UTPL

La Universidad Católica de Loja

Vicerrectorado de Modalidad Abierta y a Distancia

Estadística Descriptiva para Agronegocios

Guía didáctica

UTPL

Facultad Ciencias Exactas y Naturales

Estadística Descriptiva para Agronegocios

Guía didáctica

Carrera	PAO Nivel
Agronegocios	II

Autora:

Ximena Yadira González Rentería

ESTA_1050

UTPL

Universidad Técnica Particular de Loja

Estadística Descriptiva para Agronegocios

Guía didáctica

Ximena Yadira González Rentería

Diagramación y diseño digital

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

ISBN digital -978-9942-47-281-6

Año de edición: abril, 2025

Edición: primera edición

El autor de esta obra ha utilizado la inteligencia artificial como una herramienta complementaria. La creatividad, el criterio y la visión del autor se han mantenido intactos a lo largo de todo el proceso.

Loja-Ecuador

Índice

1. Datos de información

1.1 Presentación de la asignatura

1.2 Competencias genéricas de la UTPL

- Orientación a la innovación y a la investigación.
- Pensamiento crítico y reflexivo.

1.3 Competencias del perfil profesional

Implementar, generar e innovar procesos administrativos, económicos, tecnológicos y de producción, fundamentados en herramientas de investigación, generación, gestión y evaluación de proyectos agroproductivos en el ámbito de la cadena de valor, de manera que se fortalezca la productividad y rentabilidad de las empresas y sus productos, se mejore el posicionamiento en mercados nacionales e internacionales y se disminuya los riesgos de las organizaciones del sector agroalimentario.

1.4 Problemática que aborda la asignatura

La asignatura aborda la problemática de la gestión administrativa agroproductiva, en donde se incluyen conocimientos relacionados al uso eficiente de los recursos de la empresa (financieros, técnicos, humanos y materiales), procurando la calidad en los procesos de producción y prestación de servicios, logrando la integración de todas las áreas de la empresa, un adecuado servicio al cliente y la obtención de beneficios para la empresa; mediante la incorporación de sistemas automatizados de gestión administrativa. Este núcleo se desarrolla dentro de los campos del conocimiento de las ciencias económicas, ciencias jurídicas y derecho; para este efecto se analiza los contenidos desde los campos del conocimiento de: estadística descriptiva para agronegocios, administración financiera, administración de recursos humanos, administración financiera operativa y estructural, fundamentos de contabilidad, contabilidad financiera, derecho laboral, legislación mercantil y societaria, sistemas de producción vegetal, sistemas de producción animal, sistemas de producción acuícola, sistemas agroindustriales, seguridad y soberanía alimentaria, sistemas de información geográfica, cadenas de valor agro productivas; complementando con las asignaturas del itinerario planificación empresarial.

2. Metodología de aprendizaje

Para aportar a la consecución de los resultados de aprendizaje, en esta asignatura se aplicará la metodología de **aprendizaje basado en el pensamiento** (Thinking Based Learning, TBL). Este enfoque educativo busca desarrollar en los estudiantes habilidades de pensamiento crítico, creativo y metacognitivo para que puedan comprender profundamente los contenidos, resolver problemas de manera efectiva y tomar decisiones fundamentadas. El TBL a través de actividades estructuradas y el uso de herramientas como mapas mentales, rutinas de pensamiento y preguntas guiadas, promueve en los estudiantes un aprendizaje significativo que trasciende la memorización y les permite aplicar sus conocimientos de manera innovadora y autónoma en situaciones reales.

Así mismo, en esta asignatura se promoverá el **aprendizaje autónomo**, para lograr que los estudiantes puedan dirigir su propio proceso de aprendizaje de manera activa, reflexiva y responsable. Este tipo de aprendizaje fomenta habilidades como la autogestión, la autorregulación y el pensamiento crítico, permitiendo a los estudiantes identificar sus propias necesidades, intereses y ritmos de aprendizaje.

3. Orientaciones didácticas por resultados de aprendizaje

Primer bimestre

Resultado de aprendizaje 1:

Reconoce y caracteriza los distintos tipos de variables estadísticas.

Este resultado de aprendizaje se logrará a través de la Unidad 1. Introducción a la Estadística, misma que proporcionará los fundamentos necesarios para comprender la importancia y el uso de la estadística en el contexto de los agronegocios, reconociéndola como una herramienta esencial para analizar datos relacionados con la producción, comercialización y toma de decisiones en el sector agropecuario.

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

Durante esta primera semana, nos enfocaremos en el estudio de los *conceptos introductorios* de estadística. Analizaremos ¿qué es la estadística?, su relevancia en el análisis y organización de datos, así como su papel en la toma de decisiones informadas.

¡Empezamos!

1.1. Conceptos introductorios.

Bienvenido a la primera semana de la asignatura de Estadística descriptiva para Agronegocios. En esta etapa inicial, exploraremos algunos conceptos introductorios, así como la aplicabilidad de la estadística en diversas áreas de la vida cotidiana.

La **Estadística** es la rama de las matemáticas que se encarga de recolectar, organizar, analizar, interpretar y presentar datos, con el objetivo de extraer conclusiones útiles y tomar decisiones fundamentadas, con base en la información obtenida (Agrawal y Gopal 2013).

Probablemente usted se haya preguntado: ¿Para qué necesito la estadística en mi negocio? La respuesta es sencilla: un emprendedor exitoso debe ser capaz de comprender y utilizar la información de manera eficiente para tomar decisiones acertadas.

La estadística, como ciencia exacta y transversal, desempeña un papel fundamental, su transversalidad permite que, profesionales de diversos campos relacionados con los negocios (contadores, analistas financieros, especialistas en marketing, investigadores de mercado y economistas) la utilicen de forma directa o indirecta para organizar datos, analizarlos, interpretar resultados y tomar decisiones estratégicas.

Es muy importante que tenga en cuenta que, hay dos tipos de estadística: la estadística descriptiva y la inferencial. De momento, nos enfocaremos en la primera.

¿A qué se refiere la estadística descriptiva? Esta estadística se encarga de analizar, interpretar, resumir y presentar los resultados relacionados con un conjunto de datos, derivados de una muestra o de toda la población (Mendenhall et al., 2015).

En este concepto, surgen ahora unos términos nuevos. Usted se preguntará, *¿qué es una* **población?** *¿a qué se refiere una* **muestra?** Para aclarar sus dudas recuerde lo siguiente:

Una **población** se refiere al conjunto completo de individuos, objetos o eventos que comparten una característica en común y que son el objeto de estudio, mientras que, una **muestra** es un subconjunto representativo de la población, seleccionado para realizar análisis estadísticos (Itza Ortiz et al., 2024).

Comprendido esto, es muy importante tener en cuenta que, la muestra permite estudiar características de la población de manera más eficiente, ya que analizar toda la población puede, en muchos casos, ser bastante costoso o impracticable.

Otros términos, también son fundamentales tenerlos claros. A continuación, se presentan varios de éstos que permitirán clarificar el estudio de esta asignatura:

Las **variables** son características o atributos de los elementos de estudio que pueden tomar diferentes valores (Martínez, 2020). Una **unidad experimental** es el individuo u objeto en el que se mide una variable (Itza Ortiz et al., 2024).

Es importante considerar que, cuando se mide una sola variable en una sola unidad experimental se obtienen datos **univariados,** mientras que, cuando se miden dos variables o más, en una sola unidad experimental, los datos resultantes serán **bivariados** o **multivariados,** respectivamente (Mendenhall et al., 2015).

Semana 2

En esta semana continuamos con la Unidad 1: Introducción a la Estadística, enfocándonos en el tema tipos de variables. Se aprenderá a identificar y clasificar las variables en cualitativas y cuantitativas, comprendiendo su importancia para seleccionar métodos adecuados de análisis y aplicar correctamente herramientas estadísticas.

Unidad 1. Introducción a la estadística.

1.2. Tipos de variables

Teniendo claro qué es una variable, se puede mencionar que éstas se clasifican en dos grandes tipos: variables cualitativas y variables cuantitativas, dependiendo de si representan características categóricas o numéricas (Itza Ortiz et al., 2024).

¿Qué son las variables cualitativas?

O también llamadas variables categóricas, éstas describen características o atributos que no se miden numéricamente, sino que se clasifican en categorías. Estas categorías pueden ser **nominales** (sin un orden específico) u ordinales (con un orden lógico) (Itza Ortiz et al., 2024).

Por ejemplo, en el contexto de los agronegocios, por un lado, la **variedad de cultivos** (maíz, trigo, arroz, entre otros) es una **variable cualitativa nominal**, ya que las categorías no tienen un orden inherente. Por otro lado, la **calidad del producto**, clasificada como: baja, media o alta es una **variable cualitativa ordinal**, ya que existe un orden lógico en las categorías.

¿Qué son las variables cuantitativas?

Estas variables se expresan mediante números y representan cantidades medibles. Pueden ser de dos tipos: **discretas,** cuando toman valores enteros o finitos, o pueden ser del tipo **continuas,** cuando los valores tienen un rango infinito de números reales (Itza Ortiz et al., 2024, Martínez, 2020).

Por ejemplo, el número de hectáreas cultivadas en una finca es una **variable cuantitativa discreta**, ya que solo puede tomar valores enteros (1, 2, 3 hectáreas). Por otro lado, el peso promedio de los frutos cosechados es una **variable cuantitativa continua**, ya que puede tomar cualquier valor dentro de un rango, como 1.5 kg, 2.3 kg, etc.

Para reforzar este tema, lo invito a revisar la siguiente infografía Tipos de variables. En éste, podrá encontrar la clasificación de las variables y algunos ejemplos para poder reforzar sus conocimientos.

Tipos de variables

¿Qué le ha parecido esta síntesis de los tipos de variables? ¿Le ha ayudado a dejar la información clara? Tenga en cuenta estimado estudiante que conocer los tipos de variables es esencial para posteriormente elegir el análisis estadístico adecuado.

Actividades de aprendizaje recomendadas

Apreciado estudiante, para asegurar que ha comprendido estos contenidos, se propone realizar las siguientes actividades:

1. Revisar sobre los Tipos de variables según la escala, para reforzar los contenidos de la semana.
2. Realizar un cuadro sinóptico con los tipos de variables, en donde se propongan ejemplos de cada uno de éstos.

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

Las actividades propuestas tienen como finalidad afianzar los conocimientos que se han adquirido durante la semana. Recuerde que, las variables según la escala de medición pueden ser: nominales, ordinales, de intervalo y de razón. Las variables nominales agrupan datos en categorías sin un orden específico (ejemplo: el tipo de cultivo), mientras que las ordinales establecen un orden, pero sin distancias iguales entre categorías (ejemplo: la calidad del suelo). Por otro lado, las variables de intervalo tienen un orden y distancias iguales, pero carecen de un cero absoluto (ejemplo: la temperatura en °C) y las variables de razón poseen todas las propiedades anteriores y un cero absoluto.

Apreciado estudiante, hemos concluido el estudio de la Unidad 1 y es momento de evaluar su aprendizaje. Para esto, responda a las preguntas que se presentan a continuación. El cuestionario contiene preguntas de opción múltiple, *con una o varias opciones de respuestas correctas*. Esta actividad le servirá de repaso, así como para comprobar los conocimientos adquiridos.

Autoevaluación 1

Seleccione la/las respuesta(s) correcta(s):

1. ¿Qué se entiende por población?

   a. Una lista de valores organizados.
   b. El conjunto total de individuos u objetos de interés en un estudio.
   c. Un grupo seleccionado de elementos de una investigación.
   d. Una tabla con datos específicos.

2. ¿Qué característica principal define a una muestra?

   a. Es un conjunto completo de datos.
   b. Es una tabla organizada de valores.
   c. Es un subconjunto representativo de la población.
   d. Es cualquier grupo de elementos seleccionados al azar.

3. Un investigador quiere medir la productividad de todos los agricultores de una región, pero selecciona 50 para su análisis. ¿Qué representa este grupo de 50 agricultores?

   a. La población.
   b. La muestra.
   c. Las variables.
   d. Los datos cualitativos.

4. ¿Qué tipo de variable es la "variedad de cultivos" en una finca?

   a. Cuantitativa continua.
   b. Cuantitativa discreta.
   c. Cualitativa nominal.
   d. Cualitativa ordinal.

5. ¿Cuáles son ejemplos de variables cualitativas?

   a. La variedad de cultivos en una finca.
   b. El número de hectáreas cultivadas.
   c. El color de los frutos cosechados.
   d. El peso promedio de los sacos de arroz.

6. ¿Cuáles de las siguientes variables son cuantitativas continuas?

   a. El número de árboles en una hectárea.
   b. El rendimiento promedio de un cultivo en toneladas por hectárea.
   c. La clasificación de la calidad del suelo (buena, regular, mala).
   d. La altura de las plantas en un cultivo.

7. ¿Cuál de las siguientes variables es cuantitativa continua?

   a. El número de trabajadores en una finca.
   b. La cantidad de tractores en una cooperativa.
   c. La distancia recorrida por un tractor (en kilómetros).
   d. La calidad del suelo está clasificada como buena, regular o mala.

8. ¿Qué tipo de variable es el precio promedio de venta de un kilogramo de maíz?

   a. Cuantitativa continua.
   b. Cuantitativa discreta.
   c. Cualitativa ordinal.
   d. Cualitativa nominal.

9. Una empresa exportadora de banano desea analizar el peso promedio de todos los frutos cosechados en una semana. ¿Qué representa este análisis?

   a. Una variable cualitativa nominal.
   b. Un análisis sobre la población.
   c. Un análisis sobre una muestra.
   d. Un análisis sobre variables cualitativas.

10. Un empresario desea analizar el consumo mensual de agua en litros de 100 fincas agropecuarias. ¿Qué elementos corresponden a una población y a una muestra?

   a. El consumo de agua de todas las fincas de la región.
   b. Las 100 fincas seleccionadas para el análisis.
   c. El consumo de agua promedio de esas 100 fincas.
   d. Los datos individuales de cada finca de la región.

Ir al solucionario

Resultado de aprendizaje 2:

Construye e interpreta tablas y gráficas estadísticas.

Continuando con el estudio de esta asignatura, el presente resultado de aprendizaje se alcanzará mediante la Unidad 2: Exploración de datos, que introduce herramientas fundamentales para organizar y visualizar información de manera clara y efectiva. Los contenidos que analizaremos a continuación, están diseñados con el objetivo de que se comprenda cómo las tablas y gráficos facilitan el análisis de datos en contextos reales, con un particular énfasis en el campo de los agronegocios.

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

Durante esta semana, iniciaremos el estudio de la Unidad 2: Exploración de Datos, en donde se analizará el tema de *tabla de frecuencias*. Aprenderemos a organizar y resumir datos en tablas que faciliten su interpretación, destacando su utilidad para identificar patrones, tendencias y distribuciones en conjuntos de datos.

Unidad 2. Exploración de datos

2.1. Tabla de frecuencias

La tabla de frecuencias (o distribución de frecuencias) es una tabla que organiza y resume un conjunto de datos, mostrando cuántas veces *(frecuencia)* ocurren los diferentes valores o categorías en un conjunto de datos (Itza Ortiz et al., 2024). Estas tablas son fundamentales para analizar y visualizar patrones en los datos, de forma clara y estructurada, permitiendo identificar tendencias, variabilidad y distribuciones. Pueden utilizarse tanto para variables cuantitativas, como para variables cualitativas ordinales.

En concreto, estas tablas tienen los siguientes beneficios:

1. Facilitan la organización de datos (especialmente cuando el conjunto de datos es grande).
2. Ayudan a identificar patrones y tendencias en las observaciones.
3. Sirven como base para construir gráficos (histogramas o polígonos de frecuencia), entre otros aspectos importantes.

Seguramente en este momento, usted se preguntará, *¿En qué consiste una tabla de frecuencias?,¿cómo se construye?* Pues, ésta consiste en columnas que muestran las distintas categorías o valores de una variable, junto a las cuales se encuentra el número de veces que ocurren en un conjunto de datos. Generalmente, incluye las siguientes columnas:

- **Intervalos de datos o categorías:** Los diferentes valores o rangos de datos.
- **Frecuencia absoluta**: El número de veces que ocurre cada valor o categoría.
- **Frecuencia relativa**: La proporción o porcentaje de la frecuencia absoluta en relación al total.
- **Frecuencia acumulada:** La suma acumulativa de las frecuencias absolutas o relativas.

Para conocer más sobre la frecuencia absoluta, relativa y acumulada, y saber cómo se calculan, lo invito a revisar el siguiente documento, concretamente al Capítulo 4. Distribución de frecuencias, del texto de Posada Hernández (2016), páginas de la 45 a la 49.

¿Qué le pareció la lectura?, ¿le queda ahora más claro cómo se calculan estas frecuencias? Seguro esto lo va a ir fortaleciendo más adelante, cuando se vaya ejercitando en la construcción de estas tablas. Vamos a continuar.

Es importante mencionar que, la tabla de frecuencias puede construirse tanto con **datos no agrupados**, como con **datos agrupados**. Para el primer caso, se procede en la primera columna a ordenar de menor a mayor los diferentes valores que tiene la variable en el conjunto de datos, posterior a lo cual, se pone en las siguientes columnas la frecuencia absoluta, así como el cálculo de las frecuencias relativa y acumulada. Por otro lado, para las tablas de frecuencia para datos agrupados, que suelen emplearse cuando se maneja un número alto de datos, el proceso es diferente.

Para comprender cómo se construye este tipo de tabla, revise la infografía a continuación.

Tabla de frecuencias para datos agrupados

¿Ha quedado claro cómo se construye la tabla de frecuencias para datos agrupados? Este tipo de tablas es muy útil, por lo que se sugiere que, en caso de que alguno de los pasos no haya quedado claro, vuelva a revisar esta infografía, de manera que posterior a esto, usted esté en la capacidad de elaborar una tabla de este tipo.

