

SESIÓN MÉTODO CIENTÍFICO Y ESTADÍSTICA

CONTENIDOS:

- Filosofía y sociología de la ciencia.
- ¿Qué es el método científico?
 - Características principales (refutable, reproducible, objetivo, racional, sistemático, fáctico).
 - Pasos del método científico (planteamiento, hipótesis, experimentación, análisis, conclusión)
- Métodos de investigación: Cualitativos. Cuantitativos.
- Importancia del pensamiento estadístico en la ciencia.
- La estadística en el diseño de la investigación.
- Conceptos fundamentales de la estadística inferencial:
 - ¿Qué es la inferencia estadística?
 - Importancia.
 - Aplicaciones.
 - Ejemplos.
- El concepto de significancia.
- Diseño de experimentos.

FILOSOFÍA Y SOCIOLOGÍA DE LA CIENCIA

El método científico es el proceso sistemático y racional que permite obtener conocimiento válido y confiable sobre el mundo. En ciencia de datos, este método es esencial para diseñar experimentos, formular hipótesis y validar resultados. Además, la estadística juega un papel crucial en la interpretación de datos y la toma de decisiones basada en evidencia. Sin embargo, la ciencia no opera en aislamiento; su desarrollo está influenciado por factores filosóficos y sociales que determinan qué conocimientos se consideran válidos y cómo se difunden en la sociedad.



¿Qué es la Filosofía de la Ciencia?

La filosofía de la ciencia estudia los fundamentos, métodos e implicaciones del conocimiento científico. Algunas de sus preguntas clave incluyen:

- ¿Qué distingue el conocimiento científico de otras formas de conocimiento?
- ¿Cómo se valida una teoría?
- ¿Qué papel juega la falsabilidad en la construcción del conocimiento científico?

Diferentes corrientes filosóficas han influido en la evolución del pensamiento científico, desde el positivismo lógico, que enfatiza la observación y la verificación empírica, hasta el falsacionismo de Karl Popper, que sostiene que una teoría científica debe poder ser refutada para considerarse válida.

Sociología de la Ciencia

La sociología de la ciencia analiza cómo los factores sociales, económicos y políticos influyen en la producción y aceptación del conocimiento científico. Algunos aspectos clave incluyen:

- El papel de las comunidades científicas en la validación de teorías.
- Cómo el financiamiento y los intereses económicos pueden influir en la investigación.
- La construcción social del conocimiento y la resistencia al cambio paradigmático, como lo describió Thomas Kuhn en su teoría de las revoluciones científicas.

Por ejemplo, la aceptación del cambio climático ha estado influenciada no solo por la evidencia científica, sino también por debates políticos y económicos. Del mismo modo, en ciencia de datos, las decisiones sobre qué modelos utilizar o qué datos recolectar pueden estar condicionadas por factores sociales y éticos.

¿QUÉ ES EL MÉTODO CIENTÍFICO?

El método científico es un proceso sistemático utilizado para adquirir conocimiento mediante la observación, experimentación y análisis. Es la base de la investigación en diversas disciplinas, incluida la ciencia de datos, ya que permite obtener conclusiones fundamentadas a partir de datos empíricos.

Objetivo

El objetivo principal del método científico es validar o refutar hipótesis basadas en evidencia empírica. Al seguir este proceso, se minimizan los sesgos y se garantiza que los resultados sean verificables y confiables.

Características principales (refutable, reproducible, objetivo, racional, sistemático, fáctico).

El método científico tiene características clave que lo diferencian de otros enfoques para obtener conocimiento:

- Refutable: Las hipótesis deben poder ser probadas y, si es necesario, refutadas mediante la experimentación.
- Reproducible: Los experimentos deben poder repetirse bajo las mismas condiciones y obtener resultados similares.
- Objetivo: Se basa en hechos y datos, evitando interpretaciones subjetivas o influencias personales.
- Racional: Utiliza la lógica y el razonamiento para explicar fenómenos y llegar a conclusiones.
- Sistemático: Sigue un proceso ordenado y estructurado, lo que permite un análisis riguroso.
- Fáctico: Se basa en observaciones y datos reales obtenidos a partir de la experimentación.

