15 Clase 3: 15 - 20	
Construya el histograma respectivo	

Anexo 6. Práctica 3

Estadística para las ingenierías y arquitectura

Prof.: Juan Carlos Torres

Práctica 3 (semana 3)

Responda a cada una de las siguientes preguntas. El resolver estas preguntas le garantiza estar alcanzar los resultados de aprendizaje planteados para la tercera semana de trabajo.

1. Dado el siguiente conjunto de datos: 5,6,7,9,10,5,7,8,6,9,11,10,9,6,8,6,9,10,11	
Calcule la media	
Calcule la mediana	
Calcule la moda	
Compare la media y mediana y responda. ¿cuál describe mejor a los datos?	

2. Dado el siguiente conjunto de datos: 17, 24,23, 21, 32, 25, 27, 24, 26, 32, 31, 30, 21, 20	
Calcule la media	
Calcule la mediana	
Calcule la moda	
Compare la media y mediana y responda. ¿cuál describe mejor a los datos?	

Anexo 7. Práctica 4

Estadística para las ingenierías y arquitectura

Prof.: Juan Carlos Torres

Práctica 4 (semana 4)

Responda a cada una de las siguientes preguntas. El resolverlas le garantiza estar alcanzar los resultados de aprendizaje planteados para la cuarta semana de trabajo.

1. Dado el siguiente conjunto de datos: 6,7,9,10,5,7,8,6,9,5,11,10,9,6,8,6,9,10,11	
Calcule el rango	
Calcule la desviación media	

2. Dado el siguiente conjunto de datos: 5,6,7,9,10,5,7,8,6,9,11,10,9,6,8,6,9,10,11	
Calcule la desviación estándar, realice un proceso manual conforme se indica en el material educativo. En el denominador de la fórmula que se emplea utilice n.	

3. Dado el siguiente conjunto de datos: 17, 24,23, 21, 32, 25, 27, 24, 26, 32, 31, 30, 21, 20	
Calcule la desviación estándar, realice un proceso manual conforme se indica en el material educativo. En el denominador de la fórmula que se emplea utilice n.	

4. Una compañía de telefonía celular tiene 500.000 clientes; ha tomado 10 a los que les ha calculado los minutos hablados en el último mes, esos datos son los siguientes: 40, 48, 36, 54, 60, 45, 50, 39, 48, 45	
Calcule la desviación estándar, realice un proceso manual conforme se indica en el material educativo. Tome en cuenta que se trata de una muestra por lo que debe utilizar en el denominador de la fórmula (n - 1)	

Anexo 8. Práctica 5

Estadística para las ingenierías y arquitectura

Prof.: Juan Carlos Torres

Práctica 5 (semana 5)

Responda a cada una de las siguientes preguntas. El resolver estas preguntas le garantiza estar alcanzar los resultados de aprendizaje planteados para la quinta semana de trabajo.

1. Dado el siguiente conjunto de valores: 21, 22, 23, 27, 31, 36	
construya un diagrama de tallo y hojas:	

2. Dado el siguiente conjunto de datos: 6,6,7,8,10,11,11,13,14,14,15,16,21,24,27,28,29,30,30	
Calcule el primer cuartil	
Calcule el segundo cuartil	
Calcule el tercer quintil	
Calcule el cuarto quintil	

Anexo 9. 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.

1.1. ¿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.

1.2. 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 1.
Clasificación de probabilidad

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 en base a eso este asigna un valor de probabilidad.

1.3. 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.

1.3.1. 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.

Figura 2.
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**

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.

1.3.2. 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)

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

P (A y B) = 0,2272

La probabilidad alcanza un 22,72%.

1.4. 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, saque 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 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.

Anexo 10. Práctica 6

Estadística para las ingenierías y arquitectura

Prof.: Juan Carlos Torres

Práctica 6 (semana 6)

Responda a cada una de las siguientes preguntas. El resolver estas preguntas le garantiza estar alcanzar los resultados de aprendizaje planteados para la sexta semana de trabajo.