Actividades de aprendizaje recomendadas

Apreciado estudiante, para asegurar que ha comprendido este contenido referente a las Tablas de frecuencias, se propone realizar las siguientes actividades:

1. Revisar la Tabla de frecuencias, haciendo una lectura comprensiva.

   ¿Qué le pareció la lectura? Recuerde que, la tabla de frecuencias es una herramienta utilizada para organizar y resumir un conjunto de datos, en la que se muestra la cantidad de veces que cada valor o rango de valores aparece en una muestra. Es muy útil que durante la lectura realice alguno de los ejemplos propuestos, de manera que pueda poner en práctica los conocimientos adquiridos.

2. Realizar el siguiente ejercicio:

Un agricultor ha recolectado datos sobre el peso (en kilogramos) de 30 sacos de arroz cosechados en su finca. Los pesos son los siguientes:

Pesos (kg):

45, 50, 52, 47, 55, 50, 48, 49, 51, 54, 47, 50, 53, 49, 48, 46, 51, 50, 52, 48, 45, 47, 51, 54, 50, 48, 49, 53, 46, 52.

A partir de estos datos, construya la tabla de frecuencias

Tabla de frecuencias

INTERVALOS	Frecuencia absoluta	Frecuencia relativa	Frecuencia acumulada

Nota: Copie la tabla en un Word o cuaderno para rellenar.

Recuerde que, para realizar una tabla de datos agrupados, cuenta con un recurso de aprendizaje para su apoyo. Si es necesario, vuelva a revisarlo. Aquí se indica el paso a paso, de manera que pueda realizar el ejercicio sin ningún inconveniente.

Contenidos, recursos y actividades de aprendizaje recomendadas

Semana 4

Estimados estudiantes, durante esta semana abordaremos el tema de *representación gráfica de variables categóricas*, en donde analizaremos cómo visualizar los datos cualitativos, mediante gráficas como barras y pasteles, lo que facilitará la interpretación de frecuencias y proporciones. ¡Vamos a iniciar!

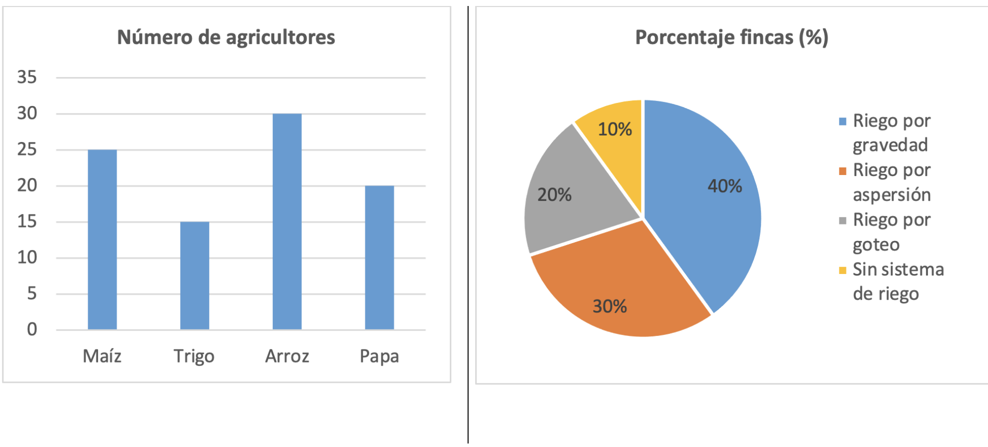
Unidad 2. Exploración de datos

2.2. Representación gráfica de variables categóricas

Continuando con el estudio de esta Unidad, destinaremos nuestros esfuerzos a revisar las gráficas para variables categóricas. Estas gráficas son representaciones visuales utilizadas para mostrar y analizar datos que están organizados en categorías, como tipos, grupos o clases (Itza Ortiz et al., 2024). Estas gráficas son esenciales porque permiten observar la distribución de las frecuencias de manera intuitiva, facilitando la comparación entre categorías. Entre las más usadas están: **las gráficas de barras y las gráficas de pastel.**

Figura 1

Ejemplos de gráficas



Nota. González, X., 2025.

A continuación, se explica cada una de éstas.

- **Gráficas de barras**: Son diagramas que usan rectángulos (barras) para representar la frecuencia, proporción o porcentaje de cada categoría. Las barras pueden ser horizontales o verticales y la longitud de cada barra es proporcional al valor que representa.

  Entre las principales características de las gráficas de barras se puede mencionar que:

a. Cada barra representa una categoría
b. Hay un espacio uniforme entre las barras para resaltar que las categorías son independientes
c. Se utilizan cuando es necesario comparar frecuencias o proporciones entre distintas categorías.

- **Gráficas de pastel:** Estas gráficas dividen un círculo en porciones que representan las proporciones o porcentajes de cada categoría en el total. Cada segmento del pastel es proporcional a la frecuencia relativa de la categoría que representa (Martínez, 2020).

  Entre las principales características de las gráficas de pastel están que:

a. Representan porcentajes o proporciones del total (100%)
b. Son ideales para mostrar la participación de cada categoría en el conjunto completo
c. No se recomiendan cuando hay muchas categorías o cuando las diferencias entre ellas son pequeñas.

Para reforzar este tema, lo invito a revisar el siguiente documento, concretamente el tema de Gráficos para datos categóricos, del texto de Obando-Bastidas y Castellanos Sánchez (2021), páginas de la 12 a la 15.

¿Qué le ha parecido la lectura? Es muy importante que, tenga claro que las gráficas de barras son ideales para realizar comparaciones directas entre categorías, mientras que las gráficas de pastel destacan la relación de cada categoría con el total.

Vamos a poner en práctica los conocimientos adquiridos. Lo invito a interactuar con el siguiente juego de arrastrar y soltar:

[Gráficas de barras vs gráficas de pastel]

¿Acertó en todas sus respuestas? Si fue así, ¡buen trabajo! Si en alguna respuesta falló, es momento de dar un nuevo vistazo a la información que hemos revisado anteriormente.

Contenidos, recursos y actividades de aprendizaje recomendadas

Semana 5

Para concluir con el estudio de la Unidad 2, nos enfocaremos durante la presente semana en el tema de *representación gráfica de variables numéricas*. Bajo este contexto, analizaremos cómo utilizar gráficas de barras, pasteles, gráficas de líneas e histogramas, para visualizar distribuciones, identificar tendencias y analizar datos cuantitativos.

Unidad 2. Exploración de datos

2.3. Representación gráfica de variables numéricas.

Estas gráficas son representaciones visuales diseñadas para analizar datos cuantitativos, es decir, aquellos que se miden y expresan numéricamente.

¿Cuál es su funcionalidad? Éstas nos ayudan a identificar patrones, tendencias, y distribuciones, facilitando el análisis estadístico y la interpretación de datos en diversos contextos. A continuación, se explicarán las gráficas más utilizadas para representar variables numéricas, entre las que destacan: **gráficas de barras, gráficas de pastel, gráficas de líneas** e **histogramas.**

- **Gráficas de barras:** Si bien estas gráficas suelen asociarse con variables categóricas, también pueden utilizarse para representar **variables numéricas discretas**, donde las barras muestran la frecuencia o valor asociado a cada número (De la Puente Viedma, 2018).

Un ejemplo que podría ser representado con una gráfica de barras es: El número de hectáreas cultivadas en distintas fincas. Cada barra representa una finca y su altura corresponde a las hectáreas cultivadas.

- **Gráficas de pastel:** Estas gráficas pueden también representar variables numéricas, esto puede darse cuando se desea mostrar la proporción que cada valor o rango tiene respecto al total

Un ejemplo que podría ser representado con una gráfica de pastel es: la distribución porcentual de ingresos generados por diferentes cultivos en una finca: maíz (40%), arroz (35%), papa (25%).

- **Gráficas de líneas:** Son ideales para analizar el comportamiento de datos numéricos continuos a lo largo de un intervalo o período, mostrando tendencias o cambios (Martínez, 2020).

Entre las principales características de las gráficas de líneas están que:

a. Se trazan puntos que representan los valores individuales y se conectan con líneas
b. Son útiles para observar tendencias, como crecimientos o disminuciones.

Para comprender cómo se podrían visualizar este tipo de datos con las gráficas de líneas, lo invito a revisar el siguiente ejercicio:

Evolución mensual del precio de maíz

¿Le ha quedado claro el uso de la gráfica de líneas? En este caso concreto, con este tipo de gráficas podrá observar tendencias en los precios del maíz, como el aumento gradual de enero a julio, el pico en julio, y la ligera disminución en los meses siguientes. Esto en la práctica profesional, puede ayudarlo a identificar patrones estacionales y planificar mejor sus estrategias de venta.

- **Histogramas:** son gráficos específicos para **variables numéricas continuas.** Éstos se utilizan para representar la distribución de los datos, dividiéndolos en intervalos (clases) y mostrando la frecuencia de cada intervalo mediante barras contiguas (Martínez, 2020).

Entre las principales características de los histogramas están que:

a. Las barras son adyacentes (sin espacios), lo que indica la continuidad de los datos;
b. Muestran la forma de la distribución (simétrica, sesgada, uniforme, etc.).

A continuación, puede revisar cómo se representa a través de un histograma datos continuos. Previamente, es muy importante que recuerde cómo se elaboran las tablas de frecuencia de datos agrupados, información que fue compartida en la semana 3.

A continuación, le invito a revisar el ejercicio para profundizar sobre lo estudiado.

Rendimiento agrícola en toneladas por hectárea

En la tabla de frecuencias que acaba de revisar, es muy importante aclarar que, las clases son excluyentes, por lo que, se ajustaron los límites superiores de cada intervalo para evitar solapamientos. Para esto, se consideró que el límite inferior está incluido y el superior está excluido en cada intervalo, excepto en el último, que incluye el límite superior. Además, es importante resaltar que, en el histograma, las barras están adyacentes (sin espacios) para reflejar la continuidad de los datos. En este, el eje x muestra los intervalos (clases), mientras que el eje y muestra la frecuencia absoluta.

Actividades de aprendizaje recomendadas

Apreciado estudiante, para asegurar que ha comprendido este contenido referente a la representación gráfica de variables numéricas, lo invito a realizar las siguientes actividades:

1. Realizar una lectura comprensiva del capítulo 5 Gráficas o diagramas, del texto de Posada Hernández 2016, concretamente las páginas de la 53 a la 64 y elaborar un cuadro sinóptico en donde resalte las características principales de las gráficas mencionadas.

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

   La actividad propuesta tiene como finalidad afianzar los conocimientos que se han adquirido durante la semana. Es necesario resaltar que, los histogramas, gráficas de barras y gráficas de pastel son herramientas fundamentales para la representación visual de datos. Elegir la representación adecuada mejora la interpretación y el análisis de la información, por lo que, es importante tener claro que, el histograma se utiliza para variables numéricas continuas, el cual permite identificar patrones en la distribución de los datos. La gráfica de barras, por otro lado, es ideal para variables categóricas, porque facilita la comparación de frecuencias entre diferentes categorías y finalmente, las gráficas de pastel muestran proporciones relativas de un todo, permitiendo visualizar la participación de cada categoría dentro del conjunto de datos.

2. Apreciado estudiante, hemos concluido el estudio de la Unidad 2 y es momento de evaluar su aprendizaje. Para esto, responda a las preguntas que se presentan a continuación. El cuestionario contiene preguntas de opción múltiple, *con una o varias opciones de respuestas correctas*. Esta actividad le servirá de repaso, así como para comprobar los conocimientos adquiridos.

Autoevaluación 2

Seleccione la/las respuesta(s) correcta(s):

1. ¿Qué representa una tabla de frecuencias?

   a. Una lista ordenada de números.
   b. Una organización de datos en categorías o intervalos y sus frecuencias.
   c. Un gráfico que muestra la relación entre variables.
   d. Un cálculo estadístico para resumir datos.

2. ¿Cuál de las siguientes características define a un histograma?

   a. Las barras son contiguas y representan frecuencias de intervalos.
   b. Las barras son separadas para destacar las categorías.
   c. Se usa para representar porcentajes únicamente.
   d. Es un gráfico circular con proporciones.

3. ¿Qué tipo de gráfico sería más adecuado para representar las ventas porcentuales de distintos productos agrícolas?

   a. Gráfica de barras.
   b. Gráfica de líneas.
   c. Gráfica de pastel.
   d. Histograma.

4. ¿Cuáles son las características de una gráfica de pastel?

   a. Muestra proporciones del total en forma de sectores.

b. Es adecuada para datos numéricos continuos.
c. Representa categorías en barras separadas.
d. Es ideal para datos categóricos con pocas categorías.

5. ¿Cuáles son los usos comunes de la gráfica de líneas?

a. Representar datos numéricos continuos a lo largo del tiempo.
b. Comparar proporciones de categorías.
c. Observar tendencias o cambios en variables.
d. Mostrar frecuencias acumuladas en un conjunto de datos.

6. ¿Qué tipo de gráfica es más útil para analizar cómo cambia el precio de un producto agrícola durante el año?

a. Gráfica de barras.
b. Gráfica de pastel.
c. Gráfica de líneas.
d. Histograma.

7. ¿En qué contexto sería más adecuado usar una gráfica de barras en lugar de un histograma?

a. Para representar la distribución de pesos de un producto.
b. Para mostrar la frecuencia de diferentes tipos de cultivos en una región.
c. Para analizar la variabilidad en el rendimiento de cultivos.
d. Para estudiar la concentración de tamaños de parcelas.

8. ¿Qué tipo de gráfica sería más adecuada para mostrar la distribución de edades de agricultores agrupadas en intervalos (20-30, 30-40, etc.)?

a. Gráfica de pastel.
b. Gráfica de barras.
c. Histograma.
d. Gráfica de líneas.

9. ¿Cuáles son las diferencias entre un histograma y una gráfica de barras?

   a. El histograma tiene barras contiguas, mientras que la gráfica de barras tiene espacios entre las barras.
   b. El histograma se usa para datos continuos, mientras que la gráfica de barras es para datos categóricos.
   c. Ambas tienen barras contiguas que representan frecuencias.
   d. El histograma refleja intervalos, mientras que la gráfica de barras muestra categorías.

10. ¿Qué características tienen las barras en una gráfica de barras para variables numéricas?

    a. Las barras son separadas para representar cada categoría.
    b. Las barras son contiguas para reflejar continuidad.
    c. La altura de cada barra representa la frecuencia o el valor.
    d. Se agrupan en intervalos de datos continuos.

Ir al solucionario

Resultado de aprendizaje 3:

Analiza e interpreta datos cuantitativos y cualitativos, que respalden propuestas de manejo y conservación de recursos naturales.

El presente resultado de aprendizaje se logrará mediante la Unidad 3: Estadísticos descriptivos, en donde se proporcionará las herramientas necesarias para resumir y analizar datos relevantes. Estas herramientas, permitirán a los estudiantes sustentar propuestas técnicas basadas en un análisis estadístico sólido.

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

Continuamos con el estudio de esta asignatura, y en esta semana iniciaremos el estudio de la *Unidad 3 Estadístico descriptivos*, en donde exploraremos el tema de medidas de tendencia central. Para esto, analizaremos la media, mediana y moda, herramientas esenciales para resumir datos y comprender su comportamiento central. ¡Iniciamos!

Unidad 3. Estadísticos descriptivos

3.1. Medidas de tendencia central

Estas medidas son herramientas estadísticas que describen el valor central o típico de un conjunto de datos (Itza Ortiz et al., 2024). Su importante radica en que permiten resumir grandes volúmenes de información y representan un punto alrededor del cual los datos se distribuyen. Las tres medidas principales son: **media, mediana y moda**, cada una con características específicas y usos particulares (Itza Ortiz et al., 2024).

A continuación, haremos una revisión de cada una de éstas, con la finalidad de que comprenda cómo deben ser calculadas, así como su utilidad.

- **Media:** O también denominada promedio o promedio aritmético, corresponde a la diferencia entre la suma de todos los valores del conjunto de datos y el número total de observaciones. Es una medida útil cuando los datos están distribuidos de manera uniforme y no presentan valores extremos (outlier) que puedan distorsionar el promedio.

$$Media = \frac{\Sigma x_i}{n}$$

De donde:

x_i = cada valor individual

n= número total de observaciones

- **Mediana:** Esta medida corresponde al valor central de un conjunto de datos, **cuando están ordenados de menor a mayor** (Itza Ortiz et al., 2024). Esta medida es útil cuando los datos presentan valores extremos, ya que no se ve afectada por ellos.

> Si el número de datos es impar, la mediana es el valor central; mientras que, si el número es par, se calcula el promedio de los dos valores centrales.

- **Moda:** Este valor o valores es aquel que más se repiten en un conjunto de datos (Itza Ortiz et al., 2024). Es útil para identificar categorías o datos más frecuentes, especialmente en variables cualitativas o datos agrupados.

> Para reforzar este tema, lo invito a revisar el siguiente documento, concretamente el capítulo 7. Medidas de tendencia central, del texto de Posada Hernández (2016), páginas de la 73 a la 85.

¿Qué le pareció la lectura? ¿han quedado claras estas medidas? Para asegurar que esto sea así, le propongo a continuación un ejercicio, en donde podrá revisar y analizar cómo se calculan estas medidas de tendencia central y su interpretación en un caso de estudio.

Ejercicio media, mediana y moda - Rendimiento agrícola

Con este ejemplo, se puede destacar que, por un lado, la **media** ayuda a calcular el rendimiento promedio por parcela para estimar la producción total. Por otro lado, la **mediana** muestra el rendimiento típico (esta medida es útil si hay valores extremos que podrían sesgar la media), mientras que, la **moda** identifica el rendimiento más común, útil para planificar estrategias enfocadas en parcelas promedio.

Contenidos, recursos y actividades de aprendizaje recomendadas

Semana 7

Continuando con el estudio de esta unidad, en la presente semana revisaremos las *medidas de dispersión y las de posición*. Para las medidas de dispersión, estudiaremos herramientas como el rango, varianza y desviación estándar, que permiten analizar la variabilidad en los datos y ayudan a evaluar la consistencia y dispersión de la información. Por otra parte, en cuanto a las medidas de posición, nos enfocaremos en los percentiles, cuartiles y deciles, herramientas útiles para identificar la ubicación relativa de los datos en un conjunto. ¡Vamos a iniciar!

