

# **Clase 05 - Inferencia estadística**

## **Curso Análisis de Datos con R para Biociencias.**

Dra. María Angélica Rueda | <https://genomics.pucv.cl/>

Pontificia Universidad Católica de Valparaíso

15 January 2023

# PLAN DE LA CLASE

## 1.- Introducción

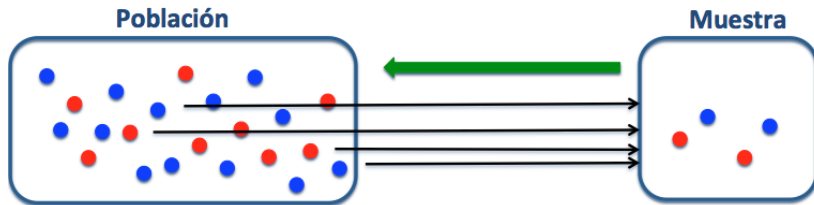
- ▶ ¿Qué es la inferencia estadística?
- ▶ ¿Cómo someter a prueba una hipótesis?
- ▶ Pruebas paramétricas
- ▶ Interpretar resultados de análisis de datos con R.

## 2.- Práctica con R y Rstudio cloud

- ▶ Someter a prueba diferentes hipótesis estadísticas.
- ▶ Realizar gráficas avanzadas con ggplot2.

# ¿QUÉ ES LA INFERENCIA ESTADÍSTICA?

**Inferencia estadística** : Son procedimientos que permiten obtener o extraer conclusiones sobre los parámetros de una población a partir de una muestra de datos tomada de ella.



¿Qué inferencia puede hacer de los datos de esta población?  
¿Qué ocurre si la muestra no es aleatoria?

# INFERENCIA ESTADÍSTICA

**¿Par qué es importante la inferencia estadística?**

- ▶ **Es más económico que hacer un Censo.**  
¿Cuántas especies hay en una bahía, en una laguna, en un bosque?
- ▶ **Bajo ciertos supuestos permite hacer afirmaciones.**  
Con fertilizante A las plantas crecen más que con fertilizante B.
- ▶ **Bajo ciertos supuestos permite hacer predicciones.**  
Mujeres con genotipos mutante del gen BCRA1 tiene 7 veces más probabilidad de tener cáncer ovárico que mujeres con genotipo normal.

# INFERENCIA ESTADÍSTICA: MÉTODOS

Los métodos de inferencia estadística que revisaremos en este curso son:

- 1. Estimación de parámetros a partir de una muestra.**

Ej. Estimación puntual o por intervalos

- 2. Pruebas de contraste de hipótesis.**

Ej. Prueba de correlación, comparación de medias, Anova.

- 3. Modelamiento predictivo.**

Ej. Regresión lineal, logística, machine learning, validación cruzada.

# PRUEBAS DE HIPÓTESIS

## Objetivo

Realizar una afirmación acerca del valor de un parámetro, usualmente contrastando con alguna hipótesis.

## Hipótesis estadísticas

*Hipótesis nula* ( $H_0$ ) es una afirmación, usualmente de igualdad.

*Hipótesis alternativa* ( $H_A$ ) es una afirmación que se deduce de la observación previa o de los antecedentes de literatura y que el investigador cree que es verdadera.

## Ejemplo

$H_0$ : El nivel medio de cortisol es  $= 15$  microgramos por decilitro.

$H_A$ : El nivel medio de cortisol es  $> 15$  microgramos por decilitro.

# ¿POR QUÉ DOS HIPÓTESIS?

- ▶ Las pruebas estadísticas tienen como propósito someter a prueba una hipótesis nula con la intención de **rechazarla** o **refutarla** (Falsacionismo de Karl Popper).
- ▶ Por lo tanto, los datos nos dirán si **existen o no** evidencias para **rechazar la hipótesis nula**.

## ¿Por qué no simplemente aceptar la hipótesis alternativa?

- ▶ Porque pueden existir otros fenómenos no conocidos o no considerados en nuestro experimento que posteriormente permitan a otro investigador rechazar nuestra hipótesis alternativa.

Lectura complementaria: El método científico según Juan José Ibáñez, 2006

# ¿CUÁNDO RECHAZAR $H_0$ ?

## Regla de decisión

Rechazo  $H_0$  cuando la evidencia observada es poco probable que ocurra bajo el supuesto de que la hipótesis sea verdadera.

Generalmente  $\alpha = 0,05$  o  $0,01$ .

Es decir, rechazamos cuando el valor del estadístico está en el 5% inferior de la función de distribución muestral.



