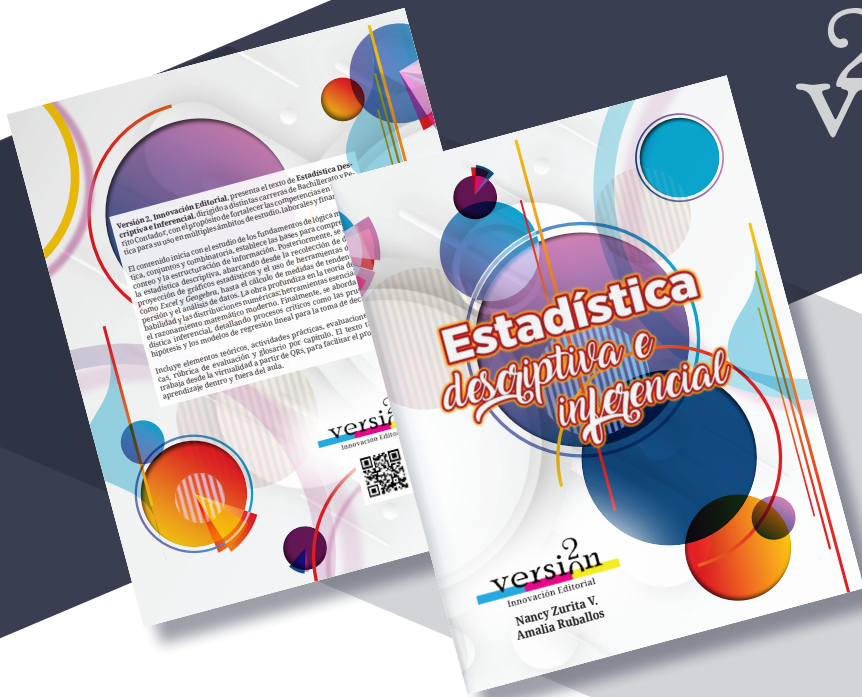


v2



Estadística Descriptiva e Inferencial
Nancy Zurita V. ~ Amalia Ruballos

Nancy Zurita Villagrán, docente especialista en Matemática, Estadística y Educación. Experiencia en análisis de datos, elaboración y validación de cuestionarios, dirección de proyectos de levantamiento de encuestas a nivel nacional, experta en la enseñanza de la matemática y en elaboración de textos educativos. Ha sido consultora en diferentes instituciones públicas y privadas, como INE, UVG, Diálogos, MIJ, Editorial Norma, Editorial Santillana, Ministerio de Educación (DIGEDUCA) entre otros. Desde el año 2011 está a cargo del Departamento de Matemática de la Universidad del Valle de Guatemala.

Amalia Ruballos

J. de Mancio, profesional en

Matemática y Estadística con sólida formación en Educación e Ingeniería con posgrados en Evaluación Educativa. Experiencia en docencia universitaria, coordinación académica y formación docente a nivel centroamericano. Ha liderado proyectos curriculares, procesos de evaluación del aprendizaje y formación en competencias digitales para la educación. Cuenta con experiencia en investigación educativa aplicada, especialmente en evaluación de programas, innovación docente y análisis cuantitativo y cualitativo de datos.



versión 2

Innovación Editorial

Versión 2. Innovación Editorial

28 calle 12-22 zona 11, Granai 2

gestion@version2.com.gt

502 2211 6808

Versión 2, Innovación Editorial, presenta el texto de **Estadística Descriptiva e Inferencial**, dirigido a las distintas carreras de Bachillerato y Perito Contador. El contenido inicia con el estudio de los fundamentos de lógica matemática, conjuntos y combinatoria, estableciendo las bases para comprender el conteo y la estructuración de información.

Posteriormente, se explora la estadística descriptiva, abarcando desde la recolección de datos, la proyección de gráficos estadísticos y el uso de herramientas digitales como Excel, hasta el cálculo de medidas de tendencia, dispersión y el análisis de datos. La obra también profundiza en la teoría de la probabilidad y las distribuciones numéricas, herramientas esenciales para el razonamiento matemático moderno. Finalmente, el texto aborda la estadística inferencial, detallando procesos críticos como las pruebas de hipótesis y los modelos de regresión lineal para la toma de decisiones.