Unidad 3. Estadísticos descriptivos

3.2. Medidas de dispersión

Estas medidas, son herramientas estadísticas que describen el grado de variabilidad o dispersión de los datos respecto a un punto central, como, por ejemplo, la media (Itza Ortiz et al., 2024, Martínez, 2020).

Las medidas de dispersión complementan a las de tendencia central (media, mediana, moda), ya que proporcionan información sobre qué tan similares o diferentes son los valores de un conjunto de datos. Cuanta mayor dispersión exista, más alejados estarán los valores entre sí o respecto al centro.

Las principales medidas de dispersión son: rango, varianza, **desviación** estándar y coeficiente de variación. A continuación, haremos una revisión de cada una de éstas.

- **Rango:** Esta medida es considerada la más simple y corresponde a la diferencia entre el valor máximo y el valor mínimo del conjunto de datos (Itza Ortiz et al., 2024). No ofrece mucha información sobre la variabilidad de los datos, ya que está basada sólo en los valores extremos. Para su cálculo, se utiliza la siguiente fórmula:

$$Rango = valor\ máximo - valor\ mínimo$$

Por ejemplo, si en una finca, los rendimientos de 5 parcelas son: 3, 5, 6, 7, 8 toneladas (t) por hectárea, el rango sería:

$$Rango = 8 - 3 = 5t/ha$$

En este caso, el rango muestra que hay una diferencia de 5 toneladas entre las parcelas con mayor y menor rendimiento.

- **Varianza:** Esta medida mide cuánto se desvían, en promedio, los datos respecto a la media, elevando al cuadrado las diferencias (Itza Ortiz et al., 2024). Es útil para analizar la dispersión total del conjunto, aunque su unidad está en el cuadrado de la original.

Es importante tener en cuenta que, dependiendo de si los datos provienen de toda una población o de una muestra, su cálculo requerirá de una fórmula diferente

A continuación, se explica a qué se refiere y cómo se calcula tanto la varianza poblacional, como la varianza muestral.

[Varianza poblacional y muestral](#)

En resumen, la *varianza poblacional* mide la dispersión exacta de todos los datos de una población respecto a su media, se utiliza cuando se dispone de información completa para evaluar la variabilidad total. Por otro lado, la *varianza muestral* estima la dispersión de una población a partir de una muestra representativa, utilizando $n-1$ como divisor para ajustar el cálculo y compensar la falta de datos completos.

- **Desviación estándar:** Está definida como la raíz cuadrada de la varianza, lo que permite expresar la dispersión en las mismas unidades que los datos originales (Itza Ortiz et al., 2024). Ésta, es una medida intuitiva para evaluar qué tan dispersos están los datos respecto a la media. Para su cálculo se utiliza la siguiente fórmula:

$$\sigma = \sqrt{\sigma^2} \text{ Para una población}$$

$$s = \sqrt{s^2} \text{ Para una muestra}$$

- **Coeficiente de variación (CV):** Es una medida que relaciona la desviación estándar con la media, para determinar qué tan homogénea o dispersa es la información, es decir, expresa la desviación estándar como un porcentaje de la media, permitiendo comparar la dispersión de diferentes conjuntos de datos. Esta medida se calcula a través de la siguiente fórmula:

$$CV = \frac{s}{x} \times 100$$

Para mejorar la comprensión de este tema, lo invito a revisar el siguiente documento, concretamente el capítulo 9. Medidas de dispersión, del texto de Posada Hernández (2016), páginas de la 96 a la 107.

¿Ha comprendido para qué sirven las medidas de dispersión y cómo se calculan? Es muy importante que refuerce la lectura si algo no ha quedado del todo claro, solamente de esta forma, podremos más adelante utilizar estos conocimientos que ha adquirido.

3.3. Medidas de posición

Para concluir con el estudio de esta Unidad y los contenidos de este bimestre, vamos a destinar nuestros esfuerzos en comprender las medidas de posición. Éstas son herramientas estadísticas que identifican la ubicación relativa de un dato o conjunto de datos dentro de una distribución. Estas medidas permiten dividir y analizar los datos en secciones específicas para entender cómo están distribuidos. Las principales son: los **cuartiles, deciles** y **percentiles** (Itza Ortiz et al., 2024).

- **Cuartil:** Corresponden a una división del número de datos entre 4.
- **Decil:** Corresponden a una división del número de datos entre 10.
- **Percentil:** Corresponden a una división del número de datos entre 100.

Entre estas medidas, nos enfocaremos ahora mismo en los cuartiles, ya que son los más utilizados.

- **Cuartiles:** Son valores que dividen un conjunto de datos en cuatro partes iguales. El **primer cuartil (Q1)** marca el 25% de los datos más bajos, el **segundo cuartil (Q2)** es la mediana y divide a los datos en dos mitades iguales (50%) y el **tercer cuartil (Q3)** marca el 75% de los datos, dejando el 25% superior.

¿Cómo se calculan? A continuación, se muestran las fórmulas:

Tabla 1
Fórmulas para calcular los cuartiles

Cuartil	Fórmula
Primer cuartil (Q1)	$Q1 = \frac{N+1}{4}$
Segundo cuartil (Q2)	$Q2 = \frac{N+1}{2}$
Tercer cuartil (Q3)	$Q3 = \frac{3(N+1)}{4}$

Nota. González, X., 2025.

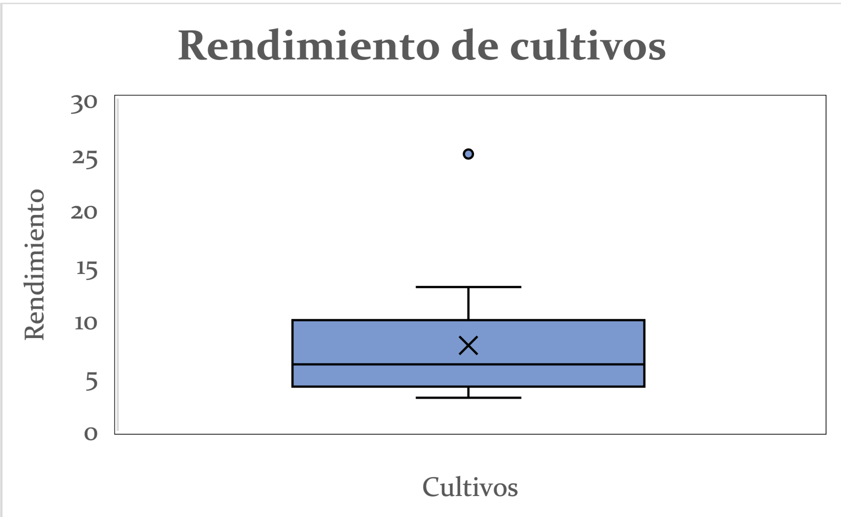
Diagramas de Cajas (Boxplot):

Ahora bien, en cuanto al diagrama de cajas, éste es una representación gráfica que resume la distribución de un conjunto de datos, destacando sus principales características: tendencia central, dispersión, y presencia de valores atípicos. Este gráfico utiliza las medidas de posición para dividir los datos en cuartiles y visualizar su distribución.

¿Cómo se compone un diagrama de cajas? Éste consta de las siguientes partes:

Figura 2

Ejemplo de diagrama de cajas o boxplot



Nota. González, X., 2025.

- **Caja:** Representa el rango intercuartílico ($Q3 - Q1$), donde se encuentra el 50% de los datos.
- **Línea dentro de la caja:** Marca la mediana ($Q2$).
- **Bigotes:** Se extienden hasta el valor más pequeño y más grande que no sea atípico.
- **Outliers o puntos fuera de los bigotes:** Representan valores atípicos.

Para reforzar estos temas, lo invito a revisar el texto de Robayo-Botiva (2020), enfatice la lectura de las páginas 6 a la 8.

¿Qué le ha parecido la lectura? Es importante destacar que, las medidas de posición permiten dividir los datos en segmentos, identificar puntos clave en la distribución, y realizar comparaciones específicas dentro del conjunto, como determinar dónde se concentran los valores o detectar extremos. Mientras que, el diagrama de cajas proporciona una representación gráfica clara y sencilla que combina estas medidas con información adicional sobre la dispersión y la simetría de los datos, destacando valores atípicos de forma visual.

Actividades de aprendizaje recomendadas

Apreciado estudiante, para reforzar los temas de esta semana, lo invito a realizar las siguientes actividades:

1. Realizar una lectura comprensiva de los temas Medidas de dispersión, y Medidas de posición para fortalecer los temas tratados.

   Es importante tener en cuenta que, las medidas de dispersión y medidas de posición son herramientas estadísticas que permiten analizar la distribución de los datos, pero cumplen funciones diferentes. Por un lado, las medidas de dispersión (como el rango, la varianza y la desviación estándar) indican cuánto se alejan los datos de la media, ayudando a comprender la variabilidad dentro del conjunto de datos, mientras que, las medidas de posición (como los cuartiles, percentiles y la mediana) dividen los datos en partes y permiten identificar valores clave en la distribución, como la tendencia central y los puntos de referencia.

2. Realizar el siguiente ejercicio:

Un agricultor ha recopilado datos sobre el rendimiento (en kilogramos por hectárea) de 20 parcelas cultivadas con maíz. Los datos son los siguientes:

540, 520, 580, 600, 560, 520, 510, 550, 570, 590, 530, 550, 580, 600, 570, 560, 550, 590, 610, 540.

Con estos datos, proceda a completar la tabla a continuación, en donde previamente deberá realizar algunos cálculos. Así mismo, responda a las preguntas propuestas:

Tabla a completar

Estadísticos descriptivos	Medidas a calcular	Preguntas de interpretación
Medidas de tendencia central	**Media:** **Mediana:** **Moda:**	*¿Cuál de estas medidas representa mejor el rendimiento típico del cultivo y por qué?*
Medidas de dispersión	**Rango:** **Varianza:** **Desviación estándar:**	*¿Qué indican estas medidas sobre la variabilidad en la producción de las parcelas?*
Medidas de posición	**Cuartiles**	*¿Cómo estas medidas pueden ayudar al agricultor a entender qué porcentaje de parcelas tiene un rendimiento superior o inferior a ciertos valores?*

Nota: copie la tabla en un Word o cuaderno para completar.

¡Ha realizado un excelente trabajo! En este ejercicio, se exploró cómo las *medidas de tendencia central* permiten tener diferentes perspectivas del rendimiento típico, en donde la **media** es la ideal para tener un panorama general y la **mediana** es útil cuando hay valores extremos. En cuanto a las medidas de dispersión, estás permitieron evaluar la variabilidad en la producción, destacando la consistencia o desigualdad entre parcelas. Finalmente, las medidas de posición como los cuartiles, permiten identificar el desempeño relativo de las parcelas, especialmente aquellas con rendimientos superiores o inferiores.

3. Apreciado estudiante, hemos concluido el estudio de la Unidad 3 y es momento de evaluar su aprendizaje. Para esto, responda a las preguntas que se presentan a continuación. El cuestionario contiene preguntas de opción múltiple, *con una o varias opciones de respuestas correctas*. Esta actividad le servirá de repaso, así como para comprobar los conocimientos adquiridos.

[Autoevaluación 3]

Seleccione la/las respuesta(s) correcta(s):

1. ¿Qué mide la mediana en un conjunto de datos?

   a. El valor promedio de los datos.
   b. El valor más frecuente en los datos.
   c. El valor central cuando los datos están ordenados.
   d. La variabilidad de los datos.

2. Si los rendimientos agrícolas de 5 parcelas son: 4, 7, 6, 8, 10 toneladas por hectárea, ¿cuál es la media?

   a. 6
   b. 7
   c. 8
   d. 10

3. En un conjunto de datos, la varianza es 9. ¿Cuál es la desviación estándar?

   a. 3
   b. 4
   c. 9

d. 27

4. Si los datos de rendimientos agrícolas están agrupados en cuartiles, ¿qué porcentaje de datos se encuentra por debajo del tercer cuartil ($Q3$)?

    a. 25%
    b. 50%
    c. 75%
    d. 100%

5. ¿Qué características son verdaderas para la media aritmética?

    a. Representa el promedio de los datos.
    b. Es sensible a valores extremos.
    c. Divide los datos en cuatro partes iguales.
    d. Utiliza todos los valores del conjunto.

6. ¿Qué información proporciona un diagrama de cajas?

    a. La dispersión total del conjunto de datos.
    b. La mediana y los cuartiles.
    c. La posición de valores atípicos.
    d. La relación entre la media y la mediana.

7. Si un agricultor desea comparar la variabilidad relativa de rendimientos entre maíz y arroz, ¿qué medida de dispersión es más adecuada?

    a. Varianza.
    b. Desviación estándar.
    c. Rango.
    d. Coeficiente de variación.

8. ¿Cuáles de las siguientes afirmaciones son verdaderas respecto a la desviación estándar?

    a. Mide la dispersión promedio de los datos respecto a la media.

b. Se calcula como la raíz cuadrada de la varianza.
c. Es una medida de posición que indica el valor central de los datos.
d. No se ve afectada por valores extremos en el conjunto de datos.

9. Si un conjunto de datos tiene valores atípicos altos, ¿qué medidas son más apropiadas para describir su tendencia central y dispersión?

   a. Media aritmética y varianza.
   b. Mediana y rango intercuartílico.
   c. Moda y desviación estándar.
   d. Mediana y desviación estándar.

10. Si se tiene una distribución de datos simétrica sin valores atípicos, ¿qué medidas son apropiadas para describir su tendencia central y dispersión?

   a. Media aritmética y desviación estándar.
   b. Mediana y rango.
   c. Moda y rango intercuartílico.
   d. Media aritmética y rango.

Ir al solucionario

Resultado de aprendizaje 1 a 3:

- Reconoce y caracteriza los distintos tipos de variables estadísticas.
- Construye e interpreta tablas y gráficas estadísticas.
- Analiza e interpreta datos cuantitativos y cualitativos, que respalden propuestas de manejo y conservación de recursos naturales.

Contenidos, recursos y actividades de aprendizaje recomendadas

Semana 8

Actividades finales del bimestre

Estimado estudiante, al concluir este primer bimestre, es momento de prepararse para la evaluación bimestral. Para ello, le recomiendo revisar nuevamente los apuntes y resúmenes que ha elaborado a lo largo de estas semanas, con el objetivo de refrescar sus conocimientos e identificar los temas que puedan requerir mayor atención. Para estos últimos, es fundamental dedicar tiempo adicional y repasar los recursos compartidos previamente.

Para fomentar su pensamiento crítico y el trabajo autónomo, le sugiero proponer ejercicios sencillos que le permitan poner en prácticas los temas que hemos revisado y que ya han sido reforzados anteriormente. Al finalizar esta sección, se propone un ejercicio para que ponga en práctica lo aprendido. Esto le permitirá tener una mayor seguridad y dominio de los conocimientos adquiridos.

Aproveche el horario de tutorías de esta semana para plantear sus dudas al docente y recibir el apoyo necesario en los temas que considere más complejos.

Durante la tutoría de la semana 8, el docente reforzará los contenidos que presenten mayor dificultad, basándose en aquellos temas que han sido compartidos por los estudiantes. Por ello, su asistencia es clave para optimizar este espacio y resolver cualquier duda. Si no le es posible asistir, el video de la tutoría estará disponible posteriormente a través de un anuncio académico.

¡Manos a la obra!

A continuación, se propone un ejercicio sencillo en donde se solicita realizar un análisis estadístico de datos correspondientes a la producción agrícola de una finca.

Una finca produce tomates y ha registrado el peso (en gramos) de 15 tomates seleccionados al azar en la última cosecha. Los datos son los siguientes:

145, 160, 155, 170, 165, 180, 150, 175, 160, 155, 170, 165, 180, 150, 175

Para estos datos, se pide calcular:

- Medidas de tendencia central: Media, mediana y moda.
- Medidas de dispersión: Rango, varianza y desviación estándar.
- Medidas de posición: Cuartiles Q1, Q2 y Q3.

Para realizar este ejercicio, es muy importante dar un vistazo nuevamente a los temas de: medidas de tendencia central, medidas de dispersión y medidas de posición, enfatizando la revisión de las fórmulas para proceder con los cálculos.

Este ejercicio, por una parte, muestra cómo calcular las medidas de tendencia central para conocer el valor promedio y más representativo de los datos. Así mismo, pone en evidencia que las medidas de dispersión ayudan a evaluar la variabilidad de los pesos de los tomates, mientras que las medidas de posición dividen los datos en secciones útiles para interpretaciones más detalladas.

Segundo bimestre

Resultado de aprendizaje 4:

Conoce y aplica leyes de las probabilidades para modelizar variables aleatorias discretas y continuas.

El presente resultado de aprendizaje se alcanzará mediante el análisis de los temas relacionados a las reglas de probabilidad (Unidad 4) y el uso de distribuciones de probabilidad (Unidad 5). Aquí, se analizarán las leyes de la probabilidad (suma y multiplicación) y aplicarán distribuciones discretas (binomial, Poisson) y continuas (normal y exponencial) para modelar fenómenos agroempresariales, como la variabilidad en rendimientos de cultivos y la demanda de productos agrícolas. Estas herramientas permitirán interpretar la incertidumbre en procesos agrícolas y tomar decisiones informadas para optimizar la producción y minimizar riesgos.

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

Vamos a iniciar este segundo bimestre y lo haremos con el tema de la probabilidad. En esta semana, nos enfocaremos en los temas de *conceptos básicos y reglas de la probabilidad*. Nos centraremos en la revisión de los principios fundamentales como experimentos aleatorios, eventos, y analizaremos las reglas de suma y multiplicación, aplicando estos conceptos para calcular probabilidades y comprender cómo se utilizan en la toma de decisiones.

4.1. Conceptos básicos

¿Ha escuchado hablar sobre probabilidades? De seguro que sí. En la vida cotidiana y en los agronegocios, constantemente enfrentamos situaciones de incertidumbre. No siempre podemos predecir con certeza si lloverá mañana, si el precio de un producto agrícola subirá o si una plaga afectará la producción de una cosecha. Sin embargo, podemos hacer predicciones fundamentadas sobre la probabilidad de que estos eventos ocurran.

La **probabilidad** es la herramienta matemática que permite cuantificar la incertidumbre (Nores, 2021).