1. En una rifa organizada por una agrupación social se van a rifar 3 premios uno a continuación de otro. Se ha logrado vender 100 boletos y el organizador compra uno.	
Calcule la probabilidad de que el organizador se gane uno de los premios.	
Determine la probabilidad de que el organizador gane el primero o el segundo premio	
Determine la probabilidad de que el organizador gane el primero y el segundo premio	

2. En una bolsa hay 10 bolitas de dos colores, cinco de cada color. Si se saca una y se la retorna a la bolsa antes de sacar otra. ¿Cuál es la probabilidad de que se tenga una de un color y otra de otro?

3. En una caja existen 2 bolas blancas y 2 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?

Anexo 11. Práctica 7

Estadística para las ingenierías y arquitectura

Prof.: Juan Carlos Torres

Práctica 1 (semana 9)

Responda a cada una de las siguientes preguntas. El resolver estas preguntas le garantiza estar alcanzar los resultados de aprendizaje planteados para la primera semana de trabajo del segundo bimestre.

1. Consulte la información relacionada con el ingreso semanal de los supervisores de turno en la industria del vidrio. La distribución de los ingresos semanales tiene una distribución de probabilidad normal, con una media de \$ 400 y una desviación estándar de \$40. ¿Cuál es la probabilidad de seleccionar a un supervisor de turno de la industria del vidrio cuyo ingreso:	
a) oscile entre \$300 y \$350?	
b) Que oscile entre 360 y \$450?	
c) Que oscile entre 425 Y 475	
d) Que sea mayor que 500	

2. Un estudio reciente con respecto a salarios de funcionarios de empresas privadas de la provincia del Guayas demostró que el salario medio mensual era de $575, con una desviación estándar de $30. Suponga que la distribución de los salarios es una distribución de probabilidad normal. Si elige un empelado privado al azar, ¿cuál es la probabilidad de que su salario sea:	
a) menor a 500	
b) mayor a 600	
c) Que oscile entre 450 y 550	
d) Que oscile entre 400 y 470	

Anexo 12. Práctica 8

Estadística para las ingenierías y arquitectura

Prof.: Juan Carlos Torres

Práctica 2 (semana 10)

Responda a cada una de las siguientes preguntas. El resolver estas preguntas le garantiza estar alcanzar los resultados de aprendizaje planteados para la segunda semana de trabajo del segundo bimestre.

1. Plantee un ejemplo de población y muestra. Debe tratarse de una situación real.

2. Plantee un ejemplo de muestreo aleatorio simple

3. Plantee un ejemplo de muestreo por conglomerados

4. Plantee un ejemplo de muestreo sistemático

5. El muestreo es recomendable por varias razones, una de ellas es la “naturaleza destructiva” de las pruebas de determinado producto. Escriba un ejemplo de este tipo. Para responder a esta pregunta revise el texto básico y de lectura a las “Razones para muestrear”

6. Revise la documentación del tema y escriba según sus propias palabras a que se llama error de muestreo.

Anexo 13. Ejercicios de Teorema central del límite

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 1

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

Nota. Torres, J., 2023.

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

Figura 1

Distribución de la población

Nota. Torres, J., 2023.

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 2

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

Nota. Torres, J., 2023.

d. Ahora obtenemos la media de las muestras:

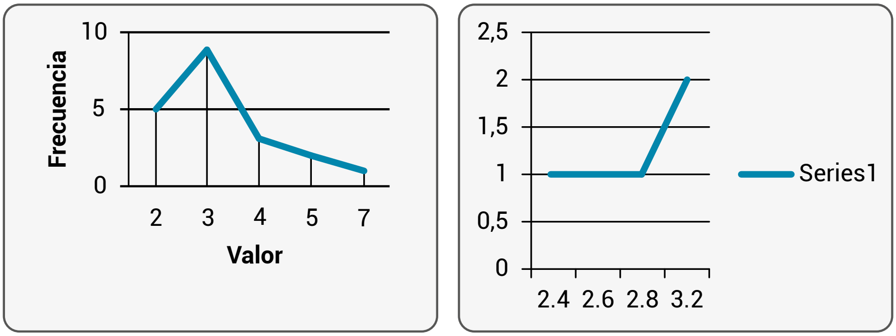
(2.4 + 3.2 + 2.8 + 3.2 + 2.6)/5 = 2.84

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 2

Comparación de gráficas



Nota. Torres, J., 2023.