CONJUNTOS Y COMBINATORIA

Conjuntos y sus operaciones. Operaciones entre conjuntos. Cardinalidad de conjuntos. Cardinalidad. **Microproyecto grupal:** Festival escolar. Técnicas de conteo. Principio de multiplicación. Diagramas de árbol. Principio de la suma. Factorial. Permutaciones y combinaciones. Permutación. **Microproyecto grupal:** Descubriendo el triángulo de Pascal y las combinaciones.

OPERACIONES ENTRE CONJUNTOS

Al igual que sumas o restas números, puedes hacer operaciones con conjuntos. Las operaciones básicas entre conjuntos son:

Operación	Definición	Representación en diagramas de Venn
Unión	A unión B es el conjunto que tiene los elementos de A o B. Se denota $A \cup B$.	
Intersección	A intersección B es el conjunto que tiene los elementos de A y B. Se denota $A \cap B$.	
Diferencia	A diferencia B es el conjunto que tiene los elementos de A y que no están en B. Se denota $A - B$.	
Complemento	El complemento de A son los elementos que están en el conjunto universo (U) y no están en A. Se denota A^c .	

Ejemplo 4

Si $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$ y U es el conjunto de los dígitos, entonces:

$$\begin{aligned} A \cup B &= \{1, 2, 3, 4, 5\} \\ A \cap B &= \{3\} \\ A^c &= \{0, 4, 5, 6, 7, 8, 9\} \\ B^c &= \{0, 1, 2, 6, 7, 8, 9\} \end{aligned}$$

Conjunto potencia

Cuando tienes un conjunto, no solo puedes trabajar con sus elementos uno por uno o en grupos. El **conjunto potencia** es el conjunto formado por todos los subconjuntos posibles que puedes construir a partir de ese conjunto, incluyendo el conjunto vacío y el mismo conjunto A.

Se denota como $P(A)$ o $\mathcal{P}(A)$.

Ejemplo 5

Supongamos que tienes el conjunto $A = \{1, 2\}$.

Los subconjuntos posibles son:

- $\emptyset$ (el conjunto vacío)
- $\{1\}$
- $\{2\}$
- $\{1, 2\}$ (el conjunto A)

Entonces:

$$P(A) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$$



REFLEXIONA

DATO CURIOSO:

El conjunto potencia es muy útil en programación, combinatoria y lógica, porque nos ayuda a analizar todas las posibilidades de un conjunto dado.



SECCIÓN 3

TÉCNICAS DE CONTEO

Cuando nos enfrentamos a situaciones en las que hay que contar posibilidades, es muy fácil olvidar opciones o confundir combinaciones. Por eso, las técnicas de conteo son herramientas poderosas: nos ayudan a organizar y calcular de manera ordenada todas las formas en que algo puede ocurrir. Estas técnicas son muy útiles en la vida cotidiana y en muchos áreas: por ejemplo, para analizar estrategias en juegos de mesa, diseñar programas de computadora que consideren distintas rutas o elecciones, planificar actividades de forma eficiente o simplemente asegurarnos de no olvidar ninguna posibilidad al resolver un problema.

PRINCIPIO DE MULTIPLICACIÓN

Si una tarea se puede hacer de m maneras y otra se puede hacer de n maneras, entonces ambas tareas juntas se pueden hacer de $m \cdot n$ maneras.

Ejemplo 1

Supón que tienes 3 camisetas (roja, azul, verde) y 2 pantalones (negro y gris).



¿Cuántos conjuntos diferentes puedes formar?

Este proceso se puede realizar en dos tareas: (1) seleccionar la camiseta y (2) seleccionar el pantalón. La tarea 1 se puede realizar de 3 maneras y la segunda de 2. Entonces por el principio de multiplicación, se tiene que hay $2 \times 3 = 6$ combinaciones posibles.