La probabilidad nos ayuda a medir la posibilidad de que ocurran ciertos eventos o sucesos, brindando una base sólida para la toma de decisiones. Esta predicción se expresa mediante un número entre 0 y 1, donde 0 indica que el evento es imposible y 1 indica que el evento es seguro.

Para comprender mejor este tema, a continuación, se presentan algunos conceptos básicos:

- **Experimento:** Es cualquier proceso que, al repetirse en las mismas condiciones, produce resultados inciertos (observación o medición).

  Ejemplo: Al lanzar una moneda, no se puede predecir con certeza si saldrá "cara" o "cruz".

- **Espacio muestral (S):** Es el conjunto de todos los posibles resultados de un experimento aleatorio (Monroy Saldívar, 2008).

  Ejemplo: En el lanzamiento de un dado, el espacio muestral es:

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

- **Evento (E):** Es cualquier subconjunto del espacio muestral, es decir, uno o varios resultados posibles de un experimento (Monroy Saldívar, 2008).

  Ejemplo:

  - **Evento A:** Obtener "cara" al lanzar una moneda.
  - **Evento B:** Obtener un número par al lanzar un dado:

  $$E=\{2,4,6\}$$

- **Probabilidad de un evento P(E):** La probabilidad de un evento es la relación entre el número de casos favorables (los que forman parte del evento) y el total de posibles resultados (el espacio muestral). Para su cálculo se utiliza la siguiente fórmula:

  $$P(E) = \frac{E}{S}$$

  De donde:

  E: Número de casos favorables.

  S: Total de posibles resultados

  Ejemplo:

  Si lanzamos un dado, la probabilidad de obtener un número par (evento B) sería:

  $$P(E) = \frac{3}{6} = 0.5$$

  Esto indica que la probabilidad de obtener un número par es del 50%

- **Eventos mutuamente excluyentes:** Son aquellos que no pueden ocurrir al mismo tiempo en un mismo experimento aleatorio (Monroy Saldívar, 2008). En otras palabras, la ocurrencia de un evento excluye la posibilidad de que ocurra el otro. Esta relación se presenta cuando los eventos no tienen intersección (es decir, no comparten resultados comunes).

Dos eventos A y B son mutuamente excluyentes si la intersección de ambos es un conjunto vacío:

$$A \cap B = \emptyset$$

Continuando con este tema, es muy importante reconocer que, en algunas ocasiones, el evento de interés se formará como una combinación de otros eventos. A continuación, pueden visualizar las tres relaciones importantes que pueden ocurrir:

Relaciones de evento

¿Le ha quedado claro estás tres relaciones? Es importante que tenga en cuenta que, cualquier evento simple en el área sombreada, es un posible resultado que aparece en el evento apropiado. En estos casos, una forma de conocer las probabilidades de la unión, intersección o el complemento, es sumar las probabilidades de todos los eventos simples asociados.

4.2. Reglas de la probabilidad

Ahora bien, para concluir con los temas de esta semana, vamos a revisar las **reglas fundamentales de la probabilidad**. Éstas son consideradas como principios que permiten calcular la probabilidad de que ocurran eventos individuales o combinados y son la base para resolver problemas de probabilidad en contextos reales, como la predicción de rendimientos agrícolas, la demanda de productos o el análisis de riesgos climáticos en los agronegocios.

- **Regla de la adición:** Esta regla permite calcular la probabilidad de que ocurra al menos uno de dos eventos *A* o *B* (Monroy Saldívar, 2008). La probabilidad de la unión de *A* y *B* se expresa como:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Por otro lado, cuando dos eventos *A* y *B* son mutuamente excluyentes, significa que no pueden ocurrir al mismo tiempo. En este caso, la probabilidad de que ocurra al menos uno de ellos es la suma de sus probabilidades individuales.

$$P(A \cup B) = P(A) + P(B)$$

- **Regla para complementos:** La probabilidad de que ocurra un evento *A* y la probabilidad de que **no ocurra A** (su complemento A^C) siempre suma 1.

  Si recuerda la representación gráfica de la relación de complemento, *A* y A^C son mutuamente excluyentes, por tanto, $A \cup A^c = S$ (todo el espacio muestral). Con base en esto, se deduce que:

$$P(A^c) = 1 - P(A)$$

- **Regla de la multiplicación**: Esta regla permite calcular la probabilidad de que dos eventos ocurran juntos (intersección de eventos). Se puede presentar tanto para *eventos independientes*, en donde la ocurrencia de uno no afecta al otro, así como para *eventos dependientes*, en donde la ocurrencia de A afecta la probabilidad de B (Monroy Saldívar, 2008).

  - Eventos independientes: $P(A \cap B) = P(A) \times P(B)$
  - Eventos dependientes: $P(A \cap B) = P(A) \times P(B|A)$

De donde:

P(B|A): es la probabilidad condicional de que ocurra *B* dado que *A* ya ocurrió.

Actividades de aprendizaje recomendadas

Estimado estudiante, para reforzar los temas de esta semana, lo invito a realizar la siguiente actividad:

1. Realizar una lectura comprensiva sobre [Temas de probabilidad](#).

La lectura propuesta pretende afianzar los temas introductorios de probabilidades, así como seguir forjando su capacidad de trabajo autónomo. Recuerde que, la probabilidad permite cuantificar la incertidumbre y predecir la ocurrencia de eventos. Se basa en la teoría de conjuntos y se expresa mediante valores entre 0 y 1, donde 0 indica imposibilidad y 1 certeza. Entre sus conceptos fundamentales se encuentran el espacio muestral, los eventos y las reglas básicas como la adición y la multiplicación. Comprender estos principios es esencial para la toma de decisiones, sobre todo en áreas como los agronegocios, donde la incertidumbre es un factor clave.

2. Apreciado estudiante, hemos concluido el estudio de la Unidad 4 y es momento de evaluar su aprendizaje. Para esto, responda a las preguntas que se presentan a continuación. El cuestionario contiene preguntas de opción múltiple, *con una o varias opciones de respuestas correctas*. Esta actividad le servirá de repaso, así como para comprobar los conocimientos adquiridos.

Autoevaluación 4

Seleccione la/las respuesta(s) correcta(s):

1. ¿Cuál de las siguientes opciones describe mejor el concepto de probabilidad?

   a. La cantidad de veces que un evento ocurre en un experimento.
   b. Una medida de certeza sobre un evento futuro, expresada entre 0 y 1.
   c. El número de posibles resultados en un experimento.
   d. La diferencia entre el número de éxitos y el número de fracasos.

2. ¿Qué se entiende por un evento aleatorio?

   a. Un evento que siempre ocurre en un experimento.
   b. Un evento que no tiene probabilidades asociadas.
   c. Un evento que nunca ocurre.

d. Un evento cuyo resultado no puede predecirse con certeza antes de realizar el experimento.

3. Si lanzamos un dado equilibrado de 6 caras, ¿cuál es la probabilidad de obtener un número impar?

   a. 1/6
   b. 2/6
   c. 3/6
   d. 4/6

4. Dos eventos A y B son mutuamente excluyentes. Si P(A) = 0.3 y P(B) = 0.5 ¿cuál es $P(A \cup B)$?

   a. 0.2
   b. 0.8
   c. 1.0
   d. 0.15

5. ¿Cuál de las siguientes afirmaciones es verdadera para eventos mutuamente excluyentes?

   a. Ambos eventos pueden ocurrir al mismo tiempo.
   b. La probabilidad de la intersección de los eventos es 1.
   c. La probabilidad de la intersección de los eventos es 0.
   d. Los eventos siempre tienen probabilidades iguales.

6. ¿Cuáles de las siguientes afirmaciones son correctas para un espacio muestral?

   a. Contiene todos los posibles resultados de un experimento aleatorio.
   b. La suma de las probabilidades de todos los eventos en el espacio muestral es igual a 1.
   c. Siempre incluye eventos que no pueden ocurrir.

d. Incluye solo eventos independientes.

7. ¿Cuáles de las siguientes afirmaciones son correctas sobre lanzar un dado equilibrado de 6 caras?

   a. La probabilidad de obtener un número par es 3/6.
   b. La probabilidad de obtener un número impar es 3/6.
   c. La probabilidad de obtener un 7 es 1/6.
   d. La probabilidad de obtener un número mayor a 6 es 0.

8. Si la probabilidad de que llueva en un día es 0.4 y la probabilidad de que un agricultor riegue su cultivo ese mismo día es 0.6, ¿cuál es la probabilidad de que ambos eventos ocurran, asumiendo independencia?

   a. 0.24
   b. 0.10
   c. 0.48
   d. 0.80

9. Si la probabilidad de que un agricultor coseche frutas maduras es 0.7, ¿cuál es la probabilidad de que no las coseche maduras?

   a. 0.7
   b. 0.3
   c. 1.0
   d. 0.5

10. Si la probabilidad de que un cultivo tenga éxito es 0.85, ¿cuál es la probabilidad de que el cultivo fracase?

   a. 0.85
   b. 0.50
   c. 0.15
   d. 0.95

Ir al solucionario

Contenidos, recursos y actividades de aprendizaje recomendadas

Semana 10

Durante esta semana, iniciaremos el estudio de la Unidad 5 Distribución de la Probabilidad, en donde nos centraremos en el análisis del tema de *distribución binomial de la probabilidad*. Para esto, estudiaremos cómo modelar experimentos aleatorios con dos posibles resultados (éxito o fracaso), reconociendo que este conocimiento será clave para resolver problemas prácticos en el campo de los agronegocios, como, por ejemplo, para evaluar la probabilidad de éxito en una serie de eventos relacionados con la producción o la comercialización.

Unidad 5. Distribución de la probabilidad

Para iniciar con el estudio de esta unidad, es necesario conocer *¿Qué es una variable aleatoria?* Una variable aleatoria es variable cuyo valor se encuentra determinado por el azar para cada resultado de un experimento (Monroy Saldívar, 2008). Los valores que toma la variable dependen del azar, y su comportamiento se describe mediante una distribución de probabilidad. Estas variables, pueden ser *discretas o continuas*, dependiendo de si pueden tomar valores aislados o cualquier valor dentro de un rango continuo (Monroy Saldívar, 2008).

En la tabla a continuación, se muestra una descripción de estas variables y algunos ejemplos.

Tabla 2
Tipos de variables aleatorias

Tipos de variables	Descripción	Ejemplos
Variable aleatoria discreta	Toma valores específicos y contables (números enteros).	• El número de semillas que germinan de un lote de 10. • La cantidad de frutas defectuosas en una muestra de 100.
Variable aleatoria continua	Toma infinitos valores posibles dentro de un intervalo continuo.	• La altura de plantas de maíz. • El peso de una fruta.

Nota. González, X., 2025.

Es muy importante tener claro estos dos tipos de variables aleatorias (discretas o continuas), porque para cada una de éstas se usa técnicas diferentes para describir sus distribuciones (Mendenhall et al. 2015). Por ahora, nos enfocaremos en las variables aleatorias discretas.

Ahora bien, se puede definir la *distribución de probabilidad* para una variable aleatoria *x*, como la distribución de frecuencia relativa construida para toda la población de mediciones (Mendenhall et al. 2015). Para el caso puntual de una **variable aleatoria discreta**, *la distribución de probabilidad* será una fórmula, tabla o gráfica en donde se presenten los posibles valores de *x* y la probabilidad *p(x)* asociada con cada valor de *x*.

En la tabla que se presenta a continuación, se muestran los requisitos para una distribución de probabilidad discreta.

Tabla 3
Requisitos para una distribución de probabilidad discreta

Requisitos	Descripción	Ejemplos
Probabilidades entre 0 y 1	Cada probabilidad debe estar entre 0 y 1 $0 \leq p(x) \leq 1$	*P(x=2)* = 0.3 (cumple)
Suma de probabilidades = 1	La suma de todas las probabilidades debe ser 1 $\Sigma P(x) = 1$	0.1+0.2+0.4+0.3 = 1
Espacio muestral definido	El conjunto de valores posibles debe estar bien definido	Para X = "número de éxitos" $S = \{0, 1, 2, 3, \ldots, n\}$

Nota. González, X., 2025.

Para reforzar este tema sobre las variables aleatorias y particularmente sobre las variables aleatorias discretas, lo invito a revisar del artículo de Nores (2021), enfatice la lectura de la página 7 a la 13.

¿Qué le ha parecido la lectura?, ¿tiene ahora mucho más claro este tema? Es importante reconocer que las variables aleatorias son esenciales para modelar fenómenos aleatorios y que particularmente, cuando hacemos referencia a las variables aleatorias discretas, estamos considerando a aquellas que toman valores específicos, como números enteros. Su comprensión es clave para aplicar distribuciones como la binomial o la de Poisson. Recuerde que puede volver a revisar este recurso, cuantas veces sea necesario. Lo importante es tener la seguridad de que está aprendiendo.

5.1. Distribución binomial de la probabilidad

Una vez que hemos conocido a las variables aleatorias discretas, es importante mencionar que hay tres distribuciones discretas de probabilidad (probabilidad binomial, de Poisson e hipergeométrica). Vamos a destinar nuestros esfuerzos para comprender las dos primeras.

Usted se preguntará *¿Qué es una distribución binomial?* Pues bien, esta es una de las distribuciones de probabilidad discreta más importantes en la estadística. Se aplica a experimentos que tienen solo dos posibles resultados: *éxito o fracaso* (Itza Ortiz et al., 2024). Esta distribución permite calcular la probabilidad de obtener una cantidad específica de éxitos en un número determinado de repeticiones o ensayos (Monroy Saldívar, 2008).

La **distribución binomial** describe la probabilidad de que ocurra un número específico de éxitos en n ensayos independientes, donde cada ensayo tiene una probabilidad constante de éxito p y de fracaso $q = 1-p$ (De la Puente Viedma, 2018).

La distribución binomial se aplica cuando se cumplen las siguientes condiciones (Itza Ortiz et al., 2024):

1. El experimento consiste en *n* intentos idénticos.
2. Cada ensayo tiene dos resultados posibles (éxito o fracaso).
3. Los intentos son independientes (el resultado de un ensayo no afecta a los demás).
4. La probabilidad de éxito (*p*) es constante en cada ensayo.

La probabilidad de obtener exactamente *k* éxitos en *n* ensayos se calcula con la siguiente fórmula:

$$P(x = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

De donde:

$\binom{n}{k} = \frac{n!}{(k!(n-k))}$ es el coeficiente binomial, que calcula de cuántas formas se pueden seleccionar *k* éxitos de *n* ensayos.

p^k es la probabilidad de obtener k éxitos.

$(1-p)^{n-k}$ es la probabilidad de obtener $n-k$ fracasos.

En la tabla que se muestra a continuación, se expone las propiedades de la distribución binomial:

Tabla 4

Propiedades de la distribución binomial

Propiedades	Descripción	Fórmula
Media (μ)	La media indica el número esperado de éxitos.	$\mu = n \times p$
Varianza (σ^2)	La varianza mide la dispersión de la variable aleatoria.	$\sigma^2 = n \times p(1-p)$
Desviación estándar (σ)		$\sigma = \sqrt{n \times p(1-p))}$

Nota. González, X., 2025.

Vamos a poner en práctica lo aprendido. A continuación, que se presenta un ejemplo de la aplicación de este tipo de distribución:

Ejemplo de Distribución binomial

¿Le quedó claro cómo se puede realizar este tipo de análisis? En este caso en concreto, el análisis realizado permite al agricultor anticipar la cantidad de semillas que germinarán. Si el objetivo del agricultor es garantizar una cierta cantidad mínima de plantas, este análisis le ayudaría a planificar su siembra de forma más eficiente.

Por favor, revise el siguiente video sobre Distribución binomial, el mismo le ayudará a fortalecer su aprendizaje sobre este tema.

¿Le pareció interesante este video? Estimados estudiantes, es necesario tener en cuenta que, la distribución binomial es utilizada para predecir la probabilidad de obtener un número determinado de éxitos en un número fijo de intentos, cada uno con dos posibles resultados: éxito o fracaso, en donde cada

intento es independiente de los demás. Seguramente, luego de ver este video, usted tendrá mucho más claro, cuándo y cómo se realizan los análisis de distribución binomial.

Contenidos, recursos y actividades de aprendizaje recomendadas

Semana 11

Continuando con otra distribución discreta de probabilidad, vamos a revisar en la presente semana la *distribución de Poisson*. Este tipo de distribución, es considerada una herramienta estadística ideal para modelar eventos poco frecuentes que ocurren a una tasa constante en un intervalo de tiempo determinado.
¡Vamos a iniciar!

Unidad 5. Distribución de la probabilidad

5.2. Distribución de la probabilidad de Poisson

Esta distribución se aplica para datos que representan el número de ocurrencias de algún suceso durante un intervalo determinado. En otras palabras, describe la probabilidad de que un evento ocurra un número de veces (*k*) en un intervalo de tiempo, espacio o área específica (Monroy Saldívar, 2008). Para esto, se deben cumplir con las siguientes condiciones:

- Los eventos ocurren de forma aleatoria e independientemente unos de otros.
- La tasa de ocurrencia de eventos (λ) es constante en el intervalo.

Si μ corresponde al número de veces que ocurre un evento en cierto período o espacio. La probabilidad constante de k sucesos de este evento es:

$$P(x=k)=\frac{\lambda^k e^{-\lambda}}{k!}$$

De donde:

$P(x = k)$ = probabilidad de *k* ocurrencias en un intervalo

λ = media o valor esperado de *k*

e = base de los logaritmos naturales (2.71828...)

La media y la desviación estándar de la variable aleatoria de Poisson x, para valores de k = 0, 1, 2, 3, ... se presentan a continuación:

Media: λ

Desviación estándar: $\sigma = \sqrt{\lambda}$

La distribución de Poisson modela situaciones en las que se cuenta **el número de eventos en intervalos específicos** y se espera que estos eventos ocurran de forma aleatoria, pero a una **tasa promedio conocida.**

En la infografía que se presenta a continuación, puede revisar algunos ejemplos prácticos que se pueden modelar mediante una variable aleatoria de Poisson.