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

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 3

Diez muestras de tamaño 5

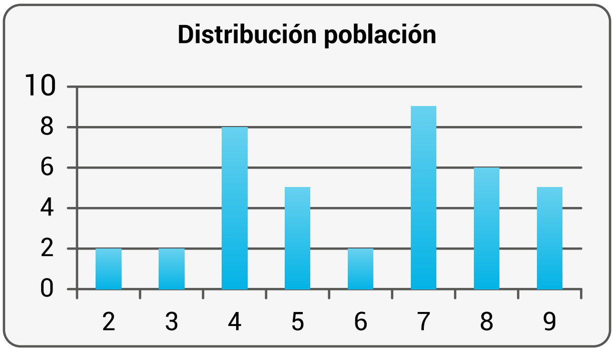
0	2	7	1	1
9	4	8	7	3
5	4	9	2	1
7	7	6	4	0
6	1	5	4	5
1	7	1	4	7
1	3	7	4	8
8	7	4	5	5
0	8	9	9	9
7	8	8	0	4

Nota. Torres, J., 2023.

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

Figura 3

Distribución de la población



Nota. Torres, J., 2023.

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

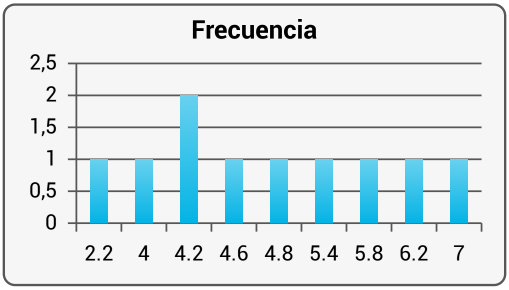
En cuanto al segundo literal tenemos la tabla siguiente cuya última columna está la media de cada muestra (media muestral), en esa tabla se puede observar la segunda columna utilizada para dibujar el histograma. Note que en el histograma se puede apreciar ya una forma normal.

Figura 4

Frecuencia de las medias muestrales

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



Nota. Torres, J., 2023.

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.

Anexo 14. Práctica 9

Estadística para las ingenierías y arquitectura

Prof.: Juan Carlos Torres

Práctica 3 (semana 11)

Responda a cada una de las siguientes preguntas. El resolver estas preguntas le garantiza estar alcanzar los resultados de aprendizaje planteados para la tercera semana de trabajo del segundo bimestre.

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 misma ciudad la desviación estándar fue 12 ¿Cuántas personas debe abarcar la muestra? La fórmula señala: n=((z.δ)/E)2 Revise los ejemplos de la guía didáctica (documento 7) para apoyarse en la resolución.

Suponga que el alcalde de su ciudad desea un cálculo de la proporción de la población que apoya su gestión. El alcalde quiere que el cálculo se encuentre a menos de 0.05 de la proporción real. Suponga un nivel de confianza de 95%. Los asesores políticos del alcalde calculan que la proporción que apoya la actual política es de 0.55.

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

Revise los ejemplos de la guía didáctica (documento 7) para apoyarse en la resolución.

Anexo 15. Práctica 10

Estadística para las ingenierías y arquitectura

Prof.: Juan Carlos Torres

Práctica 4 (semana 11)

Responda a cada una de las siguientes preguntas. El resolver estas preguntas le garantiza estar alcanzar los resultados de aprendizaje planteados para la cuarta semana de trabajo del segundo bimestre.

1. Dado el siguiente conjunto de datos: 2, 2, 3, 4, 6, 7, 8. Encuentre la media muestral utilizando todas las muestras posibles de tamaño 2. Compárela con la media aritmética de (2, 2, 3, 4, 6, 7, 8). ¿Existe diferencia? ¿Por qué?

2. Describa el teorema del límite central

Anexo 16. Práctica 11

Estadística para las ingenierías y arquitectura

Prof.: Juan Carlos Torres

Práctica 5 (semana 13)

Responda a cada una de las siguientes preguntas. El resolver estas preguntas le garantiza estar alcanzar los resultados de aprendizaje planteados para la quinta semana de trabajo del segundo bimestre.