DIAGRAMAS DE ÁRBOL

Un diagrama de árbol es un dibujo que usamos para representar visualmente todas las opciones que podemos tomar paso a paso. Cada "rama" del árbol muestra una decisión o posibilidad, y al final de cada camino se ve un resultado completo.

Ejemplo 2

Para el ejemplo anterior podríamos construir el diagrama de árbol.

Primer:

Hay que dividir el proceso en tantas etapas como sea posible. En nuestro caso solamente hay dos etapas y pantalón. Entonces quién será la primera etapa? nos sugiere utilizar la que tiene más opciones como la primera etapa.

Segundo:

Dibuja un nodo y las posibilidades para la primera etapa.

Tercero:

Dibuja un nodo y las posibilidades para la segunda etapa.



ORGANIZACIÓN DE LA INFORMACIÓN Y GRÁFICAS

2 CAPÍTULO

Estadística. Variables. Escalas de medición. Recolección y organización de la información. Variables cualitativas: tablas de frecuencias y métodos gráficos: barras, sectores (gráfico de pastel), pictograma, barras apiladas, pirámides. Excel: tablas, gráficos y tabla dinámica. Variables cuantitativas: distribuciones de frecuencias, tabla de frecuencias y métodos gráficos: Histograma, polígono de frecuencias. **Microproyecto grupal:** Explorando los hábitos estudiantiles.



Sección 1

ESTADÍSTICA

Es una disciplina cuyo término es acuñado al economista **Godtfred Achenwall** en 1760, la palabra proviene de *statum*, que significa "Estado". Además, fue el primero en mostrar tablas y gráficas que la información que presentaba. La estadística es la ciencia que se encarga de la recolección, organización, presentación, análisis e interpretación de la información con el fin de explicar fenómenos, hechos y circunstancias.



ACTIVIDAD Discute con tus compañeros, cuál de los pasos de la estadística consideran que es el más importante en el proceso?

La Estadística suele dividirse en: **Estadística descriptiva** y **Estadística inferencial**.

Descriptiva	Inferencial
Es la encargada de recolectar, procesar y presentar la información obtenida en tablas y gráficas que permitan resumir y explicar fenómenos y el comportamiento de los datos.	Mediante procesos que permiten procesar la información de una muestra para hacer predicciones generalizables para toda una población.

La estadística consiste en recolectar información para posteriormente analizarla con el afán de llegar en la toma de decisiones basadas en evidencias.

Los sujetos de estudio pueden ser individuos, familias, países, ciudades, empresas, instituciones, entre otros.

Población

Es el conjunto total o universo de sujetos u objetos de los cuales el estudio quiere saber algo.

Muestra

Es una porción o subconjunto de la población, la cual debe cumplir con las características de la población para que, al momento de realizar la inferencia, los resultados muestrales puedan ser generalizables.

Cuando se habla de población y muestra es de suma importancia definir apropiadamente los resultados que de cada una de ellas se

MÉTODOS GRÁFICOS PARA VARIABLES CUALITATIVAS

Los métodos gráficos son herramientas fundamentales para visualizar datos cualitativos, facilitando su interpretación y análisis. Los gráficos ayudan a comunicar de manera clara los resultados de una investigación y permiten identificar patrones o tendencias en los datos. A continuación, se presentan los principales métodos gráficos utilizados para representar variables cualitativas:

1. Gráfico de barras

El gráfico de barras es uno de los métodos más utilizados para representar datos cualitativos. En este gráfico, cada categoría se representa mediante una barra cuya altura es proporcional a la frecuencia de la categoría.

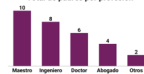
- Características:
 - Las barras pueden ser horizontales o verticales.
 - Las barras no están unidas entre sí.
 - Se puede utilizar tanto la frecuencia absoluta como la relativa o porcentual.

Ejemplo 2:

Se desea representar los datos sobre las profesiones de los padres de 25 estudiantes:

Profesión	Total de padres
Maestro	10
Ingeniero	8
Doctor	6
Abogado	4
Otros	2

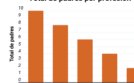
Total de padres por profesión



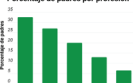
- Cada barra representa cada una de las categorías (se colocan en el eje x).
- La altura de cada barra que se coloca en el eje y, será la frecuencia con la que dicha categoría está presente.
- Todas las barras deben ser del mismo ancho.

