

# Seminario de discusión.

Tema 3: Bootstrap y Jackknife

- Métodos de permutación
- Inferencia por verosimilitud

---

MTRA. MELODY TREVIÑO RODRIGUEZ

# Bootstrap y Jackknife

---

Son técnicas estadísticas que buscan estimar la distribución de una estadística de interés mediante el **Re-muestreo** de los datos disponibles.

Ambos métodos son herramientas poderosas que permiten realizar inferencias robustas y calcular intervalos de confianza de manera más precisa, especialmente cuando no se cumplen ciertas condiciones de normalidad o cuando se dispone de muestras pequeñas.

Ambos métodos son particularmente útiles cuando se enfrentan a problemas de sesgo, varianza y distribuciones no normales, brindando una perspectiva más completa de la incertidumbre asociada con una estadística determinada. Estas técnicas son ampliamente utilizadas en la investigación y el análisis de datos para realizar inferencias más **sólidas y confiables**.

# Bootstrap y Jackknife

## Bootstrap

---

- **Concepto:** El método Bootstrap implica la creación de múltiples muestras (bootstrap muestras) a partir de los datos originales mediante el muestreo con reemplazo.
- **Proceso:**
  - Seleccionar aleatoriamente observaciones de la muestra original con reemplazo para formar una muestra bootstrap.
  - Calcular la estadística de interés en cada muestra bootstrap.
  - Repetir el proceso muchas veces para obtener una distribución de la estadística de interés.
- **Utilidad:** Proporciona una aproximación empírica de la distribución muestral de la estadística y permite construir intervalos de confianza y realizar pruebas de hipótesis de manera no paramétrica.

# Bootstrap y Jackknife

## Jackknife

---

- **Concepto:** El método Jackknife consiste en generar múltiples subconjuntos de datos eliminando una observación a la vez y calculando la estadística de interés en cada caso.
- **Proceso:**
  - Eliminar una observación a la vez para crear conjuntos de datos más pequeños.
  - Calcular la estadística de interés en cada conjunto de datos modificado.
  - Obtener una estimación puntual y el sesgo de la estadística.
- **Utilidad:** Proporciona estimaciones del sesgo y la varianza de la estadística, lo que puede ayudar a ajustar sesgos sistemáticos y mejorar la precisión de los resultados.

# Métodos de permutación

(pruebas de permutación o tests de aleatorización)

## ¿Qué son?

---

Son técnicas estadísticas utilizadas en inferencia para evaluar la significancia de un resultado observado mediante la generación de múltiples permutaciones o reordenamientos aleatorios de los datos.

Estos métodos son particularmente útiles cuando los supuestos de otros métodos paramétricos no se cumplen o cuando se dispone de muestras pequeñas.

Los métodos de permutación ofrecen una aproximación empírica a la distribución de la estadística de interés sin hacer suposiciones específicas sobre la forma de la distribución subyacente.

# Métodos de permutación

(pruebas de permutación o tests de aleatorización)

## □ Procedimiento

**Hipótesis Nula ( $H_0$ ):** Formular la hipótesis nula que se quiere probar, por ejemplo, que no hay diferencia entre dos grupos o que no hay asociación entre dos variables.

**Estadística de Interés:** Seleccionar una estadística que sea relevante para probar la hipótesis nula. Esto podría ser la diferencia entre medias, la correlación, la proporción, u otra medida según el contexto.

**Permutaciones:** Realizar todas las posibles permutaciones o reordenamientos de los datos, manteniendo constante el tamaño de las muestras. Esto implica mezclar los datos entre los grupos o permutar los valores de una variable, dependiendo de la pregunta de investigación.

**Calcular la Estadística en Cada Permutación:** Calcular la estadística de interés para cada permutación.

**Distribución Nula:** La distribución nula se forma a partir de las estadísticas calculadas en las permutaciones, y representa la variabilidad bajo la hipótesis nula.

**Comparar Estadística Observada:** Comparar la estadística observada en los datos reales con la distribución nula. Si la estadística observada se encuentra en las colas extremas de la distribución nula, se puede rechazar la hipótesis nula en favor de la alternativa.

# Inferencia por verosimilitud

---

Enfoque estadístico utilizado para estimar los parámetros de un modelo estadístico. Se basa en la verosimilitud, que mide la probabilidad de observar los datos dados ciertos valores de los parámetros del modelo.

La idea central es encontrar los valores de los parámetros que hacen que los datos observados sean más probables.

La función de verosimilitud se define como la probabilidad de observar los datos, condicionada a un conjunto específico de valores para los parámetros del modelo. La estimación por verosimilitud implica encontrar los valores de los parámetros que maximizan esta función.

La inferencia por verosimilitud es ampliamente utilizada en estadística y se aplica en diversos contextos, como la regresión, la clasificación, y la estimación de parámetros en modelos probabilísticos. Es especialmente útil cuando la distribución de los datos se ajusta a un modelo paramétrico y se busca obtener estimaciones eficientes de los parámetros.

# Inferencia por verosimilitud

## Procedimiento:

---

1. **Definir el Modelo Estadístico:** Especificar un modelo estadístico que describe la distribución de los datos.
2. **Formular la Función de Verosimilitud:** Escribir la función de verosimilitud, que es la probabilidad de observar los datos bajo diferentes valores de los parámetros del modelo.
3. **Maximizar la Función de Verosimilitud:** Encontrar los valores de los parámetros que maximizan la función de verosimilitud. Esto se puede hacer mediante métodos matemáticos o algoritmos de optimización.
4. **Obtener Estimaciones e Intervalos de Confianza:** Una vez que se han encontrado los valores estimados de los parámetros, se pueden calcular intervalos de confianza y realizar inferencias sobre los parámetros del modelo.
5. **Evaluar la Bondad de Ajuste:** Es importante evaluar la calidad del ajuste del modelo a los datos observados.



## Tarea

---

### 1) Investigar y describir un ejercicio de Inferencia por verosimilitud:

Investigar un ejemplo paso a paso que se realice en R o Python y describir lo que se hace en cada paso y aportar las conclusiones personales a las que llegaste con el resultado obtenido. Se entregará vía Moodle y ahí se especificará la dinámica.

### 2) En la próxima sesión (Sesión 6 – 24 de febrero, 2024) se llevará a cabo un **examen parcial** sobre los temas:

- Análisis de Varianza (ANOVA)
- Bootstrap y Jackknife
- Métodos de permutación e
- Inferencia por verosimilitud

### 3) Buscar al menos un artículo de investigación (papers, tesis, etc.) sobre regresión lineal y realizar un reporte de lectura. Se discutirán los hallazgos en la Sesión 7 (2 de marzo, 2024).

El artículo que estudies deberá ser diferente a de tus compañeros, en Moodle se especificará la dinámica y las fechas de entrega.