1. Plantear las hipótesis nula y alternativa para el siguiente ejemplo: La cantidad promedio de sillas que fabrica una planta es de 150 diarias con una desviación estándar de 3. Se compró nueva maquinaria y se midió la cantidad promedio de sillas que se producen ahora en un lapso de 60 días, esa media alcanzó 150.5. ¿Se puede decir que la cantidad producida actualmente es mayor; es decir la compra de maquinaria permitió incrementar la producción?

2. Plantear las hipótesis nula y alternativa para el siguiente ejemplo: Una fábrica de bandas de carro tiene un producto (una banda) con una resistencia promedio de 1000 kilos, la desviación estándar es de 15 kilos. Se están cambiando los materiales con los que se fabrican dichas bandas y se hacen mediciones periódicas. La última medición de 100 bandas arrojó una resistencia promedio de 1010 kilos. ¿Se puede concluir que ha existido un cambio en los valores de resistencia de las bandas?

3. Para el siguiente ejemplo determine el paso 4 en la verificación de hipótesis, escriba cual será la regla de decisión, apóyese de un gráfico: La cantidad promedio de sillas que fabrica una planta es de 150 diarias con una desviación estándar de 3. Se compró nueva maquinaria y se midió la cantidad promedio de sillas que se producen ahora en un lapso de 60 días, esa media alcanzó 150.5. Se recomienda trabajar con un nivel de significación de 0.05. ¿Se puede decir que la cantidad producida actualmente es mayor; es decir la compra de maquinaria permitió incrementar la producción?

4. Para el siguiente ejemplo determine el paso 4 en la verificación de hipótesis, escriba cual será la regla de decisión, apóyese de un gráfico: Una fábrica de bandas de carro tiene un producto (una banda) con una resistencia promedio de 1000 kilos, la desviación estándar es de 15 horas. Se están cambiando los materiales con los que se fabrican dichas bandas y se hacen mediciones periódicas. La última medición de 100 bandas arrojó una resistencia promedio de 1010 kilos. Se recomienda trabajar con un nivel de significación de 0.10. ¿Se puede concluir que ha existido un cambio en los valores de resistencia de las bandas?

Anexo 16. Práctica 17

Estadística para las ingenierías y arquitectura

Prof.: Juan Carlos Torres

Práctica 6 (semana 14)

Responda a cada una de las siguientes preguntas. El resolver estas preguntas le garantiza estar alcanzar los resultados de aprendizaje planteados para la sexta semana de trabajo.

<table>
<tr><td colspan="2">1. El gerente de ventas de una librería de textos universitarios, afirma que sus vendedores realizan en promedio 40 llamadas de ventas a la semana a profesores con una desviación estándar de 3 ventas. Varios representantes señalan que el cálculo es muy bajo. Una muestra aleatoria de 28 representantes de ventas revela que la cantidad media de llamadas que se realizó la semana pasada fue de 42.

Con el nivel de significancia de 0.05, ¿puede concluir que la cantidad media de llamadas semanales ha cambiado?</td></tr>
<tr><td>Hipótesis nula y alternativa</td><td></td></tr>
<tr><td>Nivel de significancia con el que trabajará</td><td></td></tr>
<tr><td>Estadístico que utilizará</td><td></td></tr>
<tr><td>Gráfica del problema y regla de decisión</td><td></td></tr>
</table>

Resultado y conclusión. Respuesta a la pregunta que plantea el problema	

2. La vida media de una batería de un reloj digital es de 305 días. Las vidas medias de las baterías se rigen por la distribución normal. Hace poco se modificó la batería para que tuviera mayor duración. Una muestra de 20 baterías modificadas exhibió una vida media de 311 días con una desviación estándar de 12 días. ¿La modificación incrementó la vida media de la batería? Nivel de significancia de 0.05	
Hipótesis nula y alternativa	
Nivel de significancia con el que trabajará	
Estadístico que utilizará	
Gráfica del problema y regla de decisión	
Resultado y conclusión. Respuesta a la pregunta que plantea el problema	