Nota: es recomendable agregar el nombre a los ejes, así como de guardar una proporción del 75% del eje x en el eje y. No olvides que puedes cambiar las frecuencias absolutas por las frecuencias relativas o porcentuales, según la información que quieras dar.

Total de padres por profesión

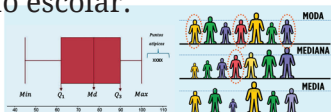


Porcentaje de padres por profesión



INTERPRETACIÓN DE LA INFORMACIÓN

Medidas de tendencia central. Datos sin agrupar: media aritmética, geométrica y armónica, mediana, moda. Datos agrupados: media aritmética, mediana, moda. Medidas de posición: datos sin agrupar, datos agrupados en tabla de frecuencias simples y en intervalos. Medidas de dispersión: rango, intercuartílico, varianza y desviación estándar. Medidas de forma: coeficientes de asimetría y curtosis, diagrama de caja y bigote. **Proyecto estadístico:** Mi entorno escolar.



Sección 1

Medidas de tendencia central, medidas de posición y medidas de variabilidad

El análisis de datos a través de las medidas de tendencia central, posición, variación y forma permite obtener información relevante sobre las características de los datos y facilita la toma de decisiones informadas. Estas medidas ayudan a describir y resumir los datos, identificar patrones y tendencias, y comprender la distribución de los valores en un conjunto de datos.

MEDIDAS DE TENDENCIA CENTRAL

Las medidas de tendencia central permiten identificar un valor representativo o promedio dentro de un conjunto de datos. Entre las más comunes se encuentran la media, la mediana, la moda, así como otras medidas específicas como la media geométrica y la media armónica.

La media aritmética y la mediana son especialmente útiles para resumir la información de forma eficaz. La media aritmética mide adecuadamente datos simétricos y sin valores extremos, mientras que la mediana es preferible cuando existen valores atípicos o distribuciones sesgadas. Por su parte, la media geométrica es útil para calcular promedios de tasas de crecimiento o proporciones, mientras que la media armónica se emplea en el análisis de velocidades, tasas o razones. Todas las medidas serán calculadas para datos simples y agrupados en intervalos.

A. MEDIDAS DE TENDENCIA CENTRAL PARA DATOS SIN AGRUPAR

MEDIA ARITMÉTICA O PROMEDIO ($\bar{x}$)

La media es el valor promedio ($\bar{x}$) de un conjunto de datos y se obtiene sumando todos los valores y dividiendo el resultado entre la cantidad total de datos.

- Pasos para calcular la media para datos sin agrupar:
- Sumar todos los valores ($\sum x_i$).
 - Dividir la suma entre la cantidad total de datos (n).

La fórmula que puede utilizarse es:

$$\bar{x} = \left(\sum x_i \right) / n$$

Ejemplo 1

Un grupo de 10 estudiantes obtuvo las siguientes calificaciones en un examen: 85, 90, 78, 92, 88, 80, 88, 91, 87, 89.

Cálculo:

$$\bar{x} = \frac{785 + 90 + 78 + 92 + 88 + 80 + 88 + 91 + 87 + 89}{10}$$

$$= 866/10 = 86.6$$

$$\bar{x} = 86.6$$

Interpretación: la calificación promedio de los estudiantes es 86.6.

Calcular la media aritmética del número diario de celular e internet.

Ejercicio 4: Cantidad de mensajes enviados por día. Se registró cuántos mensajes envían los adolescentes por día en redes sociales:

Intervalo (mensajes)	Frecuencia (f)
0-49	4
50-99	6
100-149	8
150-199	5
200-249	2

Intervalo (minutos)	Frecuencia (f)	x_i	$f \cdot x_i$
0-49	4		
50-99	6		
100-149	8		
150-199	5		
200-249	2		
Total	25		

Calcular la media aritmética del número de mensajes enviados.