Pasos del método científico (planteamiento, hipótesis, experimentación, análisis, conclusión)

El método científico sigue una serie de pasos que garantizan la validez y confiabilidad de los resultados:

1. Planteamiento del Problema: Identificar una pregunta o problema a investigar.

2. Formulación de Hipótesis: Proponer una posible explicación basada en conocimientos previos.
3. Experimentación: Diseñar y realizar pruebas para recopilar datos y evaluar la hipótesis.
4. Análisis de Datos: Interpretar los resultados mediante herramientas estadísticas o modelos científicos.
5. Conclusión: Determinar si la hipótesis es aceptada, rechazada o si necesita ser modificada.
 - Conclusión: Aceptar, rechazar o modificar la hipótesis.

Importancia del Método Científico en Ciencia de Datos


El método científico es fundamental en la ciencia de datos, ya que permite tomar decisiones basadas en evidencia en lugar de suposiciones. Desde la formulación de preguntas hasta el análisis de datos y la validación de modelos, cada paso del método científico se aplica en el proceso de extracción de conocimiento a partir de los datos.

Ejemplo en Ciencia de Datos

Supongamos que un analista de datos desea investigar si el tiempo de estudio influye en las calificaciones de los estudiantes:

- Problema: ¿El tiempo de estudio afecta el rendimiento académico?
- Hipótesis: A mayor tiempo de estudio, mejores calificaciones.
- Experimentación: Recopilar datos sobre el tiempo de estudio y las calificaciones de un grupo de estudiantes.
- Análisis: Aplicar técnicas estadísticas como correlación o regresión para determinar la relación entre ambas variables.
- Conclusión: Si los resultados muestran una relación positiva y significativa, la hipótesis se acepta; de lo contrario, se rechaza o se replantea.

Este enfoque permite validar afirmaciones con base en datos reales, evitando sesgos y proporcionando información confiable para la toma de decisiones.



MÉTODOS DE INVESTIGACIÓN: CUALITATIVOS. CUANTITATIVOS

Los métodos de investigación permiten explorar, analizar y obtener conocimiento sobre fenómenos en diversas disciplinas. Dependiendo del enfoque del estudio, se pueden emplear métodos cualitativos, cuantitativos o una combinación de ambos (método mixto).

Métodos Cualitativos

Los métodos cualitativos buscan comprender y describir fenómenos a través de la interpretación de datos no numéricos. Se centran en el significado, la experiencia y la percepción de los individuos en un contexto determinado.

- Enfoque: Exploración profunda de fenómenos sociales, culturales y psicológicos.
- Ejemplo de técnicas: Entrevistas a profundidad, grupos focales, observaciones etnográficas, análisis de discurso.
- Aplicación: Estudios sobre comportamiento humano, percepciones y experiencias subjetivas.

Métodos Cuantitativos

Los métodos cuantitativos se basan en la recolección y análisis de datos numéricos para identificar patrones, relaciones y tendencias en un fenómeno. Permiten realizar mediciones objetivas y generalizables.

- Enfoque: Medición de variables, análisis estadístico y comprobación de hipótesis.
- Ejemplo de técnicas: Encuestas con escalas numéricas, experimentos controlados, análisis de datos secundarios.
- Aplicación: Estudios en ciencias exactas, economía, psicometría y epidemiología.



IMPORTANCIA DEL PENSAMIENTO ESTADÍSTICO EN LA CIENCIA

El pensamiento estadístico es fundamental en la investigación, ya que proporciona herramientas para interpretar datos y validar resultados.

- Toma de Decisiones Basada en Datos: Permite extraer conclusiones confiables a partir del análisis de información.
- Reducción de Incertidumbre: Cuantifica la incertidumbre y el error en mediciones, mejorando la precisión del estudio.
- Validación de Resultados: Determina si los hallazgos son estadísticamente significativos o si pueden deberse al azar.

LA ESTADÍSTICA EN EL DISEÑO DE LA INVESTIGACIÓN

El diseño de investigación requiere una planificación cuidadosa para garantizar la validez y confiabilidad de los resultados.

- Diseño Experimental: Define cómo se recolectarán y analizarán los datos para probar hipótesis.
- Muestreo: Selección de una muestra representativa que permita generalizar los resultados a una población más amplia.
- Control de Variables: Minimización de sesgos y factores externos que puedan influir en los resultados.

En conclusión, tanto los métodos cualitativos como cuantitativos tienen un papel esencial en la investigación. La elección del método depende de la naturaleza del problema a estudiar y de los objetivos del análisis.



CONCEPTOS FUNDAMENTALES DE LA ESTADÍSTICA INFERENCIAL

La inferencia estadística es el proceso mediante el cual se realizan generalizaciones o conclusiones acerca de una población, basándose en los datos obtenidos de una muestra representativa. Es una herramienta fundamental en estadística, ya que permite hacer afirmaciones sobre grandes grupos sin tener que estudiar todos los elementos de una población.

