

# Estadística descriptiva. Ciencia de Datos. Probabilidad

*Ciencias Exactas y Naturales | Estadística*

## Descripción del Curso

Esta unidad se centra en el uso de herramientas tecnológicas (hojas de cálculo o software de programación) para realizar cálculos estadísticos y generar visualizaciones, además de comunicar resultados de forma clara para audiencias no técnicas, destacando limitaciones y recomendaciones.

En el marco de la asignatura Estadística, la Unidad 7, Herramientas Tecnológicas y Comunicación de Resultados, se integra para fortalecer la capacidad de convertir datos en evidencia comprensible y accionable. El curso combina fundamentos estadísticos con prácticas orientadas a la manipulación de datos y la presentación de hallazgos a públicos diversos. A través de actividades con hojas de cálculo y/o lenguajes de programación, los estudiantes adquieren habilidades para calcular medidas descriptivas, construir visualizaciones adecuadas y seleccionar el tipo de gráfico más eficaz según el tipo de dato y el objetivo de comunicación.

El objetivo general de la unidad es: Utilizar herramientas tecnológicas para realizar cálculos estadísticos descriptivos y generar visualizaciones, comunicando resultados de manera clara y responsable a audiencias no técnicas.

Específicos:

- Operar con hojas de cálculo o software de programación para calcular medidas descriptivas y crear visualizaciones adecuadas.
- Seleccionar el tipo de gráfico adecuado para cada tipo de dato y objetivo de comunicación.
- Redactar interpretaciones de resultados enfatizando supuestos, limitaciones y recomendaciones para públicos no técnicos.

La unidad también fomenta el desarrollo de una comunicación responsable y ética: presentar resultados con claridad, reconociendo incertidumbres y limitaciones, y sostener las recomendaciones en evidencia numérica, de modo que los estudiantes sean capaces de justificar decisiones ante audiencias diversas y con distintos niveles de alfabetización estadística.

## Competencias

- Analizar conjuntos de datos y realizar cálculos descriptivos utilizando herramientas tecnológicas.
- Crear visualizaciones efectivas y apropiadas para comunicar hallazgos.
- Interpretar resultados considerando supuestos, limitaciones y incertidumbres.
- Comunicar de forma clara y responsable a audiencias no técnicas, adaptando el mensaje al contexto y al nivel de alfabetización estadística.
- Aplicar principios de ética y responsabilidad en la presentación y uso de los datos.
- Integrar conocimientos de estadística con habilidades digitales para tomar decisiones basadas en evidencia.

## Requerimientos

- Conocimientos básicos de estadística descriptiva y gráficos fundamentales.
- Acceso a herramientas tecnológicas: hojas de cálculo (por ejemplo, Excel o Google Sheets) o software de programación (R, Python) y, opcionalmente, entornos de visualización (Tableau, Power BI).
- Computadora con conexión a internet y capacidad para instalar o ejecutar software necesario.
- Capacidad para interpretar gráficos y comunicar hallazgos de manera clara a audiencias no técnicas.
- Participación en actividades prácticas y entregas dentro de las fechas establecidas y compromiso con la calidad de las visualizaciones y la interpretación.

## Unidades del Curso

### Unidad 1: Unidad 1: Fundamentos de Estadística Descriptiva, Probabilidad y Ciencia de Datos

#### Objetivos de Aprendizaje

- Identificar tipos de datos y variables (cuantitativas y cualitativas) y sus escalas de medición.
- Explicar y aplicar conceptos básicos de probabilidad (evento, probabilidad, complementos) y las reglas de suma y producto en contextos simples.
- Relacionar la estadística descriptiva con las etapas iniciales de un proyecto de ciencia de datos y su utilidad para la toma de decisiones.

#### Contenidos Temáticos

1. **Tipo de datos y escalas de medición** – Descripción corta: qué son variables cualitativas vs. cuantitativas y sus escalas (nominal, ordinal, intervalo/razón).
2. **Introducción a la estadística descriptiva** – Descripción corta: conceptos de tendencia central y dispersión como fundamentos para resumir datos.
3. **Probabilidad básica y reglas clave** – Descripción corta: evento, probabilidad, complementos y reglas de suma y producto aplicadas a escenarios simples.

#### Actividades

- **Actividad 1: Exploración de tipos de datos** — Identificar y clasificar variables en un conjunto de datos real disponible (p. ej., datos de satisfacción o rendimiento). Breve explicación de qué mide cada variable y la escala correspondiente. Principales aprendizajes: clasificación de datos y elección de métodos adecuados.
- **Actividad 2: Cálculo de probabilidades básicas** — Resolver ejercicios simples que involucren eventos y reglas de suma y producto (dados o cartas). Puntos clave: aplicar correctamente las reglas y entender resultados en contextos simples.

- **Actividad 3: Conexión con ciencia de datos** — Discusión guiada sobre qué rol juegan las descripciones y probabilidades en un flujo de análisis de datos, con ejemplos de proyectos reales.