MEDIANA (Md o Me)

La mediana es una medida de tendencia central que representa el valor central de un conjunto de datos ordenados. Es especialmente útil cuando los datos contienen valores extremos o atípicos, ya que no se ve afectada por ellos como ocurre con la media. La media

na divide al conjunto de datos en dos partes iguales:

- El 50% de los datos son menores o iguales a la mediana.
- El otro 50% son mayores o iguales a ella.

Para datos en tabla de frecuencias simples:

1. Calcular el número total de datos N sumando todas las frecuencias.
2. Identificar $N/2$ si es par o impar.
3. Determinar la posición de la mediana como en datos simples.
4. Acumular las frecuencias hasta encontrar en qué valor se la posición de la mediana.

Ejemplo 2

Una profesora registró la cantidad de horas semanales dedicadas al deporte por sus 20 estudiantes:

Horas (x_i)	Frecuencia (f)
1	3
2	5
3	4
4	6
5	2
Total	20

Paso 1. Para la calcular la mediana primero debemos calcular las frecuencias acumuladas.

Horas (x_i)	Frecuencia (f)	F_{ac}
1	3	3
2	5	8
3	4	12
4	6	18
5	2	20
Total	20	

Paso 2. Ahora se debe ubicar la posición de la mediana mediante

$$N/2 = 20/2 = 10$$

PROBABILIDAD

Espacios muestrales y probabilidad. Espacios muestrales. Conceptos básicos de la probabilidad. **Microproyecto grupal:** ¿Qué tan cerca está la práctica de la teoría? Reglas de probabilidad. Regla de la adición. Regla del producto. Distribución de probabilidad. Distribuciones discretas. Distribución binomial. Distribución normal. **Proyecto estadístico:** Uso de *GeoGebra Classic* en línea para distribuciones normales.

Ejemplo 3

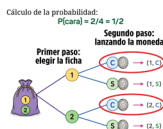
Para el espacio muestral del Ejemplo 1, ¿cuál es la probabilidad de obtener color rojo e la ruleta. Podemos ver en la tabla que hay 2 casos posibles del total de 4: $\{C, R\}$ de $\{C, R, B, G\}$.

	R (Rojo)	A (Azul)
C (Cara)	(C,R)	(C,A)
S (Sello)	(S,R)	(S,A)

Cálculo de la probabilidad:
 $P(\text{rojo}) = 2/4 = 1/2$

Ejemplo 4

Para el espacio muestral del Ejemplo 2, ¿cuál es la probabilidad de obtener cara al lanzar la moneda. Podemos ver en el árbol que hay 2 casos posibles del total de 4: $\{C, C\}$ de $\{C, C, S, S\}$.

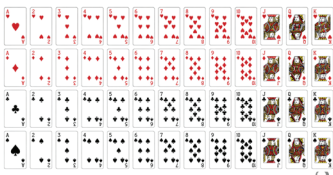


Ejemplo 5

Se extrae una carta al azar de una baraja estándar de 52 cartas. Se desea calcular la probabilidad de obtener un 6 o una carta de corazones.

Paso 1: Calcular el espacio muestral

En una baraja estándar hay un total de 52 cartas, organizadas en 4 palos: corazones, diamantes, tréboles y picas. Cada palo contiene 13 cartas, que se dividen en cartas numeradas del 1 al 10, más las cartas especiales: As (A), Jota (J), Reina (Q) y Rey (K). Cada carta es única porque combina un valor (por ejemplo, 6, Q o A) y un palo (por ejemplo, corazones o picas). Por ejemplo, el "6 de corazones" es diferente del "6 de trébol". Esto permite formar el espacio muestral completo como el conjunto de las 52 combinaciones posibles.



DISTRIBUCIONES DISCRETAS

Una distribución de probabilidad discreta describe cómo se asignan probabilidades a los distintos valores posibles de una variable aleatoria que solo puede tomar valores específicos y contables. Estos valores suelen ser números enteros que miden de contar ocurrencias, como el número de éxitos en varios intentos o la cantidad de veces que ocurre un evento. En este tipo de distribución, cada valor posible tiene una probabilidad asociada, y la suma de todas estas probabilidades es igual a 1. Esto permite analizar y predecir la probabilidad de cada resultado en situaciones donde los resultados no son continuos, sino separados y bien definidos.