Importancia de la Inferencia Estadística

La inferencia estadística es crucial porque:

- **Predicciones:** Permite realizar predicciones sobre eventos futuros o comportamientos de una población a partir de datos actuales.
- **Toma de decisiones:** Facilita la toma de decisiones fundamentadas en datos, minimizando el riesgo de error en situaciones de incertidumbre.

Aplicaciones de la Inferencia Estadística


La inferencia estadística tiene aplicaciones clave en diversas disciplinas:

- **Medicina:** En la evaluación de la efectividad de tratamientos, donde se realizan ensayos clínicos y se usa la inferencia para extrapolar los resultados a la población en general.
- **Negocios:** En la predicción de ventas futuras, identificando patrones a partir de datos históricos y utilizando inferencias para estimar el comportamiento futuro del mercado.
- **Política y Sociología:** Para realizar estudios y encuestas que permitan hacer proyecciones sobre el comportamiento de la población.

Ejemplo de Inferencia Estadística

Imaginemos que un laboratorio ha desarrollado un nuevo fármaco para reducir la presión arterial y desea probar su efectividad. La hipótesis que se plantea podría ser:

- **Hipótesis:** El nuevo fármaco reduce significativamente la presión arterial en pacientes con hipertensión.



Para probar esta hipótesis, se toma una muestra representativa de pacientes y se mide su presión arterial antes y después del tratamiento. A través de una prueba estadística (por ejemplo, una prueba t para muestras emparejadas), se determina si la reducción de la presión arterial es estadísticamente significativa, lo que nos permitiría hacer una generalización sobre la efectividad del fármaco para toda la población de pacientes con hipertensión.

EL CONCEPTO DE SIGNIFICANCIA

La significancia estadística es un concepto clave en la inferencia estadística que nos ayuda a determinar si los resultados de un experimento o análisis son probablemente reales o si podrían haberse obtenido por azar. Se evalúa utilizando el valor p (p-value), que es la probabilidad de obtener un resultado igual o más extremo que el observado, bajo la suposición de que la hipótesis nula es verdadera.

Valor p

El valor p es una medida de la probabilidad de error en un experimento. Si p es muy bajo, indica que es poco probable que el resultado se haya producido por azar y, por lo tanto, se puede rechazar la hipótesis nula en favor de la hipótesis alternativa.

$p < 0.05$: Resultado significativo.

Ilustración 1 Valor p

Ejemplo: Si obtenemos un valor p de 0.03, esto significa que hay solo un 3% de probabilidad de que los resultados obtenidos se deban al azar, lo que sugiere que los resultados son estadísticamente significativos. En este caso, podemos rechazar la hipótesis nula (por ejemplo, que no haya efecto) con un nivel de confianza del 97%.



DISEÑO DE EXPERIMENTOS

El diseño de experimentos es un proceso sistemático utilizado para planificar, estructurar y llevar a cabo experimentos con el objetivo de obtener datos válidos y confiables. Este diseño se enfoca en controlar y manipular variables para observar sus efectos, ayudando a establecer relaciones causales y a tomar decisiones basadas en evidencia.

Objetivo

1. Controlar variables: Asegurar que las variables que no se están investigando no afecten los resultados.
2. Observar efectos: Medir cómo las modificaciones en las variables independientes afectan las variables dependientes.
3. Reducir sesgos y errores: Minimizar los factores externos que puedan influir en los resultados, asegurando que los datos sean representativos.

Componentes

1. Hipótesis: Declaración clara que se quiere probar o refutar a través del experimento.
2. Variables:
 - Independientes: Son las variables que el experimentador manipula para estudiar su efecto sobre otras variables.
 - Dependientes: Son las variables que se miden y que se espera que cambien debido a las modificaciones de las variables independientes.
3. Grupos:
 - Grupo Experimental: El grupo que recibe el tratamiento o manipulación.
 - Grupo de Control: El grupo que no recibe el tratamiento o que recibe un tratamiento estándar, utilizado para comparar y medir los efectos del tratamiento.
4. Replicación: Repetir el experimento bajo las mismas condiciones para garantizar la validez y consistencia de los resultados.
5. Aleatorización: Asignar al azar las unidades experimentales a los grupos para evitar sesgos.

Ejemplo en Ciencia de Datos

En ciencia de datos, se diseñan experimentos para evaluar el desempeño de modelos y algoritmos.