## Evaluación

Evaluación formativa mediante: (i) ejercicios cortos de clasificación de datos y cálculos de probabilidades; (ii) participación en debates sobre el papel de la estadística en ciencia de datos; (iii) un cuestionario corto que combine conceptos de tipos de datos, medidas básicas y reglas de probabilidad básica.

## Unidad 2: Medidas de Tendencia Central y Dispersión I

### Objetivos de Aprendizaje

- Calcular media y mediana de conjuntos de datos y comparar sus resultados en distintas distribuciones.
- Calcular rango, varianza y desviación típica y comprender su interpretación práctica.
- Relacionar las medidas con contextos reales para explicar cuándo una medida es más adecuada que otra.

### Contenidos Temáticos

1. **Media y mediana** – Descripción corta: cómo se calculan y en qué situaciones capturan la tendencia central de un conjunto de datos.
2. **Rango, varianza y desviación típica** – Descripción corta: conceptos de dispersión y su interpretación en datos reales.
3. **Interpretación contextual de medidas** – Descripción corta: elección de medidas según características de la distribución (simetría, sesgo, presencia de valores extremos).

### Actividades

- **Actividad 1: Cálculos prácticos de tendencia central** — Trabajar con un conjunto de datos pequeño para calcular media y mediana, comparar resultados ante distribuciones sesgadas vs. simétricas.
- **Actividad 2: Medidas de dispersión en contextos reales** — Calcular rango, varianza y desviación típica y discutir qué dicen sobre la variabilidad de los datos.
- **Actividad 3: Análisis interpretativo** — Presentar una interpretación de las medidas con un caso aplicado (p. ej., puntuaciones de un examen o tiempos de respuesta de usuarios).

## Evaluación

Evaluación formativa basada en ejercicios de cálculo y en una breve actividad de interpretación de un caso real, alineada con los objetivos 1 y 3 de la unidad.

## Unidad 3: Distribución de Datos y Gráficos Descriptivos

### Objetivos de Aprendizaje

- Construir tablas de frecuencia y frecuencias relativas para conjuntos de datos.
- Crear y leer histogramas, diagramas de caja (boxplot) y gráficos de barras para resumir la distribución.
- Interpretar la forma de la distribución, identificar sesgos y detectar outliers.

## Contenidos Temáticos

1. **Tablas de frecuencia y frecuencias relativas** – Descripción corta: cómo organizar datos en tablas y calcular porcentajes.
2. **Histogramas** – Descripción corta: interpretación de la densidad, la dispersión y la forma de la distribución a través de barras.
3. **Diagramas de caja (boxplot) y gráficos de barras** – Descripción corta: lectura de cuartiles, valores atípicos y comparación entre grupos.

## Actividades

- **Actividad 1: Construcción de tablas de frecuencia** — Tomar un conjunto de datos y construir tablas de frecuencia y frecuencias relativas, interpretando resultados.
- **Actividad 2: Gráficos descriptivos** — Crear histogramas, boxplots y gráficos de barras con software o Excel/Google Sheets y analizar la forma de la distribución y posibles sesgos.
- **Actividad 3: Análisis de outliers** — Identificar posibles valores atípicos y discutir su impacto en las medidas resumidas.

## Evaluación

Evaluación mediante ejercicios de construcción de tablas y gráficos y una breve interpretación de la distribución de un conjunto de datos, conectando con el objetivo 2.

## Unidad 4: Unidad 4: Probabilidad Básica y Aplicaciones en Ciencia de Datos

### Objetivos de Aprendizaje

- Calcular la probabilidad de un evento en un experimento sencillo.
- Aplicar las reglas de suma y producto para resolver problemas de probabilidad contextualizados.
- Interpretar resultados probabilísticos en términos de decisiones dentro de un contexto de ciencia de datos.

## Contenidos Temáticos

1. **Probabilidad de un evento** – Descripción corta: definición y cálculo básico de probabilidades simples.
2. **Reglas de suma y producto** – Descripción corta: uso de la regla de la suma para eventos mutuamente excluyentes y la regla del producto para eventos independientes.

3. **Aplicaciones prácticas** – Descripción corta: resolución de escenarios simples relevantes para ciencia de datos (p. ej., combinaciones de características, decisiones basadas en probabilidades).

## Actividades

- **Actividad 1: Cálculo de probabilidades simples** — Ejercicios con experiencias simples (flip de moneda, dados) para practicar  $P(A)$ ,  $P(B)$ ,  $P(A \cap B)$  y  $P(A \cup B)$ .
- **Actividad 2: Problemas contextualizados** — Resolver problemas donde se use la probabilidad para apoyar decisiones de negocio o de ciencia de datos (p. ej., probabilidad de ocurrencia de un evento en un conjunto de registros).
- **Actividad 3: Discusión de límites** — Analizar supuestos de probabilidad en el contexto de datos reales y su impacto en conclusiones.