[Ejemplos prácticos: Distribución de Poisson]

¿Le ha quedado más claro en qué casos se puede modelar con la distribución de Poisson? Recuerde que, la distribución de Poisson es una herramienta poderosa para modelar la cantidad de eventos aleatorios, que ocurren en un intervalo de tiempo o espacio. Es muy importante que comprenda estos conceptos básicos, para que posteriormente pueda aplicarlos en ejemplos sencillos.

Con esta información entendida, vamos a poner en práctica lo aprendido. En el recurso que se presenta a continuación, revise un ejemplo de la aplicación de este tipo de distribución:

[Control de plagas en un cultivo de tomate]

¿Qué le pareció este ejemplo? Este caso de estudio ha permitido aplicar la distribución de Poisson para predecir la aparición de insectos en un cultivo de tomate. Estimados estudiantes, es muy importante que, aprendan a utilizar este tipo de probabilidad e interpretar sus resultados, a fin de que se pueda tomar decisiones informadas, como, por ejemplo, en el control de plagas.

Actividades de aprendizaje recomendadas

Estimado estudiante, para reforzar los temas de esta semana, lo invito a realizar las siguientes actividades:

1. Realizar una lectura comprensiva sobre Distribución binomial y Distribución de Poisson para reforzar sus conocimientos.

   ¿Aportó la lectura a mejorar la comprensión de estos temas? Espero que así sea. Recuerde que, la distribución binomial y la distribución de Poisson son dos modelos probabilísticos que describen eventos discretos, pero se aplican en contextos diferentes. Por un lado, la distribución binomial se usa cuando hay un número fijo de ensayos, mientras que la distribución de Poisson es útil cuando se cuentan eventos en un intervalo sin un número predefinido de intentos.

2. Analice los siguientes ejemplos y clasifique a los mismos (marcando con una x), de acuerdo al tipo de distribución que podría aplicarse: Binomial o Poisson. Elabore una justificación para su elección.

Binomial o Poisson

Ejemplos	Distribución Binomial	Distribución de Poisson	Justificación
El número de plantas de maíz que germinan en una caja de 100 semillas.			
El número de llamadas telefónicas que recibe un centro de atención al cliente en una hora.			
El número de caras que se obtienen al lanzar una moneda 10 veces.			
El número de defectos en un rollo de tela de 100 metros.			
El número de días lluviosos en un mes.			
El número de clientes que compran un producto específico en una tienda en un día.			
El número de errores tipográficos en una página de un libro.			
El número de accidentes automovilísticos en una carretera en un año.			
El número de bombillas que se queman en una habitación con 20 bombillas en un mes.			

Ejemplos	Distribución Binomial	Distribución de Poisson	Justificación
El número de veces que se saca un as al extraer una carta de una baraja 5 veces.			

Nota: copie la tabla en un Word o cuaderno para completar.

¡Ha realizado un buen trabajo! Esta actividad le ha permitido aplicar los conocimientos adquiridos sobre las distribuciones binomial y de Poisson a situaciones reales. Es necesario recordar que, la distribución binomial es ideal para un número fijo de ensayos con dos posibles resultados, mientras que la de Poisson se encarga de modelar eventos que ocurren a una tasa constante en un intervalo. Con esto, dé un vistazo a las justificaciones que ha planteado y corrobore que está en lo correcto.

Contenidos, recursos y actividades de aprendizaje recomendadas

Semana 12

Estimado estudiante, vamos a continuar con el estudio de esta Unidad y en esta semana, es momento de analizar la primera distribución que utiliza variables aleatorias continuas: *la distribución normal*. En este apartado, revisaremos cómo esta distribución en forma de campana describe una amplia variedad de fenómenos naturales y sociales. Aprenderemos a calcular probabilidades utilizando la tabla de la distribución normal estándar y a interpretar los resultados en el contexto de problemas reales. ¡Vamos a iniciar!

Unidad 5. Distribución de la probabilidad

5.3. Distribución normal de la probabilidad

Esta distribución, es una de las distribuciones más importantes en estadística y probabilidad (Itza Ortiz et al., 2024). Se utiliza para describir fenómenos naturales, sociales y de negocios, especialmente en los Agronegocios. Muchos procesos, como la altura de las plantas, el peso de las frutas o los rendimientos agrícolas, se distribuyen de forma normal, permitiendo realizar predicciones, establecer controles de calidad y tomar decisiones basadas en datos.

Para poder iniciar con esta revisión, hay algunos conceptos importantes que tenemos que abordar, para poder asegurar la comprensión de este tema.

A continuación, se detallan los conceptos de la Distribución normal.

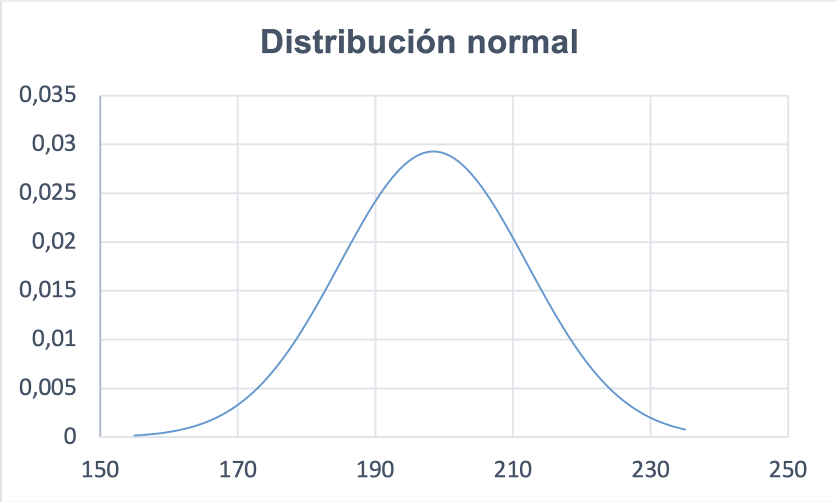
Conceptos de la Distribución normal

¿Han quedado claros estos conceptos? Es muy importante que vuelva a revisarlos, en caso de que alguno de éstos no esté del todo claro. Esto le permitirá comprender lo que veremos a continuación.

La **distribución normal** es una *distribución continua* que se caracteriza por su forma de campana simétrica (campana de Gauss) (Figura 10). Los datos se agrupan alrededor de la media (μ), y la dispersión de los datos está controlada por la desviación estándar (σ) (Monroy Saldívar, 2008).

Figura 3

Ejemplo de una distribución normal



Nota. González, X., 2025.

A continuación, se mencionan algunas características de las variables aleatorias continuas (Mendenhall et al. 2015):

- El área bajo una distribución continua de probabilidad es igual a 1.
- La probabilidad de que *x* caiga en un intervalo particular, por ejemplo, de *a* a *b* es igual al área bajo la curva entre los dos puntos *a* y *b*.
- $P(x = a) = 0$. Lo que implica que, $P(x \geq a) = P(x > a) y P(x \leq a) = P(x < a)$.

Dicho esto, es importante conocer que las distribuciones de probabilidad continua, pueden tomar varias formas, pero un gran número de las variables aleatorias observadas, poseen una distribución de frecuencias que tiene la forma de un montículo, o es aproximadamente una distribución normal de probabilidad. Esta distribución se representa por la función de densidad de probabilidad, que se muestra a continuación:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

De donde:

x = valor de la variable aleatoria.

μ = media (centro de la distribución).

σ = desviación estándar (define la amplitud de la curva).

e = constante de Euler (e=2,718).

La **distribución normal** se utiliza para describir situaciones en las que la mayoría de los valores se encuentran cerca de la media, con menos valores extremos a ambos lados (Itza Ortiz et al., 2024).

Otro tema relevante dentro de la distribución normal es el procedimiento de estandarización. La estandarización se realiza para facilitar el cálculo de probabilidades y la comparación entre diferentes conjuntos de datos. Este procedimiento permite transformar una variable normal x con media μ y desviación estándar σ en una **variable normal estándar z**, que tiene media 0 y desviación estándar 1 (Mendenhall et al. 2015, Monroy Saldívar, 2008). La estandarización se basa en la fórmula:

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

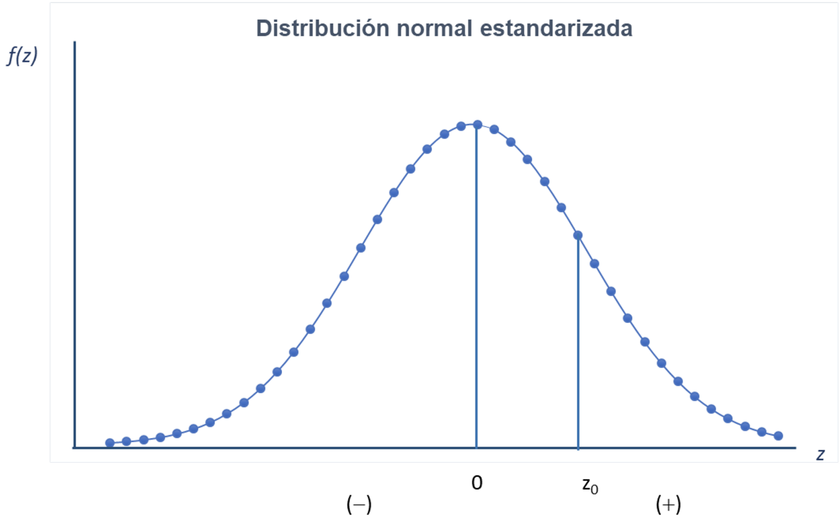
De esta fórmula, podemos extraer las siguientes conclusiones:

- Cuando x es menor que la media μ, el valor de z es negativo.
- Cuando x es mayor que la media μ, el valor de z es positivo.
- Cuando $x = \mu$, el valor de z = 0.

La distribución de probabilidad para z (Figura 4), se denomina **distribución normal estandarizada** porque su media es 0 y su desviación estándar es 1.

Figura 4

Distribución normal estandarizada



Nota. González, X., 2025.

Es así que, los valores de z del lado izquierdo de la curva son negativos, mientras que los del lado derecho son positivos. El **área acumulada** en la distribución normal estandarizada se refiere a la probabilidad acumulada desde el extremo izquierdo de la curva $(-\infty)$ hasta un punto z específico en el eje horizontal. El área acumulada se interpreta como la probabilidad de que la variable aleatoria z tome un valor menor o igual a z_0.

$$P(z \leq z_0)$$

Una vez que se ha comprendido cómo se calcula el área acumulada, su valor se va a encontrar en la tabla de la distribución normal estándar.

La **tabla de la distribución normal estandarizada** la puede revisar en el documento de Distribución normal. Revise, por favor.

Ahora que conoció a la tabla de la distribución normal estándar, es importante comentar que ésta es una herramienta esencial para calcular probabilidades asociadas a variables aleatorias que siguen una distribución normal. ¿Cómo se usa? El primer paso es transformar la variable normal a una variable estándar (la fórmula la puede volver a revisar en los contenidos anteriores). Con el valor de Z obtenido, se busca en la tabla el área bajo la curva normal estándar a la izquierda de ese valor. Este área representa la probabilidad de que la variable Z sea menor o igual al valor que se está buscando. *¿Cómo se interpreta este resultado?* El valor que se encuentra en la tabla indica la probabilidad de que ocurra un evento determinado.

A continuación, se presenta un ejercicio, en donde se muestra en un caso de estudio, como se puede analizar la probabilidad utilizando la distribución normal:

Control de calidad del peso de naranjas en una empacadora agrícola

¿Qué le pareció el ejercicio? ¿Comprendió el proceso? Este caso permite aplicar los conceptos de la distribución normal en un problema real de control de calidad agrícola.

La comprensión de la distribución normal permite optimizar la producción y garantizar que la mayor cantidad de productos cumpla con los requisitos de calidad.

5.4. Distribución exponencial

Ahora bien, para concluir con los temas de la semana vamos a revisar la segunda distribución que utiliza variables aleatorias continuas: la distribución exponencial. Esta distribución de probabilidad continua describe el tiempo entre eventos en un proceso que ocurre de manera aleatoria, constante e independiente. Es ampliamente utilizada en áreas como la agricultura, logística, biología y gestión de riesgos para modelar tiempos de espera, como el tiempo hasta la próxima falla de un equipo o el tiempo entre llegadas de clientes.

La distribución exponencial está asociada a los procesos de Poisson, que modelan el número de eventos que ocurren en un intervalo de tiempo o espacio. En este contexto:

- La distribución de Poisson modela el número de eventos en un intervalo.
- La distribución exponencial modela el tiempo entre eventos en un proceso de Poisson.

La **distribución exponencial** puede derivarse de un proceso experimental de Poisson, pero en este caso, tomando como variable aleatoria el tiempo que tarda en producirse un suceso.

Es decir, la distribución exponencial describe el tiempo entre eventos en un proceso que ocurre a una tasa constante (λ). La variable aleatoria x representa el tiempo hasta que ocurre el próximo evento.

La función de densidad de probabilidad para la distribución exponencial está dada por:

$$f(x) = \lambda e^{-\lambda x}, x \geq 0$$

De donde:

x = tiempo de eventos.

λ = tasa promedio de ocurrencia de eventos $(\lambda > 0)$

e = base de logaritmo natural

Por otro lado, la fórmula para calcular una probabilidad acumulada de la distribución exponencial es la siguiente:

$$f(x) = 1 - e^{-\lambda x}, x \geq 0$$

En el recurso que se muestra a continuación, puede revisar algunas de las propiedades más relevantes de la distribución exponencial.

[Propiedades de la Distribución exponencial]

¿Ha comprendido las propiedades de este tipo de distribución? Recuerde que la distribución exponencial es clave para modelar el tiempo entre eventos en procesos continuos, como el tiempo de espera entre llegadas de clientes o fallos en sistemas. Entre sus propiedades destacan su carácter continuo, la ausencia de memoria (la probabilidad de un evento no depende del tiempo transcurrido) y su relación directa con la distribución de Poisson. Si ha quedado alguna duda, vuelva a revisar el recurso.

Es muy importante que entienda estas propiedades, para luego estar seguro en qué caso es aplicable este tipo de distribución.

Actividades de aprendizaje recomendadas

Estimado estudiante, para reforzar los temas de esta semana, lo invito a realizar la siguiente actividad:

1. Realizar una lectura comprensiva sobre Variables aleatorias continuas y revisar la información relacionada a la distribución normal y la distribución exponencial.

   A través de esta lectura, se puede resaltar que la distribución normal es simétrica (en forma de campana) y es utilizada para modelar fenómenos naturales con datos centrados alrededor de una media, mientras que, la distribución exponencial es *asimétrica y decreciente*, utilizada principalmente para modelar tiempos entre eventos aleatorios. Se puede decir que, mientras la distribución normal describe variabilidad en datos, la distribución exponencial representa eventos con tasas constantes.

2. Apreciado estudiante, hemos concluido el estudio de la Unidad 5 y es momento de evaluar su aprendizaje. Para esto, responda a las preguntas que se presentan a continuación. El cuestionario contiene preguntas de opción múltiple, con *una o varias opciones de respuestas correctas*. Esta actividad le servirá de repaso, así como para comprobar los conocimientos adquiridos.

Autoevaluación 5

Seleccione la/las respuesta(s) correcta(s):

1. ¿Cuál de las siguientes afirmaciones es verdadera para una distribución binomial?

    a. Modela el tiempo entre eventos en un proceso continuo.
    b. Describe el número de éxitos en un número fijo de ensayos independientes.
    c. Requiere que los datos sean simétricos.
    d. Modela variables continuas.

2. ¿Cuál de las siguientes situaciones puede modelarse con la distribución de Poisson?

    a. La cantidad de clientes que llegan a una tienda por hora.
    b. La probabilidad de que un dado saque un número par en 10 lanzamientos.
    c. El tiempo entre llegadas de camiones a un almacén.
    d. El peso promedio de frutas en una muestra.

3. ¿Qué describe mejor una distribución normal?

    a. Una distribución discreta con datos asimétricos.
    b. Una distribución continua en forma de campana simétrica.
    c. Una distribución discreta para conteos en intervalos.
    d. Una distribución continua que mide tiempos entre eventos.

4. ¿Cuál de las siguientes opciones describe mejor la distribución exponencial?

    a. Mide el número de eventos en un intervalo de tiempo.
    b. Modela el número de éxitos en ensayos independientes.
    c. Modela la dispersión de datos alrededor de una media.
    d. Modela el tiempo entre eventos en un proceso continuo.

5. En una finca, se sabe que el 80% de las semillas germinan exitosamente. Si se plantan 10 semillas, ¿cuál es la probabilidad de que exactamente 8 germinen?

   a. Binomial con *n = 10, p = 0.8,* calcular *P(x = 8).*
   b. Poisson con $\lambda = 8$, calcular *P(x = 8).*
   c. Exponencial con $\lambda = 8$, calcular *P(x = 8).*
   d. Normal con $\mu = 8$, calcular *P(x = 8).*

6. Un agricultor registra un promedio de 5 plagas por hectárea. ¿Cuáles de las siguientes afirmaciones son correctas sobre este problema?

   a. Se puede usar una distribución Poisson para modelar el número de plagas.
   b. $\lambda = 5$ representa la tasa promedio de plagas por hectárea.
   c. Se puede usar una distribución exponencial para modelar este problema.
   d. El número de plagas sigue una distribución normal.

7. El peso promedio de manzanas en un lote es de 200 g, con una desviación estándar de 15 g. Si los pesos se distribuyen normalmente, ¿cuáles de las siguientes afirmaciones son correctas?

   a. El 68% de las manzanas pesan entre 185 g y 215 g.
   b. Este problema se puede modelar con una distribución normal con $\mu = 200, \sigma = 15$.
   c. Este problema se puede modelar con una distribución binomial.
   d. Los pesos de las manzanas siguen una distribución Poisson.

8. ¿Cuáles de las siguientes afirmaciones son correctas sobre las distribuciones Poisson y exponencial?

    a. La distribución Poisson modela el número de eventos, y la exponencial modela el tiempo entre ellos.
    b. Ambas distribuciones son discretas.
    c. Si los tiempos entre eventos siguen una exponencial, el número de eventos sigue una Poisson.
    d. La tasa promedio (λ) es un parámetro compartido entre ambas.