Ejemplo 3

Para construir la distribución de probabilidad para el lanzamiento de un dado donde la variable aleatoria X es tal que:

$$X \in \{1, 2, 3, 4, 5, 6\}$$

y como todos los resultados son igualmente probables,

$P(X = x) = 1/6$ para $x = 1, 2, 3, 4, 5, 6$

entonces la tabla de distribución de probabilidad es así:

X	1	2	3	4	5	6
P(X)	1/6	1/6	1/6	1/6	1/6	1/6

Esto significa que cada número tiene la misma probabilidad de ocurrir, y la suma de todas las probabilidades es:

$$1/6 + 1/6 + 1/6 + 1/6 + 1/6 + 1/6 = 1$$



Ejemplo 4

Se lanzan dos dados y se define la variable aleatoria X como la suma de los puntos obtenidos.

Los valores posibles de X son:

$$2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$$

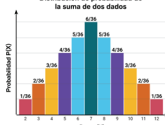
Sin embargo, no todos tienen la misma probabilidad. Por ejemplo:

- La suma 7 se puede obtener de muchas formas: $\{1,6\}$, $\{2,5\}$, $\{3,4\}$, $\{4,3\}$, $\{5,2\}$, $\{6,1\}$
- La suma 2 solo se obtiene con $\{1,1\}$
- La suma 12 solo se obtiene con $\{6,6\}$

Entonces la tabla de distribución de probabilidad es así:

x	P(X)
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Distribución de probabilidad de la suma de dos dados



ESTADÍSTICA INFERENCIAL

5 CAPÍTULO

Estadística inferencial. Distribución muestral. Error estándar. Estimación de parámetros. Intervalos de confianza: estimación puntual, para la media y para la proporción. Pruebas de hipótesis. Errores en la toma de decisión: tipo I y tipo II. Pasos para prueba de hipótesis. Prueba de hipótesis para la media y para la proporción. **Proyecto estadístico:** Satisfacción y rendimiento en el aprendizaje virtual. Correlación y regresión lineal simple. Correlación lineal. Coeficiente de correlación de Pearson. Regresión lineal simple. **Proyecto estadístico:** Evaluación del efecto de la medicación sobre el apetito.

Si se toma una muestra de $n = 64$ baterías, ¿cuál es el error estándar de la media muestral?

$$EE_{\bar{x}} = \frac{4}{\sqrt{64}} = \frac{4}{8} = 0.5$$

Interpretación:

El promedio de vida útil obtenido a partir de muestras de 64 baterías fluctuará, en promedio, 0.5 meses respecto al valor poblacional.

Por lo tanto, la muestra tiene una buena precisión para estimar la media real.

ACTIVIDAD

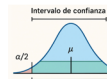


Tu y otro compañero respondan las siguientes preguntas considerando el ejemplo anterior:

1. ¿Qué pasaría con el error estándar si en lugar de 64 baterías se tomaran solo 16?

2. ¿Por qué es importante el tamaño de muestra en la precisión de una estimación?

3. ¿Crees que el error estándar mide "errores" o el sentido de equivocaciones humanas? Explica.



En muchas situaciones reales no es posible estudiar a toda la población, por lo que necesitamos inferir sus características a partir de una muestra.

La estimación de parámetros permite precisamente eso: utilizar los datos obtenidos para aproximar valores desconocidos, como la media o la proporción poblacional, y expresar el grado de confianza en esas estimaciones.

Sección 2

ESTIMACIÓN DE PARÁMETROS INTERVALOS DE CONFIANZA

Esta sección te enseñará a construir e interpretar intervalos de confianza, comprendiendo que toda estimación implica cierta incertidumbre. Aquí aprenderás a convertir los resultados muestrales en conclusiones confiables sobre la población.

ESTIMACIÓN PUNTUAL

Como mencionamos antes, la estimación puntual asigna un único valor numérico (obtenido de la muestra) para aproximar un parámetro poblacional desconocido.