## Evaluación

Evaluación a través de ejercicios de probabilidad básica y un mini-proyecto de aplicación en un escenario contextualizado para demostrar comprensión de conceptos y su uso práctico.

## Unidad 5: Unidad 5: Probabilidad Condicional e Independencia

### Objetivos de Aprendizaje

- Definir y calcular  $P(A|B)$  y  $P(A \cap B)$  a partir de tablas de frecuencias o escenarios dados.
- Explicar qué significa independencia entre eventos y cómo detectarla en datos.
- Aplicar probabilidades condicionales para interpretar relaciones entre variables en contextos de ciencia de datos.

### Contenidos Temáticos

1. **Probabilidad condicional** – Descripción corta: definición, fórmula y ejemplos prácticos.
2. **Independencia de eventos** – Descripción corta: criterios, interpretación y diferencias con dependencia.
3. **Aplicación en ciencia de datos** – Descripción corta: uso de probabilidades condicionadas para entender relaciones y tomar decisiones en datos reales.

## Actividades

- **Actividad 1: Cálculo de probabilidades condicionadas** — Utilizar tablas de contingencia para calcular  $P(A|B)$  y  $P(A \cap B)$  y discutir interpretación.
- **Actividad 2: Identificación de independencia** — Análisis de escenarios para decidir si dos eventos son independientes y qué implica para modelos simples.
- **Actividad 3: Problemas de ciencia de datos** — Aplicar conceptos a ejemplos de datos (p. ej., correlaciones simples entre características) para comprender relaciones condicionales.

## Evaluación

Evaluación mediante ejercicios prácticos de probabilidades condicionales e independencia y una breve actividad de interpretación en un caso de ciencia de datos.

## **Unidad 6: Unidad 6: Flujo Básico de Análisis de Datos y Reproducibilidad**

### **Objetivos de Aprendizaje**

- Diseñar un flujo de trabajo de análisis de datos desde la obtención hasta la comunicación de resultados.
- Aplicar prácticas de limpieza y preprocesamiento para preparar datos para el análisis descriptivo.
- Crear productos reproducibles (documentación, notebooks, código) que permitan replicar el análisis.

### **Contenidos Temáticos**

1. **Recolección y calidad de datos** – Descripción corta: fuentes de datos, muestreo, integridad y ética de datos.
2. **Limpieza y preparación de datos** – Descripción corta: manejo de valores faltantes, normalización y estandarización.
3. **Análisis descriptivo y comunicación** – Descripción corta: resumen de resultados y comunicación a audiencias no técnicas, con énfasis en reproducibilidad.

### **Actividades**

- **Actividad 1: Diseño de flujo de datos** — Crear un diagrama simple de un flujo de análisis desde recolección hasta reporte, identificando entradas, procesos y salidas.
- **Actividad 2: Limpieza de datos** — Practicar técnicas de manejo de valores faltantes y estandarización en un dataset real o simulado.
- **Actividad 3: Producto reproducible** — Desarrollar un notebook o script con pasos documentados para reproducir el análisis y compartir la versión final.

### **Evaluación**

Evaluación basada en (i) la calidad del diagrama de flujo, (ii) un informe corto de limpieza y análisis descriptivo y (iii) un repositorio reproducible con notas de versión y documentación.

## **Unidad 7: Unidad 7: Herramientas Tecnológicas y Comunicación de Resultados**

### **Objetivos de Aprendizaje**

- Operar con hojas de cálculo o software de programación para calcular medidas descriptivas y crear visualizaciones adecuadas.
- Seleccionar el tipo de gráfico adecuado para cada tipo de dato y objetivo de comunicación.
- Redactar interpretaciones de resultados enfatizando supuestos, limitaciones y recomendaciones para públicos no técnicos.

## Contenidos Temáticos

1. **Herramientas y entornos** – Descripción corta: hojas de cálculo y/o lenguajes de programación para estadística básica y visualización.
2. **Visualización adecuada** – Descripción corta: criterios para elegir gráficos y buenas prácticas de visualización.
3. **Comunicación de resultados** – Descripción corta: redactar informes y presentaciones para audiencias no técnicas, con énfasis en limitaciones y recomendaciones.

## Actividades

- **Actividad 1: Cálculos y gráficos con herramientas** — Realizar cálculos descriptivos y crear visualizaciones (p. ej., media, mediana, desviación típica, histogramas) en Excel/Sheets o en un lenguaje de programación básico.
- **Actividad 2: Elección de gráfica adecuada** — Ejercicios para decidir qué gráfico usar según el tipo de dato y el objetivo comunicativo.
- **Actividad 3: Informe final para audiencia no técnica** — Preparar un breve informe que explique resultados, supuestos, limitaciones y recomendaciones en lenguaje claro.

## Evaluación

Evaluación a través de un proyecto corto donde se apliquen herramientas para cálculos y visualización y se elabore una presentación o informe para un público no técnico.