9. Un agricultor registra un promedio de 4 plagas por hectárea en su campo. ¿Cuál es la probabilidad de encontrar exactamente 6 plagas en una hectárea?

    a. *P(x = 6)* para una distribución binomial con n = 6, p = 4.
    b. *P(x = 6)* para una distribución de Poisson con $\lambda = 4$.
    c. *P(x = 6)* para una distribución normal con μ = 6, σ = 1.
    d. *P(x = 6)* para una distribución exponencial con $\lambda = 4$.

10. ¿Cuáles de las siguientes características describen correctamente la distribución Poisson?

    a. Modela el número de eventos en un intervalo fijo de tiempo o espacio.
    b. Requiere que los eventos sean dependientes entre sí.
    c. La tasa de eventos promedio (λ) es constante.
    d. Modela variables continuas.

Ir al solucionario

Resultado de aprendizaje 5:

Conoce los principios básicos de la estimación y el contraste de hipótesis estadísticas.

El presente resultado de aprendizaje se logrará mediante la Unidad 6: Distribuciones Muestrales, en donde se analizarán conceptos clave como la media muestral, el error estándar y el teorema central del límite, aplicándolos a estimaciones y contrastes en el sector de los agronegocios. A través de ejemplos prácticos, analizarán datos reales para tomar decisiones informadas sobre la calidad y manejo de productos agrícolas.

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

Durante esta semana iniciaremos el estudio de la Unidad 6: Distribuciones Muestrales, y revisaremos los *tipos de muestreo*. Aprenderemos a identificar y aplicar los diferentes métodos de muestreo, como el aleatorio simple, estratificado y por conglomerados, esenciales para garantizar que los datos recopilados representen adecuadamente a la población en análisis estadísticos.

Unidad 6. Distribuciones muestrales

Para iniciar con el estudio de esta Unidad, es muy importante reconocer el hecho de que, para analizar determinadas características de una **población,** se lo puede hacer a través de diversas **muestras** que pueden extraerse de ella.

¿Recuerda el concepto de población y muestra? Para refrescar estos conocimientos, por favor, revise nuevamente los contenidos de la semana 1, esto le ayudará a comprender la información que revisaremos a continuación.

6.1. Tipos de muestreo

Antes de revisar los tipos de muestreo, es muy importante tener claro ¿Qué es un muestreo?

El muestreo es el proceso de seleccionar un subconjunto de individuos o elementos de una población para estudiarlos y hacer inferencias sobre la población total.

En este recurso que se presenta a continuación, se muestran los principales tipos de **muestreo probabilístico**: *Muestreo aleatorio simple, muestreo aleatorio sistemático, muestreo aleatorio estratificado y muestreo por conglomerados* (De la Puente Viedma, 2018). En éstos, cada elemento de la población tiene una probabilidad conocida y no nula de ser seleccionado.

Le invito a revisar el recurso sobre los tipos de muestreo.

Tipos de muestreo

¿Han quedado claros estos cuatro tipos de muestreo? ¿revisó los ejemplos? Es importante, estimado estudiante, que revise con detenimiento este recurso, tanto para entender a qué se refiere cada tipo de muestreo, así como para entender en la práctica, cómo se pueden utilizar.

Así mismo, se cuenta con el **muestreo No Probabilístico,** en donde la selección no se basa en probabilidades conocidas, sino en criterios subjetivos o prácticos. Dentro de este tipo de muestreo está el de conveniencia, en el que se seleccionan los elementos más accesibles y por cuotas, en donde se establece un número fijo de elementos con ciertas características.

El tipo de muestreo afecta la representatividad de la muestra y, por ende, la precisión y validez de las inferencias.

6.2. Distribuciones muestrales y Teorema del límite central

Continuando con la explicación, es importante tener en cuenta que las distribuciones muestrales son una herramienta clave en la estadística inferencial, ya que nos permiten entender cómo las características de una muestra (como la media) se distribuyen, facilitando así la estimación de parámetros poblacionales y el contraste de hipótesis.

Las distribuciones muestrales cuentan con propiedades, mismas que se mencionan a continuación:

- La forma de la distribución muestral depende de la distribución de la población y del tamaño de la muestra.
- A medida que el tamaño de la muestra aumenta, la distribución muestral tiende a ser más simétrica y concentrada alrededor del valor verdadero del parámetro poblacional.

Teorema del límite central:

En cuanto al teorema del límite central, se puede mencionar que éste es un pilar fundamental de la estadística, que está rigurosamente demostrado con precisión matemática. Este teorema establece que al extraer muestras de tamaño n de una población con media μ y desviación estándar σ, y al calcular las medias muestrales, la distribución de esas medias se aproxima a una distribución normal si n es suficientemente grande (De la Puente Viedma, 2018).

No es relevante conocer la distribución original de la población, lo crucial es que, independientemente de esta, la distribución de las medias muestrales tiende a aproximarse a una distribución normal.

Este teorema indica que, independientemente de la distribución de la población, la distribución muestral de la media:

- Se aproxima a una distribución normal si el tamaño de la muestra n es suficientemente grande (n ≥ 30 es una regla común).
- Tiene una media igual a la media poblacional (μ).
- Tiene una desviación estándar igual al error estándar de la media (σ/n).

Para reforzar sus conocimientos sobre este tema, lo invito a revisar el artículo sobre El teorema del límite central.

¿Le quedó claro este tema? Es necesario puntualizar que el Teorema del Límite Central es uno de los conceptos más importantes en estadística, ya que establece que, independientemente de la distribución de una población, la distribución de las medias muestrales tiende a ser normal, cuando el tamaño de la muestra es suficientemente grande. Es muy importante que, si algo no quedó claro, vuelva a revisar la información, esto ayudará a mejorar su proceso de enseñanza aprendizaje.

6.3. La distribución muestral de la media

Ahora bien, tenemos claro que cada muestra de tamaño n que podemos extraer de una población proporciona una media. En este caso, la *distribución muestral* de la media describe cómo se distribuyen las medias de todas las muestras posibles de tamaño *n* extraídas de una población (Llinás Solano, 2017).

Si una muestra aleatoria de *n* mediciones se selecciona de una población con media μ, y desviación estándar μ, la distribución muestral de la media muestral $\bar{x}$ tendrá media μ y desviación estándar $\frac{\sigma}{\sqrt{n}}$ (Mendenhall et al., 2015).

Es importante tener en cuenta que, si la población tiene una distribución normal, la distribución muestral de $\bar{x}$ sigue también una distribución normal. Mientras que, si la población no sigue una distribución normal, pero n > 30, se aplica el *Teorema del límite central* y la distribución muestral de medias se aproximará también a la normalidad (Mendenhall et al., 2015).

La desviación estándar de $\bar{x}$ dada por $\frac{\sigma}{\sqrt{n}}$ se conoce como **error estándar de la media** (SE) (Mendenhall et al., 2015).

Es necesario reconocer que, a mayor tamaño de la muestra, menor es el error estándar (SE), lo que significa que las medias muestrales están más cerca de la media poblacional.

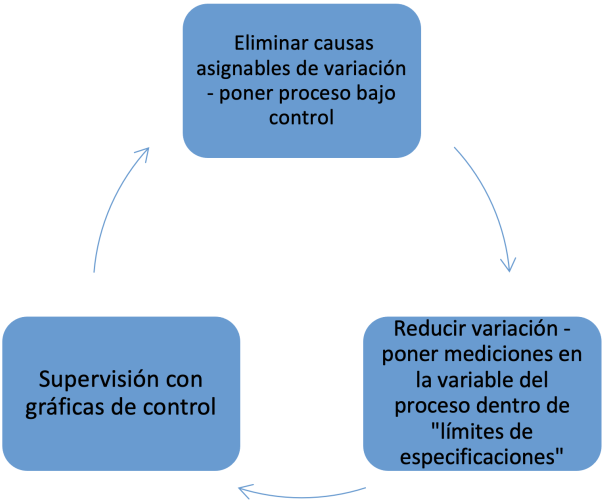
6.4. Control estadístico de procesos

Finalmente, vamos a revisar una aplicación muestral, relacionada al control estadístico de procesos. Esta es una metodología utilizada para monitorear y controlar un proceso mediante técnicas estadísticas. Su objetivo principal es garantizar que el proceso opere de manera eficiente, produciendo productos que cumplen con las especificaciones de calidad deseadas (Mendenhall et al., 2015).

En la figura posterior, se presenta cómo se da el control estadístico de procesos.

Figura 5

Control estadístico de procesos



Nota. Adaptado de *Introducción a la probabilidad y estadística (14ª. edición)* [Ilustración], por Mendenhall, W., Beaver, R. y Beaver, B., 2015, CENGAGE Learning, CC BY 4.0.

Información importante:

- Si la causa de un cambio en la variable **es asignable**: puede ser localizada y corregida.
- Cambios fortuitos debido a la alteración del ambiente de producción (no controlable): **variación aleatoria.**
- Si la variación en la variable de un proceso es solo aleatoria: **proceso en control.**

¿Quedó claro este tema? Es importante recapitular que el control estadístico de procesos es una herramienta indispensable para mantener y mejorar la calidad en sistemas de producción. Al identificar y gestionar variaciones en los procesos, éste permite a las organizaciones tomar decisiones informadas, mejorar la satisfacción del cliente y reducir costos operativos.

Actividades de aprendizaje recomendadas

Estimado estudiante, para reforzar los temas de esta unidad, lo invito a realizar la siguiente actividad:

1. Realiza una lectura comprensiva acerca de las Distribuciones muestrales.

   La lectura recomendada refuerza el hecho que, las distribuciones muestrales describen cómo se comportan las estadísticas de una muestra al extraer múltiples muestras de una población y que son fundamentales para la inferencia estadística. Éstas permiten estimar parámetros poblacionales y calcular probabilidades, así como incluyen la distribución de la *media muestral y la distribución de proporciones, que se basan en el Teorema del Límite Central.*

2. Resuelva el siguiente ejercicio:

   Imagine que trabaja para una empresa agrícola que evalúa el rendimiento de cultivos en diferentes parcelas. La empresa cuenta con los datos de rendimiento por hectárea de una población grande (Anexo 1).

   Realizar lo siguiente:

   a. Seleccionar 30 valores al azar de la población para construir una muestra y calcular la media muestral.
   b. Repetir este proceso al menos 10 veces y registre las medias muestrales.
   c. Para reconocer la distribución muestral, construya un histograma que represente las medias obtenidas en las muestras y reflexione sobre la forma del histograma.
   d. ¿Cómo se reconoce el impacto del tamaño de la muestra? Realice el mismo procedimiento con muestras de tamaños 10 y 50.
   e. Analice cómo cambia la dispersión y la forma de la distribución muestral al variar el tamaño de la muestra.

f. Finalmente, responda a la siguiente pregunta: ¿Cómo respalda esta actividad el Teorema del Límite Central?

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

¡Excelente trabajo! Esta actividad nos permite ver con claridad cómo a medida que tomamos muestras y calculamos sus medias, las distribuciones muestrales tienden a tener una forma aproximadamente normal, incluso si la población original no lo es, tal como predice el Teorema del Límite Central. Además, en esta actividad pudo constatar cómo el hecho de aumentar el tamaño de la muestra, la variabilidad (desviación estándar de las medias) disminuyó, lo que refleja una mayor precisión en las estimaciones. Este ejercicio nos permite comprender cómo realizar inferencias estadísticas confiables y cómo el tamaño de la muestra afecta los resultados.

3. Apreciado estudiante, hemos concluido el estudio de la Unidad 6 y es momento de evaluar su aprendizaje. Para esto, responda a las preguntas que se presentan a continuación. El cuestionario contiene preguntas de opción múltiple, *con una o varias opciones de respuestas correctas*. Esta actividad le servirá de repaso, así como para comprobar los conocimientos adquiridos.

Autoevaluación 6

Seleccione la/las respuesta(s) correcta(s):

1. ¿Cuál de las siguientes afirmaciones describe el muestreo aleatorio simple?

   a. Se seleccionan elementos a intervalos regulares de una lista ordenada.
   b. Todos los elementos de la población tienen la misma probabilidad de ser seleccionados.
   c. La población se divide en subgrupos homogéneos y se eligen muestras de cada uno.

d. Se seleccionan elementos según su accesibilidad.

2. ¿Cuál es la principal diferencia entre el muestreo estratificado y el muestreo por conglomerados?

   a. En el estratificado, los elementos se seleccionan en intervalos fijos.
   b. En el conglomerado, se seleccionan elementos individuales de cada subgrupo.
   c. En el conglomerado, la selección es completamente al azar.
   d. En el estratificado, los subgrupos son homogéneos; en el conglomerado, son heterogéneos.

3. ¿Qué describe una distribución muestral?

   a. La variabilidad de los datos dentro de una población.
   b. La distribución de un estadístico calculado a partir de todas las posibles muestras de un tamaño específico.
   c. El comportamiento de una variable aleatoria en una población infinita.
   d. La frecuencia relativa de los datos observados en una muestra.

4. ¿Cuál de las siguientes afirmaciones describe mejor el Teorema del Límite Central?

   a. La distribución muestral de la media se aproxima a una distribución normal si n es suficientemente grande.
   b. La media de una población es siempre igual a la media de las muestras.
   c. Todos los datos de una población tienden a ser simétricos si el tamaño muestral es grande.
   d. La desviación estándar de las muestras es constante independientemente de su tamaño.

5. ¿Por qué el Teorema del Límite Central es crucial en estadística inferencial?

    a. Permite realizar estimaciones y pruebas de hipótesis usando métodos basados en la distribución normal.
    b. Justifica el uso de muestras para inferir sobre poblaciones no normales.
    c. Asume que todas las poblaciones tienen distribución normal.
    d. Asegura que todas las muestras tengan el mismo tamaño.

6. De las siguientes afirmaciones, elija las que son verdaderas sobre la distribución muestral de la media.

    a. Tiene una media igual a la media poblacional (μ).
    b. Es siempre normal, independientemente del tamaño de la muestra.
    c. Su error estándar disminuye al aumentar el tamaño de la muestra.
    d. Su desviación estándar es igual a la desviación estándar poblacional.

7. ¿Cuál de las siguientes es una propiedad de la distribución muestral de la media?

    a. Su media es igual a la media poblacional.
    b. Su desviación estándar es igual a la desviación estándar poblacional.
    c. Su forma es siempre normal, independientemente del tamaño de la muestra.
    d. Se calcula únicamente para datos cualitativos.

8. Si $n = 36$ y la desviación estándar poblacional es 15, ¿cuál es el error estándar de la media?

    a. 15
    b. 2.5

c. 3.0
d. 1.5

9. Si un proceso está bajo control estadístico, ¿qué implica?

    a. Todas las muestras están dentro de los límites de control.
    b. No hay causas especiales de variación.
    c. La calidad del producto siempre cumple con las especificaciones.
    d. La variación en los datos es únicamente aleatoria.

10. ¿Cuáles de las siguientes herramientas son comunes en el Control Estadístico de Procesos?

    a. Gráficas de control.
    b. Histogramas para visualizar la distribución de datos.
    c. Diagramas de flujo.
    d. Diagramas de Pareto para identificar problemas clave.

Ir al solucionario

Contenidos, recursos y actividades de aprendizaje recomendadas

Semana 14

Antes de iniciar con el estudio de esta Unidad, es necesario tener en cuenta que, en las unidades anteriores se exploraron los fundamentos de la estadística descriptiva, como medidas de tendencia central, dispersión y visualización de datos. Ahora, con el tema de intervalos de confianza, se inicia el estudio de la inferencia estadística, una herramienta clave para analizar datos muestrales y hacer estimaciones sobre parámetros poblacionales. Este enfoque permite considerar la incertidumbre inherente al muestreo, fortaleciendo las aplicaciones estadísticas en contextos prácticos, como los agronegocios.

La **inferencia estadística** se ocupa de tomar decisiones o hacer predicciones sobre *parámetros*, es decir, las medidas numéricas descriptivas que caracterizan a una población, entre las que destacan: media poblacional (μ), desviación estándar (σ) y proporción binomial p (Mendenhall et al., 2015).

En este sentido, los métodos para hacer inferencias acerca de parámetros poblacionales pueden considerar tanto a la *estimación*, como a la *prueba de hipótesis. ¿A qué se refieren estas categorías?*

La **estimación** se refiere al hecho de predecir o estimar el valor del parámetro, mientras que, la **prueba de hipótesis** se refiere a tomar una decisión acerca del valor de un parámetro, con base en alguna idea preconcebida acerca de cuál podría ser su valor (Mendenhall et al., 2015).

Dicho esto, para estimar el valor de un parámetro poblacional se puede usar información de la muestra, a través de un **estimador.** *¿Qué es un estimador?* Es una regla, que menciona cómo calcular una estimación basada en información de la muestra (Mendenhall et al., 2015).

Estos estimadores se usan de dos formas distintas. Por un lado, está la estimación puntual, y por otro la estimación de intervalo. Ahora, fijaremos nuestra atención en el primer grupo.

Unidad 7. Intervalo de confianza

7.1. Estimación puntual

La estimación puntual es un valor único que se utiliza para aproximar un parámetro poblacional desconocido (Llinás Solano, 2017). Por ejemplo, la media muestral ($\underline{x}$) es una estimación puntual de la media poblacional (μ). Si bien proporciona una estimación precisa, no indica cuánta incertidumbre existe alrededor de este valor.

Por ejemplo: Si se mide el peso promedio de 50 frutas y se obtiene $\underline{x}$ = 200 g, esto es una estimación puntual de μ, el peso promedio de todas las frutas en la población.

Es muy importante saber cómo se comporta el estimador en un muestreo repetido, que está descrito por su distribución muestral. Esta distribución, justamente, provee de información que se puede usar para seleccionar al mejor estimador (Mendenhall et al., 2015). A continuación, se mencionan las características relevantes para hacer esta selección:

1. La distribución muestral del estimador puntual, debe estar centrada sobre el valor verdadero del parámetro que se estimará. El estimador insesgado es aquel que no debe subestimar o sobreestimar de manera constante al parámetro de interés (Llinás Solano, 2017).
2. La dispersión de la distribución muestral debe ser tan pequeña como sea posible.

La distancia entre la estimación y el parámetro estimado se denomina: **error de estimación** (Mendenhall et al., 2015).