Un ejemplo podría ser:

- Hipótesis: Un nuevo algoritmo de clasificación mejora la precisión de un modelo.
- Variables: El algoritmo es la variable independiente (lo que se manipula), y la precisión del modelo es la variable dependiente (lo que se mide).
- Grupos: Se comparan dos grupos: uno que utiliza el algoritmo actual y otro que utiliza el nuevo algoritmo.
- Replicación: El experimento se repite varias veces para asegurarse de que los resultados son consistentes y no se deben a una coincidencia o variabilidad aleatoria.

Pasos Principales en el Diseño de Experimentación:

1. Plantear una pregunta o problema de investigación.
2. Formular una hipótesis a probar.
3. Seleccionar las variables a manipular (independientes) y las que se medirán (dependientes).
4. Asignar aleatoriamente a los sujetos u objetos a los diferentes grupos de experimentación.
5. Realizar el experimento, recopilando datos de forma sistemática.
6. Analizar los resultados y sacar conclusiones basadas en los datos obtenidos.

Relación con el Método Científico y la Estadística

El método científico y la estadística son esenciales en la ciencia de datos para realizar investigaciones válidas. Mientras que el método científico proporciona un marco estructurado para llevar a cabo la investigación de manera sistemática, la estadística permite analizar, interpretar y validar los resultados obtenidos. En conjunto, ambas herramientas aseguran que las conclusiones sean basadas en evidencia objetiva y no en suposiciones o intuiciones, formando la base para una toma de decisiones informada y confiable en diversos campos.

El método científico y la estadística son herramientas fundamentales en ciencia de datos. El primero proporciona un marco para la investigación sistemática, mientras que la segunda permite analizar y validar los resultados. Juntos, forman la base para la toma de decisiones basada en evidencia.

ACTIVIDAD PRÁCTICA GUIADA: FORMULACIÓN DE HIPÓTESIS Y DISEÑO DE UN EXPERIMENTO

En esta actividad práctica, los alumnos aprenderán a formular una hipótesis sobre un problema de su interés y a diseñar un experimento para probarla. Este proceso es fundamental en el método científico y en la ciencia de datos, ya que permite validar o refutar ideas basadas en evidencia empírica.

Paso 1: Identificar un Problema de Interés

Instrucción: Piensa en un problema o pregunta que te interese. Puede ser relacionado con tu área de estudio, trabajo o vida cotidiana.

Ejemplo:

- Problema: ¿El tiempo de estudio afecta las calificaciones de los estudiantes?
- Contexto: En un curso de ciencia de datos, los estudiantes tienen diferentes hábitos de estudio.

Paso 2: Formular una Hipótesis

Instrucción: Propón una hipótesis que responda al problema. Una hipótesis es una afirmación que puede ser probada o refutada.

Estructura de la Hipótesis:

- Hipótesis Nula (H_0): No hay relación entre las variables.
- Hipótesis Alternativa (H_1): Existe una relación entre las variables.

Ejemplo:

- Hipótesis Nula (H_0): El tiempo de estudio no afecta las calificaciones de los estudiantes.
- Hipótesis Alternativa (H_1): A mayor tiempo de estudio, mejores calificaciones.

Paso 3: Definir las Variables

Instrucción: Identifica las variables involucradas en tu hipótesis.

- Variable Independiente: La que manipulas o cambias (tiempo de estudio).
- Variable Dependiente: La que mides u observas (calificaciones).

Ejemplo:

- Variable Independiente: Horas de estudio por semana.
- Variable Dependiente: Calificación final en el curso.

Paso 4: Diseñar el Experimento

Instrucción: Diseña un experimento para probar tu hipótesis. Considera los siguientes aspectos:

- Población y Muestra: ¿Quiénes participarán en el experimento?
- Recolección de Datos: ¿Cómo medirás las variables?
- Control de Variables: ¿Cómo evitarás que otros factores afecten los resultados?

Ejemplo:

- Población: Estudiantes de un curso de ciencia de datos.
- Muestra: 50 estudiantes seleccionados al azar.
- Recolección de Datos:
 - Registrar las horas de estudio semanales de cada estudiante.
 - Obtener las calificaciones finales del curso.
- Control de Variables:
 - Asegurar que todos los estudiantes tengan acceso a los mismos recursos (libros, profesores, etc.).
 - Excluir a estudiantes con circunstancias especiales (enfermedades, etc.).

Paso 5: Realizar el Análisis de Datos

Instrucción: Define cómo analizarás los datos para probar tu hipótesis.

