

# **Subject: Probability and Statistics**

## **Class VIII: Pruebas estadísticas paramétricas**



**UdeA**

**Bioengineering**

**Francisco José Campuzano Cardona**

Bioengineerer, MSc in Engineering

Veamos algunas pruebas no paramétricas para una sola muestra:

### Pruebas de Normalidad

La normalidad se puede analizar a través de pruebas de hipótesis, en este caso se compara una muestra, con una distribución estándar, la normal.

- Kolmogorov-Smirnoff ( $n > 50$ )
- Shapiro-Wilks ( $n < 50$ )

***H<sub>0</sub>**: Los datos provienen de una distribución normal*

***H<sub>1</sub>**: Los datos no provienen de una distribución normal*

Como la normalidad es un supuesto de muchas pruebas paramétricas, se usa estas pruebas para probarlo.

## Transformaciones Box-Cox

Como se mencionó en clases pasadas, los datos atípicos en ocasiones no se pueden corregir, y tampoco eliminar. Este resulta ser el caso más frecuente. Por este motivo se tienen alternativas, es hacer una transformación Box-Cox, o transformaciones de potencias. Box-Cox solo aplica para variables positivas, en otro casos se debe hacer una transformación de potencias por otro método como Yeo-Johnson

Sea  $X$  una variable con datos atípicos, entonces sea la siguiente transformación:

$$y = \frac{x^{\lambda} - 1}{\lambda}, \lambda \neq 0$$

$$y = \ln(x), \lambda = 0$$

Donde  $\lambda$  es cualquier número que permita mejorar la distribución de datos.

## Normalización

Otra alternativa para mejorar la distribución de datos es la normalización. Se selecciona un valor apropiado respecto del cual normalizar. Por ejemplo normalizar respecto de la media, o de algún valor de referencia que existe de la variable, o respecto del valor máx. no atípico.

Sea  $X$  una variable con datos atípicos, entonces sea la siguiente transformación:

$$Y = \frac{X}{A}$$

Donde  $A$  es el valor conveniente respecto del que se decide la normalización.

## Prueba Bionomial

Es utilizada para evaluar si la proporción de éxitos en una muestra difiere de un valor esperado. Se basa en la distribución binomial, que modela la probabilidad de obtener exactamente  $k$  éxitos en  $n$  ensayos independientes, donde cada ensayo tiene solo dos posibles resultados (éxito o fracaso). Es eficiente con pocos datos.

***H<sub>0</sub>:*** La proporción de éxitos es  $p$

***H<sub>1</sub>:*** La proporción de éxitos no es  $p$  ( se puede evaluar para menor, mayor, o diferente)

## Prueba de signos – Comparación a una mediana

Es un caso especial de la **prueba binomial**. Busca compara la mediana de la muestra con un valor de referencia sin asumir normalidad. La prueba realiza la diferencia de todos los valores a la mediana de interés, y determina cuales son mayores y cuales menores, y hace un ajuste a una distribución binomial para determinar la probabilidad de éxito o fallo, es decir, de que sean mayores o menores.

***H<sub>0</sub>**: La mediana de las diferencias es 0*

***H<sub>1</sub>**: La mediana de las diferencias es diferentes de 0*

Ejemplo biomédico:

Se mide la saturación de oxígeno en sangre de 20 pacientes con EPOC y se evalúa si la mediana es menor a 95%, sin asumir normalidad.

## Prueba de signos – Muestras pareadas

También se puede usar en muestras pareadas, si se evalúa el efecto antes y después de algún tratamiento, se calcula la diferencia y se determina el número de mayores y menores, finalmente se compara con una mediana determinada.

**H<sub>0</sub>:** La mediana de las diferencias es 0

**H<sub>1</sub>:** La mediana de las diferencias es diferente de cero

Ejemplo biomédico:

Se quiere evaluar si el uso de un medicamento mejora el peso de pacientes obesos. Entonces se toman 20 pacientes, se les pesa antes y después del tratamiento, se determina el número de éxito., se hace una prueba binomial, con  $n$  = al número de diferencias positivas y negativas (0 se ignora) y  $p=0.5$  (la mediana es cero)

## Prueba Wilcoxon

Es una prueba alternativa a la prueba de signos. Tampoco asume normalidad. Es una alternativa a la prueba T para una muestra, cuando los datos no son normales.