Hay que tener en cuenta que, para cualquier estimador puntual con una distribución normal, la regla empírica dice que aproximadamente el 95% de todas las estimaciones puntuales estarán a no más de dos (1.96) desviaciones estándar de la media de esa distribución (Mendenhall et al., 2015).

Para reforzar sus conocimientos en torno a la estimación puntual y para conocer a qué se refiere la estimación por intervalos, revise sobre la Estimación puntual y estimación por intervalos.

¿Le quedaron claros estos contenidos? Es necesario tener claro que, la estimación es una herramienta que permite aproximar parámetros poblacionales con base en datos muestrales. La *estimación puntual*, por su

parte, utiliza un único valor (como la media muestral) para representar el parámetro poblacional, mientras que, la *estimación por intervalos* amplía esta idea proporcionando un rango de valores dentro del cual es probable que se encuentre el parámetro poblacional, con un nivel de confianza específico (como el 95%). Recuerde, una lectura comprensiva le ayudará a lograr el cometido de estos recursos, que no es otro que complementar y reforzar la explicación, de manera que podamos continuar con el estudio.

7.2. Construcción e interpretación de intervalos de confianza.

Para continuar con este tema, es necesario conocer que cuando la distribución muestral de un estimador puntual es aproximadamente normal, se puede construir un estimador de intervalo o un **intervalo de confianza**. Un intervalo de confianza amplía la estimación puntual para incluir un rango de valores en el cual se espera que se encuentre el parámetro poblacional, con un nivel de confianza específico ($1-\alpha$, donde α es el nivel de significancia) (Llinás Solano, 2017).

Para construir un intervalo de confianza, se requiere:

1. La estimación puntual del parámetro de interés.
2. La desviación estándar de la muestra.
3. El tamaño de la muestra.
4. Un valor crítico que depende del nivel de confianza deseado (normalmente obtenido de una distribución t o z).

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

De donde:
$\bar{x}$ = media muestral.

$z_{\alpha/2}$ = valor crítico de la distribución normal para el nivel de confianza deseado (por ejemplo, 1.96 para un 95% de confianza)

σ = desviación estándar de la muestra.

n = tamaño de la muestra.

¿Cómo se interpreta un intervalo de confianza? Un intervalo de confianza, por ejemplo, del 95% indica que, en el largo plazo, el 95% de los intervalos construidos de esta manera contendrán la verdadera media poblacional.

En el recurso a continuación, se presenta un ejemplo para que pueda constatar cómo se puede calcular un intervalo de confianza.

Intervalos de confianza

¿Le quedó claro este ejemplo? Recuerde que, los intervalos de confianza son una herramienta clave para estimar parámetros poblacionales con un nivel de certeza definido. Proveen no solo una estimación puntual, sino un rango que refleja la incertidumbre inherente al muestreo.

7.3. Intervalos de confianza de muestra grande (1- α) para una media poblacional μ

Por otra parte, cuando el tamaño de la muestra es grande (n ≥ 30), el Teorema del Límite Central asegura que la distribución de la media muestral se aproxima a una distribución normal, incluso si la población original no es normal. En este caso, el intervalo de confianza se calcula utilizando la fórmula anterior.

Hay algunas consideraciones importantes que se deben tener en cuenta (Llinás Solano, 2017):

- Si la desviación estándar poblacional (σ) es desconocida, se utiliza la desviación estándar muestral (s).
- Este método es confiable para muestras grandes, independientemente de la forma de la población.

Ejemplo:

Si una muestra de n = 100 tiene una media muestral de 50 kg, una desviación estándar muestral de 5 kg, y se desea un intervalo de confianza del 90% (z = 1.645):

$$IC = 50 \pm 1.645 \frac{5}{\sqrt{100}} = 50 \pm 0.8225$$

$$IC = (49.18, 50.82)$$

¿Qué nos dice este intervalo de confianza? Este intervalo de confianza lo que menciona es que hay un 90% de confianza de que el peso promedio poblacional esté entre 49.18 kg y 50.82 kg.

7.4. Intervalos de confianza para la diferencia entre dos medias poblacionales

Ya para finalizar esta Unidad, vamos a revisar que cuando se comparan dos poblaciones, el intervalo de confianza para la diferencia entre sus medias $(\mu 1 - \mu 2)$ es útil para determinar si existe una diferencia estadísticamente significativa entre ellas (Llinás Solano, 2017).

$$IC = (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

De donde:

$\bar{x}_1, \bar{x}_2$ = Medias muestrales de las dos medias
σ_1^2, σ_2^2 = Desviaciones estándar de las dos poblaciones. El error estándar también se puede estimar como se menciona a continuación, cuando los tamaños muestrales son grandes:

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

n_1, n_2 = Tamaño de las dos muestras

En el ejemplo que se muestra a continuación, se presenta un ejemplo para que pueda constatar cómo se puede calcular un intervalo de confianza para la diferencia entre dos medias poblacionales.

Intervalos de confianza para la diferencia entre dos medias

¿Con este ejemplo le quedó claro este tipo de intervalo? Recuerde que, este tipo de intervalo puede ser utilizado para comparar dos poblaciones y determinar si hay una diferencia significativa entre sus medias.

Actividades de aprendizaje recomendadas

Estimado estudiante, para reforzar los temas de esta unidad, lo invito a realizar las siguientes actividades:

1. Realizar una lectura comprensiva sobre los Intervalos de confianza.

   Recuerde que, los intervalos de confianza permiten estimar un parámetro poblacional a partir de una muestra, proporcionando un rango de valores en el que probablemente se encuentra el verdadero valor del parámetro con un nivel de confianza determinado. Es muy importante su interpretación correcta, más aún si de esto depende la toma de decisiones, esto evitará emitir conclusiones erróneas sobre la población estudiada.

2. Realizar el siguiente quiz, en donde pondrá en práctica los contenidos abordados en esta unidad. ¿Está listo? Vamos a comenzar:
   El Reto de los Intervalos de Confianza

   Felicidades por completar "El Reto de los Intervalos de Confianza" En esta actividad, usted reforzó sus habilidades para calcular, interpretar y aplicar intervalos de confianza en diferentes contextos. Recuerde, estos intervalos permiten incorporar incertidumbre en las estimaciones, haciendo los análisis más precisos. ¡Continúa practicando para dominar estos conceptos!

3. Apreciado estudiante, hemos concluido el estudio de la Unidad 7 y es momento de evaluar su aprendizaje. Para esto, responda a las preguntas que se presentan a continuación. El cuestionario contiene preguntas de opción múltiple, con *una o varias opciones de respuestas correctas*. Esta actividad le servirá de repaso, así como para comprobar los conocimientos adquiridos.

[Autoevaluación 7]

Seleccione la/las respuesta(s) correcta(s):

1. ¿Qué es una estimación puntual?

    a. Un rango de valores que incluye un parámetro poblacional con un nivel de confianza.
    b. Un único valor utilizado para aproximar un parámetro poblacional.
    c. La diferencia entre dos medias muestrales.
    d. Una herramienta para comparar dos poblaciones.

2. ¿Cuál de los siguientes componentes es necesario para construir un intervalo de confianza para la media poblacional?

    a. Media muestral, tamaño de muestra y nivel de confianza.
    b. Media muestral, mediana y desviación estándar.
    c. Desviación estándar, moda y rango.
    d. Mediana, tamaño de muestra y nivel de confianza.

3. ¿Qué significa un intervalo de confianza del 95%?

    a. Que el parámetro poblacional está definitivamente dentro del intervalo calculado.
    b. Que el 95% de los datos de la población están dentro del intervalo.
    c. Que el 95% de los intervalos construidos contendrán el parámetro poblacional verdadero.
    d. Que la media muestral se encuentra dentro del intervalo.

4. ¿Por qué se puede utilizar un intervalo de confianza basado en la distribución normal para muestras grandes?

   a. Porque la distribución de la población siempre es normal.
   b. Porque la media muestral sigue una distribución normal para tamaños grandes, según el Teorema del Límite Central.
   c. Porque la muestra siempre es representativa de la población.
   d. Porque los datos son cualitativos.

5. Si una muestra grande tiene una media de 50, un error estándar de 2 y un nivel de confianza del 95%, ¿cuáles son los pasos correctos para calcular el intervalo?

   a. Multiplicar z = 1.96 por el error estándar.
   b. Sumar y restar el margen de error a la media muestral.
   c. Dividir la media por el error estándar.
   d. Usar un nivel de confianza mayor para reducir el intervalo.

6. ¿Cuáles de las siguientes afirmaciones son ciertas para intervalos de confianza sobre la diferencia entre dos medias?

   a. Se utiliza el error estándar combinado de ambas muestras.
   b. Siempre se calcula con un nivel de confianza del 99%.
   c. El intervalo indica dónde probablemente se encuentra la diferencia entre las medias poblacionales.
   d. Se utiliza únicamente para datos cualitativos.

7. ¿Qué sucede cuando se incrementa el nivel de confianza de un intervalo?

   a. El intervalo se vuelve más estrecho.
   b. El margen de error disminuye.
   c. El intervalo se vuelve más amplio.
   d. Hay mayor certeza de que el parámetro poblacional está incluido.

8. Un intervalo de confianza del 95% para la diferencia de medias entre dos poblaciones es (1.5, 4.8). ¿Qué conclusiones son válidas?

   a. La diferencia promedio entre las poblaciones está entre 1.5 y 4.8.
   b. Existe evidencia estadística de que las medias de las dos poblaciones son diferentes.
   c. No hay diferencia significativa entre las dos poblaciones.
   d. La diferencia de medias siempre es 4.8.

9. ¿Qué representa el intervalo de confianza para la diferencia entre dos medias?

   a. Un rango de valores donde probablemente se encuentran las medias muestrales.
   b. Un rango de valores donde probablemente se encuentra la diferencia entre las medias poblacionales.
   c. La suma de los errores estándar de las dos muestras.
   d. La probabilidad de que las muestras sean iguales.

10. Un intervalo de confianza del 95% para la media poblacional es (10.5, 15.3). ¿Qué se puede concluir?

   a. Hay un 95% de confianza de que la media poblacional está entre 10.5 y 15.3.
   b. La media poblacional está definitivamente dentro de este rango.
   c. El 95% de los datos de la población están dentro de este rango.
   d. Este intervalo incluye el valor verdadero de la media poblacional en el 95% de los casos.

Ir al solucionario

Semana 15

Finalmente, en esta semana exploraremos la Unidad 8: Prueba de Hipótesis, y nos enfocaremos en dos temas clave: *procedimientos de una prueba estadística de hipótesis*, en donde aprenderemos los pasos sistemáticos para evaluar afirmaciones sobre parámetros poblacionales, así como, revisaremos el tema *prueba de muestra grande acerca de una media poblacional*, en donde aplicaremos estos procedimientos en escenarios donde el tamaño de la muestra permite utilizar la distribución normal, desarrollando habilidades prácticas para resolver problemas relacionados con los agronegocios ¡Vamos a iniciar!

Unidad 8. Prueba de hipótesis

Estimado estudiante, estamos por concluir el estudio de esta asignatura y para hacerlo, vamos a hablar de la prueba de hipótesis. *¿En qué consiste una prueba de hipótesis?* Pues, una prueba de hipótesis se encarga de recopilar datos de una muestra y evaluarlos (Llinás Solano, 2017). Luego de esto, el estadístico decide si existen o no pruebas suficientes, basándose en el análisis de los datos, para rechazar la hipótesis nula.

Para comprender la prueba de hipótesis es muy importante que se familiarice con cierta terminología, para asegurar la completa comprensión de este tema. Por favor, revise la infografía que se muestra a continuación.

[Conceptos clave para la prueba de hipótesis]

¿Le quedaron claros estos conceptos? Si es necesario, vuelva a revisarlos. Es muy importante, para que comprenda las partes de la prueba de hipótesis que veremos a continuación.

8.1. Procedimientos de una prueba estadística de hipótesis

Una vez que ha comprendido la terminología compartida en el último recurso, vamos a revisar el procedimiento de una prueba de hipótesis. Este procedimiento tiene que ver con un conjunto de pasos sistemáticos para evaluar una afirmación sobre un parámetro poblacional (como una media) utilizando datos muestrales. Éste permite determinar si hay evidencia estadística suficiente para rechazar una hipótesis inicial (hipótesis nula).

En la infografía que se presenta a continuación, podrá revisar sus pasos:

Pasos del Procedimiento de una Prueba de Hipótesis

¿Qué le parecieron estos pasos? ¿los ha comprendido? Es importante reconocer que la prueba de hipótesis es una herramienta fundamental para validar afirmaciones sobre parámetros poblacionales. Este procedimiento permite realizar pruebas sobre una media (o comparar dos medias poblacionales), permitiendo tomar decisiones basadas en evidencia estadística.

8.2. Una prueba de muestra grande acerca de una media poblacional

Por otro lado, la prueba de muestra grande acerca de una media poblacional es una herramienta estadística utilizada para evaluar si la media de una población es igual a un valor específico (μ_0), basado en los datos obtenidos de una muestra grande ($n \geq 30$). Esta prueba es especialmente útil cuando se conoce la desviación estándar de la población (σ) o cuando el tamaño de la muestra es suficiente para aplicar el Teorema del Límite Central, que permite asumir que la distribución de la media muestral se aproxima a una distribución normal.

Es importante comentar que, el procedimiento para realizar la prueba es el mismo que se explicó en el último recurso. *¿Lo recuerda?* Puede darle un vistazo, en caso de que algún paso no lo recuerde.

Una vez que ha revisado nuevamente este procedimiento, es importante destacar el paso en donde se calcula el **valor *p*.** Recuerde que, este valor p o también denominado nivel de significancia, es el valor más pequeño de α para el cual H_0 se puede rechazar (Mendenhall et al., 2015). Pero ¿qué tan pequeño debe ser el valor de p antes de que se decida rechazar H_0?

Se puede usar una escala de cálculo para clasificar los resultados, en este sentido.

a. Si el valor de p es menor que 0.01, H_0 se rechaza, por lo que los resultados son **altamente significativos,**
b. Si el valor de p está entre 0.01 y 0.05, H_0 se rechaza, aquí los resultados son **estadísticamente significativos,**
c. Si el valor de p está entre 0.05 y 0.10, H_0 por lo general no se rechaza, en este caso los resultados son solo tendentes hacia **significancia estadística,**
d. Si el valor de p es mayor que 0.10, H_0 no se rechaza, en este caso los resultados no son **estadísticamente significativos** (Mendenhall et al., 2015).

Es importante destacar que, como lo hemos visto, en una prueba de hipótesis las decisiones se basan en datos muestrales, lo que introduce la posibilidad de cometer errores. Estos errores ocurren porque la muestra puede no representar perfectamente a la población, o debido a la naturaleza probabilística de las pruebas estadísticas. Los dos tipos principales de errores son el error tipo I y el error tipo II. *¿A qué se refieren estos errores?*

El **error tipo I** (α) ocurre cuando se rechaza la hipótesis nula (H_0) cuando en realidad es verdadera. Este error implica concluir que hay un efecto o diferencia cuando no lo hay. Por otra parte, el **error tipo II** (β) ocurre si se acepta la hipótesis nula (H_0) cuando es falsa y alguna hipótesis alternativa es verdadera. Este error implica no detectar un efecto o diferencia que realmente existe (Llinás Solano, 2017, Mendenhall et al., 2015).

Actividades de aprendizaje recomendadas

Estimado estudiante, para reforzar los temas de esta unidad, lo invito a realizar las siguientes actividades:

1. Realizar una lectura comprensiva acerca de las Pruebas de hipótesis con una muestra.

   Es importante recordar que, las pruebas de hipótesis con una muestra permiten evaluar afirmaciones sobre un parámetro poblacional utilizando datos muestrales. ¿Qué necesitamos para esto? Se requiere establecer la hipótesis nula (H_0) y la hipótesis alternativa (H_1), para poder determinar si hay suficiente evidencia estadística para rechazar H_0. Factores como el nivel de significancia y el valor p son esenciales para la interpretación.

2. Lea el siguiente caso y complete la información en la tabla que se muestra a continuación:

   Una empresa láctea afirma que el tiempo promedio de conservación de su yogur es de 20 días. Una muestra de n = 50 productos, tiene una media de 19.5 días y una desviación estándar de 2. Los distribuidores quieren verificar si el tiempo promedio es diferente al declarado.

Responder a las siguientes preguntas:

- Plantear la H_0 y H_a, considerando que el problema indica una prueba bilateral.
- Reflexionar: ¿Cómo afecta esta decisión a los consumidores y a la empresa?

Tabla a completar

H_0	
H_a	
¿Cómo afecta esta decisión a los consumidores y la empresa?	

Nota: copie la tabla en un Word o cuaderno para rellenar.

¡Felicidades por completar el ejercicio! En esta actividad pudo practicar la formulación de hipótesis nula y alternativa, lo cual es fundamental para realizar pruebas estadísticas. Es muy importante que recuerde que, la hipótesis nula (H_0) representa la afirmación inicial que se busca probar, mientras que la hipótesis alternativa (H_a) plantea una diferencia o cambio. En el caso del yogur, rechazar H_0 significa que el tiempo promedio de conservación difiere del valor declarado, lo que afecta la confianza del consumidor.

3. Apreciado estudiante, hemos concluido el estudio de la Unidad 8 y es momento de evaluar su aprendizaje. Para esto, responda a las preguntas que se presentan a continuación. El cuestionario contiene preguntas de opción múltiple, *con una o varias opciones de respuestas correctas*. Esta actividad le servirá de repaso, así como para comprobar los conocimientos adquiridos.

Autoevaluación 8

Seleccione la/las respuesta(s) correcta(s):

1. ¿Qué es una prueba de hipótesis?

   a. Un procedimiento para calcular una media muestral.
   b. Una herramienta para comparar datos cualitativos.
   c. Un procedimiento para tomar decisiones sobre un parámetro poblacional usando datos muestrales.
   d. Una fórmula para estimar parámetros poblacionales.