- Método Estadístico: ¿Qué prueba usarás? (por ejemplo, correlación, regresión lineal).
- Herramientas: ¿Qué software o librerías usarás? (por ejemplo, Python, Pandas, SciPy).

Ejemplo:

- Método Estadístico: Correlación de Pearson para medir la relación entre horas de estudio y calificaciones.
- Herramientas: Usar Python con las librerías Pandas y SciPy.

```
import pandas as pd
from scipy.stats import pearsonr

# Datos de ejemplo
datos = {
    'Horas_Estudio': [10, 15, 20, 25, 30, 35, 40, 45, 50, 55],
    'Calificaciones': [60, 65, 70, 75, 80, 85, 90, 95, 100, 105]
}

df = pd.DataFrame(datos)

# Calcular correlación
correlacion, p_valor = pearsonr(df['Horas_Estudio'], df['Calificaciones'])
print(f"Correlación: {correlacion}, p-valor: {p_valor}")
```

Ilustración 2 Ejemplo método estadístico con Python y SciPy

Explicación:

Importar librerías

```
import pandas as pd
from scipy.stats import pearsonr
```

Ilustración 3 Importar librerías

- pandas: Se utiliza para manejar datos en forma de tablas (DataFrames).
- pearsonr: Función de la librería scipy.stats que calcula la correlación de Pearson y su valor p.

Crear un Conjunto de Datos

```
# Datos de ejemplo
datos = {
    'Horas_Estudio': [10, 15, 20, 25, 30, 35, 40, 45, 50, 55],
    'Calificaciones': [60, 65, 70, 75, 80, 85, 90, 95, 100, 105]
}
```

Ilustración 4 Creación conjunto de datos

Se crea un diccionario llamado datos que contiene dos listas:

- Horas_Estudio: Representa el número de horas que un estudiante dedica al estudio por semana.
- Calificaciones: Representa las calificaciones obtenidas por los estudiantes.

Convertir datos en DataFrame

```
df = pd.DataFrame(datos)
```

Ilustración 5 Convertir datos en DataFrame

- Se convierte el diccionario datos en un DataFrame de Pandas. Un DataFrame es una estructura de datos tabular similar a una hoja de cálculo.
- El DataFrame tendrá dos columnas: Horas_Estudio y Calificaciones.

Calcular la Correlación de Pearson

```
# Calcular correlación  
correlacion, p_valor = pearsonr(df['Horas_Estudio'], df['Calificaciones'])
```

Ilustración 6 Calcular correlación de Pearson

pearsonr: Calcula la correlación de Pearson entre las dos columnas del DataFrame.

- correlación: Es un valor entre -1 y 1 que indica la fuerza y dirección de la relación lineal.
 - 1: Correlación positiva perfecta.
 - -1: Correlación negativa perfecta.
 - 0: No hay correlación lineal.
- p_valor: Es un valor que indica la significancia estadística de la correlación.
 - Si $p < 0.05$, la correlación es estadísticamente significativa

Mostrar los Resultados

```
print(f"Correlación: {correlacion}, p-valor: {p_valor}")
```

Ilustración 7 Mostrar resultados

Se imprime el valor de la correlación y el p-valor

Salida

```
Correlación: 1.0, p-valor: 0.0
```

Ilustración 8 Salida

Paso 6: Interpretar los Resultados

Instrucción: Analiza los resultados del experimento y determina si apoyan o refutan tu hipótesis.

- Correlación: Si es cercana a 1 o -1, indica una relación fuerte.
- p-valor: Si es menor que 0.05, el resultado es estadísticamente significativo.

Ejemplo:

- Correlación = 1.0: Significa que a medida que aumentan las horas de estudio, las calificaciones también aumentan de manera lineal.
- p-valor = 0.0: Confirma que esta relación no es producto del azar, sino que es estadísticamente significativa.



Paso 7: Concluir y Comunicar los Hallazgos

Instrucción: Escribe una conclusión basada en los resultados y comunica tus hallazgos de manera clara.

Ejemplo:

Conclusión: Los resultados apoyan la hipótesis de que, a mayor tiempo de estudio, mejores calificaciones. Esto sugiere que el tiempo de estudio es un factor importante en el rendimiento académico.

Conclusión del Ejercicio

- Aprendiste a formular una hipótesis y diseñar un experimento para probarla.
- Practicaste la identificación de variables, recolección de datos y análisis estadístico.
- Este proceso es esencial para la investigación científica y la toma de decisiones basada en datos.