***H<sub>0</sub>**: La mediana de las observaciones es  $M$*

***H<sub>1</sub>**: La mediana de las observaciones es diferente de  $M$*



## Prueba T para dos muestras:

Evalúa si la diferencia de la media de dos muestras es significativamente diferentes de cero, es decir si las medias son significativamente diferentes. Tiene el supuesto de normalidad en las variables, homocedasticidad e independencia:

***H<sub>0</sub>**: Las medias de las muestras son iguales*

***H<sub>1</sub>**: Las medias de las muestras son diferentes*

Ejemplo: Se quiere determinar si la adición de celulosa a películas de PVA aumenta la rigidez. Entonces se compara un grupo control, con un grupo que tenga celulosa.

## Prueba T de comparaciones múltiples.

Evalúa si la diferencia de la media es significativamente diferente de cero, en más de dos muestras, haciendo comparaciones por pares. Tiene el supuesto de normalidad en las variables, independencia y varianza constante.

***H<sub>0</sub>**: Las medias de las muestras son iguales*

***H<sub>1</sub>**: Las medias de las muestras son diferentes*

Requiere de correcciones: Ejemplo Bonferroni, Holm... porque la probabilidad de cometer un error tipo I aumenta al repetirse la prueba varias veces.

## Prueba T para muestras pareadas

Se usa cuando los sujetos experimentales son medidos dos veces en diferentes momentos. Se busca ver si hay diferencias antes y después de un tratamiento. Aquí no hay lugar al supuesto de independencia. La distribución de las diferencias debe ser normal.

***H<sub>0</sub>**: Las medias antes y después son iguales*

***H<sub>1</sub>**: Las medias antes y después son diferentes*

Ejemplo: Se quiere determinar si el ejercicio cambia significativamente la frecuencia cardiaca. Entonces se toma la FC antes y después del ejercicio.

## Prueba Wilcoxon para muestras pareadas

Se usa cuando los sujetos experimentales son medidos dos veces en diferentes momentos. Se busca ver si hay diferencias antes y después de un tratamiento. Pero no se cumple con el supuesto de normalidad en la diferencia.

***H<sub>0</sub>**: Las medias antes y después son iguales*

***H<sub>1</sub>**: Las medias antes y después son diferentes*

Ejemplo: Se quiere determinar si el ejercicio físico cambia significativamente la frecuencia cardiaca. Entonces se toma la FC antes y después de hacer ejercicio.

## Prueba U de Mann-Whitney o Prueba Suma de Rangos de Wilcoxon

Se usa para comparar la mediana de distribuciones que no presentan normalidad, pero que son independientes.

***H<sub>0</sub>**: Las medianas son iguales*

***H<sub>1</sub>**: Las medianas son diferentes*

Ejemplo: Se requiere comparar si la grasa abdominal es diferente en pacientes blancos y no blancos, pero no se encuentra una distribución normal en los datos.

## Validación de la homocedasticidad

La homocedasticidad se puede analizar a través de pruebas de hipótesis

- Levene (Robusta ante distribuciones menos normales)
- Barttlet (Supuesto de normalidad estricto)

***H<sub>0</sub>:*** Los dos grupos de datos presentan varianza constante

***H<sub>1</sub>:*** Los dos grupos de datos no presentan varianza constante.

## Ejercicio de clase:

En la carpeta correspondiente a la clase encontrará un enunciado para el ejercicio de clase del día de hoy. El ejercicio no tiene nota, pero se recomienda realizar el ejercicio en clase.

# Subject: Probability and Statistics



**UdeA**

**Bioengineering**

¡Thanks!

Francisco José Campuzano Cardona

Bioengineering. MSc in Engineering