2. ¿Cuál de las siguientes es una afirmación sobre la hipótesis nula (H_0)?

   a. Representa la afirmación que se busca demostrar.
   b. Se presume verdadera hasta que se demuestre lo contrario.
   c. Indica una diferencia significativa entre parámetros poblacionales.
   d. Solo se utiliza en pruebas bilaterales.

3. ¿Qué representa el nivel de significancia (α) en una prueba de hipótesis?

   a. La probabilidad de cometer un error tipo I.
   b. La probabilidad de cometer un error tipo II.
   c. El intervalo en el que se encuentra el parámetro poblacional.
   d. La media de la muestra.

4. ¿Qué es la región crítica en una prueba de hipótesis?

   a. El rango de valores donde se acepta H_0.
   b. El rango de valores donde se rechaza H_0.
   c. La probabilidad de cometer un error tipo II.
   d. El valor exacto del estadístico de prueba.

5. ¿Qué indica un valor p en una prueba de hipótesis?

   a. La probabilidad de que H_0 sea verdadera.
   b. La probabilidad de aceptar H_a.
   c. La probabilidad de que el parámetro poblacional esté en la región crítica.
   d. La probabilidad de observar un estadístico tan extremo, suponiendo que H0 es verdadera.

6. ¿Cuándo se rechaza H_0 en una prueba de hipótesis?

   a. Cuando el estadístico de prueba está en la región crítica.
   b. Cuando el valor p es mayor que α.
   c. Cuando el nivel de significancia es bajo.
   d. Cuando el tamaño de la muestra es grande.

7. Si el valor p de una prueba es 0.03 y $\alpha = 0.05$, ¿qué decisión se toma?

   a. Se acepta H_0.
   b. Se rechaza H_0.
   c. No se puede tomar una decisión.
   d. Se ajusta el nivel de significancia.

8. ¿Cuál es una condición para realizar una prueba de muestra grande acerca de una media poblacional?

   a. La muestra debe ser menor que 30.
   b. La población debe ser normal.
   c. La desviación estándar de la población debe ser conocida.
   d. El tamaño de la muestra no afecta el resultado.

9. ¿Cuándo se utiliza la fórmula del estadístico z en una prueba de hipótesis?

   a. Cuando el tamaño de la muestra es pequeño (n < 30).
   b. Cuando la población no es normal.

c. Cuando la desviación estándar poblacional es conocida.
d. Cuando se comparan dos proporciones.

10. En una prueba z, si el valor p es 0.01 y α = 0.05, ¿qué decisión se toma?

a. Se rechaza H_0.
b. No se rechaza H_0.
c. El resultado es inconcluso.
d. Se debe aumentar el tamaño de la muestra.

Ir al solucionario

Resultado de aprendizaje 4 y 5:

- Conoce y aplica leyes de las probabilidades para modelizar variables aleatorias discretas y continuas.
- Conoce los principios básicos de la estimación y el contraste de hipótesis estadísticas.

Contenidos, recursos y actividades de aprendizaje recomendadas

Semana 16

Actividades finales del bimestre

Estimado estudiante, al concluir este segundo bimestre, es momento de prepararse para la evaluación bimestral. Para ello, le recomiendo revisar nuevamente los apuntes y resúmenes que ha elaborado a lo largo de estas semanas, con el objetivo de refrescar sus conocimientos e identificar los temas que puedan requerir mayor atención. Para estos últimos, es fundamental dedicar tiempo adicional y repasar los recursos compartidos previamente.

Aproveche el horario de tutorías de esta semana para plantear sus dudas al docente y recibir el apoyo necesario en los temas que considere más complejos. Durante la tutoría de la semana 16, el docente reforzará los contenidos que presenten mayor dificultad, basándose en aquellos temas que han sido compartidos por los estudiantes. Por ello, su asistencia es clave para optimizar este espacio y resolver cualquier duda. Si no le es posible asistir, el video de la tutoría estará disponible posteriormente a través de un anuncio académico.

¡Hacemos una última actividad! Lo invito a realizar el siguiente quiz que le ayudará a realizar un pequeño repaso de los contenidos del bimestre.

Breakout estadística descriptiva

¿Qué le ha parecido el juego? ¿ha respondido correctamente a las preguntas propuestas? Es muy importante que refuerce este ejercicio y como se mencionó anteriormente, haga una revisión de los contenidos que usted considera necesitan un nuevo repaso, esto le asegurará un buen desenvolvimiento en la evaluación bimestral.

4. Autoevaluaciones

Autoevaluación 1

Pregunta	Respuesta	Retroalimentación
1	b	En estadística, la población representa el conjunto completo que se quiere analizar o sobre el cual se requiere obtener conclusiones.
2	c	Una muestra debe ser representativa para reflejar adecuadamente las características de la población.
3	b	Una muestra es un grupo seleccionado de una población para facilitar el análisis estadístico.
4	c	La variedad de cultivos es una variable cualitativa nominal porque representa categorías sin un orden inherente.
5	a y c	Las variables cualitativas describen características o atributos no numéricos, como variedades o colores. Las hectáreas cultivadas y el peso promedio son cuantitativas.
6	b y d	Las variables cuantitativas continuas pueden tomar valores en un rango continuo, como alturas y rendimientos.
7	c	Las variables cuantitativas continuas pueden tomar cualquier valor dentro de un rango, como la distancia en kilómetros.
8	a	El precio promedio es cuantitativo continuo porque puede tomar valores decimales, como 1.25 usd o 1.30 usd por kilogramo.
9	b	El peso promedio de todos los frutos cosechados representa un análisis de la población, ya que incluye a todos los frutos.
10	a y b	La población incluye todas las fincas de la región, mientras que las 100 fincas seleccionadas constituyen la muestra representativa para el estudio.

Ir a la autoevaluación

Autoevaluación 2

Pregunta	Respuesta	Retroalimentación
1	b	Una tabla de frecuencias organiza los datos para mostrar cuántas veces ocurren valores o categorías específicas, facilitando el análisis.
2	a	Un histograma representa la distribución de datos numéricos continuos y utiliza barras contiguas para reflejar la continuidad de los intervalos.
3	c	Una gráfica de pastel es ideal para mostrar proporciones o porcentajes del total, destacando cómo se distribuyen las ventas entre los productos.
4	a y d	Las gráficas de pastel son útiles para representar proporciones de un total en forma visual y clara, pero no son adecuadas para datos numéricos continuos.
5	a y c	La gráfica de líneas es ideal para analizar cambios en datos numéricos continuos a lo largo del tiempo, pero no se usa para comparar proporciones ni frecuencias acumuladas.
6	c	Una gráfica de líneas conecta puntos secuenciales y es ideal para analizar cambios en datos a lo largo del tiempo, como precios.
7	b	La gráfica de barras es ideal para datos categóricos, como los tipos de cultivos, mientras que el histograma se usa para datos continuos.
8	c	El histograma es adecuado para representar la distribución de datos agrupados en intervalos, como edades, ya que muestra frecuencias en barras contiguas.
9	a, b y d	Un histograma es específico para datos continuos y utiliza barras contiguas para reflejar intervalos, mientras que la gráfica de barras compara frecuencias de categorías con barras separadas.
10	a y c	En gráficas de barras para variables numéricas, las barras están separadas y su altura representa frecuencias o valores asociados a categorías discretas.

Ir a la autoevaluación

Autoevaluación 3

Pregunta	Respuesta	Retroalimentación
1	c	La mediana divide al conjunto de datos en dos partes iguales, siendo el valor central en un conjunto ordenado.
2	b	La media se calcula sumando los valores y dividiendo por el número total de datos
3	a	La desviación estándar es la raíz cuadrada de la varianza.
4	c	El tercer cuartil ($Q3$) marca el límite por debajo del cual se encuentra el 75% de los datos.
5	a, b y d	La media es el promedio, considera todos los valores, pero puede verse afectada por valores atípicos.
6	a, b y c	Un diagrama de cajas muestra la dispersión, los cuartiles, la mediana y ayuda a identificar valores atípicos.
7	d	El coeficiente de variación permite comparar la variabilidad relativa de dos conjuntos de datos con diferentes escalas.
8	a y b	La desviación estándar es una medida de dispersión que indica cuánto se alejan, en promedio, los datos de la media. Se calcula tomando la raíz cuadrada de la varianza. No es una medida de posición y sí puede verse afectada por valores extremos.
9	b y d	La mediana es resistente a valores atípicos y proporciona una mejor representación del centro en distribuciones sesgadas. El rango intercuartílico mide la dispersión central y también es resistente a valores extremos. La desviación estándar puede utilizarse, pero es sensible a valores atípicos.
10	a y d	En distribuciones simétricas sin valores atípicos, la media aritmética es una buena medida de tendencia central, y la desviación estándar es adecuada para medir la dispersión. El rango también puede proporcionar información sobre la amplitud de los datos. La mediana y el rango intercuartílico son más útiles en distribuciones asimétricas o con valores atípicos.

Ir a la autoevaluación

Autoevaluación 4

Pregunta	Respuesta	Retroalimentación
1	b	La probabilidad mide la certeza de que ocurra un evento, expresada como un número entre 0 (imposible) y 1 (seguro).
2	d	Un evento aleatorio tiene resultados inciertos antes de realizar el experimento, pero cada posible resultado tiene una probabilidad asociada.
3	c	Hay 3 números impares (1, 3, 5) en un dado de 6 caras. Por lo tanto, la probabilidad es 6. Por lo tanto, la probabilidad es de 3/6 = 0.5
4	b	Para eventos mutuamente excluyentes, $P(A \cup B) = P(A) + P(B)$. Por lo tanto, $P(A \cup B) = 0.3 + 0.5 = 0.8$
5	c	Los eventos mutuamente excluyentes no pueden ocurrir simultáneamente.
6	a y b	Un espacio muestral es el conjunto de todos los resultados posibles (a), y la suma de las probabilidades de todos los eventos en este espacio es 1 (b).
7	a y b	Hay 3 números pares y 3 impares en un dado de 6 caras, por lo que sus probabilidades son iguales (3/6).
8	a	Para eventos independientes $P(A \cap B) = P(A) \times P(B)$. Por lo tanto, $P(A \cap B) = 0.4 \times 0.6 = 0.24$
9	b	La probabilidad de que un evento no ocurra es el complemento de la probabilidad de que ocurra. En este caso, $P(no\ maduras) = 1 - 0.7 = 0.3$
10	c	La probabilidad de que un evento no ocurra es el complemento de su probabilidad. Aquí, $P(Fracaso) = 1 - 0.85 = 0.15$

Ir a la autoevaluación

Autoevaluación 5

Pregunta	Respuesta	Retroalimentación
1	b	La distribución binomial modela el número de éxitos en un número fijo de ensayos independientes, donde la probabilidad de éxito es constante.
2	a	La distribución de Poisson modela la cantidad de eventos que ocurren en un intervalo de tiempo o espacio.
3	b	La distribución normal es una distribución continua con forma de campana, simétrica alrededor de la media.
4	d	La distribución exponencial mide el tiempo entre eventos en un proceso continuo que ocurre de manera constante e independiente.
5	a	Este problema se ajusta a una distribución binomial porque modela un número fijo de intentos (n = 10) con una probabilidad constante de éxito (p = 0.8).
6	a y b	La distribución Poisson es adecuada para modelar conteos de eventos como las plagas, con $\lambda = 5$ como tasa promedio.
7	a y b	Este es un problema típico de distribución normal, donde la mayoría de los datos están dentro de $\mu \pm \sigma$
8	a y c	La Poisson y la exponencial están relacionadas: una mide conteos (Poisson) y la otra mide tiempos entre eventos (exponencial), ambas con el mismo λ.
9	b	La distribución de Poisson modela el número de eventos (plagas) que ocurren en un intervalo (hectárea) con una tasa promedio constante $(\lambda = 4)$
10	a y c	La distribución de Poisson modela conteos de eventos en un intervalo fijo, con una tasa constante (λ), y los eventos deben ser independientes.

Ir a la autoevaluación

Autoevaluación 6

Pregunta	Respuesta	Retroalimentación
1	b	El muestreo aleatorio simple garantiza que todos los elementos de la población tengan igual probabilidad de ser seleccionados, sin criterios adicionales.
2	d	En el muestreo estratificado, los subgrupos son homogéneos y se toma una muestra de cada uno, mientras que, en el conglomerado, los grupos son heterogéneos y se seleccionan completos.
3	b	Una distribución muestral describe cómo se distribuyen los valores de un estadístico, como la media o la proporción, calculados a partir de muestras repetidas.
4	a	El Teorema del Límite Central establece que la distribución muestral de la media se aproxima a una normal, independientemente de la distribución de la población, cuando *n* es grande.
5	a y b	El Teorema del Límite Central permite realizar análisis estadísticos confiables, incluso cuando la población no es normal.
6	a y c	La media muestral tiene una media igual a la poblacional, y su error estándar depende del tamaño de la muestra, disminuyendo conforme *n* aumenta.
7	a	La distribución muestral de la media tiene una media igual a la media poblacional, pero su desviación estándar es menor, dependiendo del tamaño de la muestra.
8	b	El error estándar de la media se calcula como $\frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{36}} = 2.5$
9	b y d	Un proceso bajo control estadístico no presenta variaciones atribuibles a causas especiales, solo a causas comunes o aleatorias.
10	a y d	Las gráficas de control y los diagramas de Pareto son herramientas clave para monitorear procesos y priorizar problemas en el control estadístico de procesos.

Ir a la autoevaluación

Autoevaluación 7

Pregunta	Respuesta	Retroalimentación
1	b	Una estimación puntual es un único valor que se utiliza para aproximar un parámetro poblacional, como la media muestral para estimar la media poblacional.
2	a	Para construir un intervalo de confianza, se necesita la media muestral, el tamaño de la muestra, el nivel de confianza y la desviación estándar.
3	c	Un intervalo de confianza del 95% indica que, en el largo plazo, el 95% de los intervalos construidos incluirán el valor verdadero del parámetro poblacional.
4	b	Según el Teorema del Límite Central, la distribución de la media muestral se aproxima a una normal cuando el tamaño de la muestra es suficientemente grande, independientemente de la distribución de la población
5	a y b	Se calcula el margen de error multiplicando *z* por el error estándar y luego se suma y resta este valor de la media muestral para obtener el intervalo.
6	a y c	El intervalo de confianza para la diferencia entre medias utiliza el error estándar combinado y estima el rango probable de la diferencia entre las medias poblacionales.
7	c y d	Un nivel de confianza mayor asegura más certeza de incluir el parámetro poblacional, pero hace que el intervalo sea más amplio para reflejar mayor incertidumbre.
8	a y b	El intervalo sugiere que la diferencia promedio está entre 1.5 y 4.8, y como el intervalo no incluye 0, hay evidencia de una diferencia significativa entre las poblaciones.
9	b	Este intervalo estima el rango donde probablemente se encuentra la diferencia entre las medias de las poblaciones de las cuales provienen las muestras.
10	a	Un intervalo de confianza del 95% indica que, en el largo plazo, el 95% de los intervalos construidos incluirán el valor verdadero del parámetro poblacional.

Ir a la autoevaluación

Autoevaluación 8

Pregunta	Respuesta	Retroalimentación
1	c	Una prueba de hipótesis es un método estadístico para evaluar afirmaciones sobre un parámetro poblacional utilizando datos muestrales.
2	b	La hipótesis nula (H_0) es la afirmación inicial que se considera verdadera a menos que haya suficiente evidencia para rechazarla.
3	a	El nivel de significancia (α) es la probabilidad de rechazar H_0 cuando es verdadera (error tipo I).
4	b	La región crítica es el rango de valores del estadístico de prueba donde se rechaza la hipótesis nula.
5	d	El valor *p* mide la probabilidad de obtener un estadístico de prueba igual o más extremo que el observado, asumiendo que H_0 es verdadera.
6	a	Se rechaza H_0 si el estadístico de prueba está en la región crítica o si el valor $p \leq \alpha$.
7	b	Dado que el valor *p* (0.03) es menor que α (0.05), se rechaza H_0
8	c	La prueba de muestra grande requiere que la desviación estándar poblacional (σ) sea conocida.
9	c	El estadístico *z* se utiliza en pruebas de hipótesis cuando σ es conocida y el tamaño de la muestra es grande.
10	a	Dado que el valor *p* (0.01) es menor que α (0.05), se rechaza H_0.

Ir a la autoevaluación

5. Referencias bibliográficas

Agrawal, A., Gopal, K. (2013). *Principles of Statistics and Reporting of Data. In: Biomonitoring of Water and Waste Water*. Springer. https://doi.org/10.1007/978-81-322-0864-8_9

De la Puente Vidma, C. (2018). *Estadística descriptiva e inferencial*. Ediciones IDT CB.

Itza Ortiz, M. F., Escárcega Ávila, A. M., Carrera Chávez, J. M., Orozco Lucero, E. y Beristain Ruiz, D. M. (2024). *Estadística descriptiva: principios básicos de estudio.* (Primera edición). Universidad Autónoma de Ciudad Juárez.

Martínez, E. (2020). *Estadística*. (Primera edición). Ediciones UAPA.

Mendenhall, W., Beaver, R. y Beaver, B. (2015). *Introducción a la probabilidad y estadística* (14ª. edición). CENGAGE Learning.

Monroy Saldívar, S. (2008). *Estadística Descriptiva*. (Primera edición). Instituto Politécnico Nacional.

Llinás Solano, H. (2017). *Estadística Inferencial*. (Novena reimpresión). Universidad del Norte Editorial.

Nores, M.L. (2021). *Un recorrido por las distribuciones de probabilidad*. Revista de Educación Matemática. 36, 2 (Jul. 2021), 7–43. DOI: https://doi.org/10.33044/revem.30460.

Posada Hernández, G. J. (2016). *Elementos básicos de estadística descriptiva para el análisis de datos.* Editorial Fundación Universitaria Luis Amigó. https://www.funlam.edu.co/uploads/fondoeditorial/120_Ebook-elementos_basicos.pdf

Robayo-Botiva, D. M. (2020). *Medidas de posición en variable continua y discreta – datos desagrupados con Microsoft Excel* (Generación de contenidos impresos N.° 19). Bogotá: Ediciones Universidad Cooperativa de Colombia. doi: https://doi.org/10.16925/gcgp.26

6. Anexos