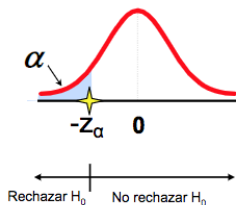
# PRUEBA DE HIPÓTESIS: UNA COLA O DOS COLAS

## Prueba unilateral izquierda

Ejemplo:

$$H_0: \mu \geq 3$$

$$H_A: \mu < 3$$

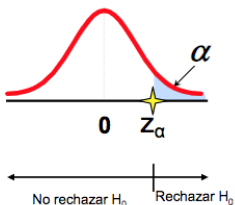


## Prueba unilateral derecha

Ejemplo:

$$H_0: \mu \leq 3$$

$$H_A: \mu > 3$$

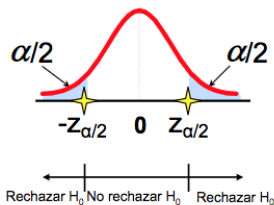


## Prueba bilateral

Ejemplo:

$$H_0: \mu = 3$$

$$H_A: \mu \neq 3$$



# ¿PUEDO COMETER UN ERROR EN LAS PRUEBAS DE HIPÓTESIS?

Por supuesto, siempre es posible llegar a una conclusión incorrecta.

## Tipos de errores

Tipo I ( $\alpha$ ) y tipo II ( $\beta$ ), ambos están inversamente relacionados.

Decisión	$H_0$ es cierta	$H_0$ es falsa
<i>Aceptamos <math>H_0</math></i>	Decisión correcta	Error tipo II
<i>Rechazamos <math>H_0</math></i>	Error tipo I	Decisión correcta

# TIPOS DE PRUEBAS ESTADÍSTICAS

Según la forma de la distribución de la variable aleatoria.

## 1. Métodos paramétricos

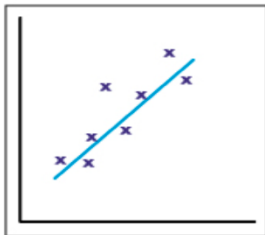
- ▶ Las pruebas de hipótesis usualmente asumen una distribución normal de la variable aleatoria.
- ▶ Útil para la mayoría de las variables cuantitativas continuas.

## 2. Métodos NO paramétricos

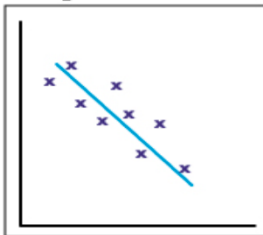
- ▶ Las pruebas de hipótesis no asumen una distribución normal de la variable aleatoria.
- ▶ Útil para todas las variables, incluyendo cuantitativas discretas y cualitativas.

# CORRELACIÓN ENTRE VARIABLES

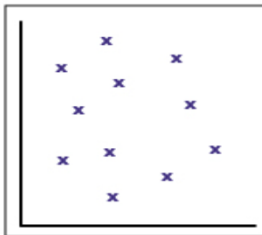
**Positive correlation**



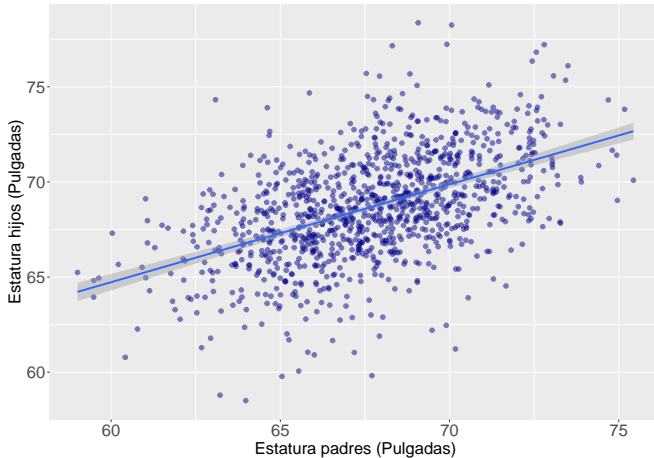
**Negative correlation**



**No correlation**



# ESTUDIO DE CASO: RELACIÓN ESTATURA PADRES - HIJOS



**Figure 1:** Relación estatura de padres e hijos evaluados por Karl Pearson en sus estudios de regresión.

# HIPÓTESIS PRUEBA DE CORRELACIÓN

## Hipótesis

$H_0 : \rho = 0$  ausencia de correlación.

$H_1 : \rho \neq 0$  existencia de correlación.

## Supuestos:

- 1) Las variables X e Y son continuas y su relación es lineal.
- 2) La distribución conjunta de (X,Y) es una distribución Bivariable normal.

# PRUEBA DE CORRELACIÓN DE PEARSON

```
cor.test(father.son$fheight, father.son$sheight)
```

```
##
```

```
## Pearson's product-moment correlation
```

```
##
```

```
## data: father.son$fheight and father.son$sheight
```

```
## t = 19.006, df = 1076, p-value < 2.2e-16
```

```
## alternative hypothesis: true correlation is not equal to 0
```

```
## 95 percent confidence interval:
```

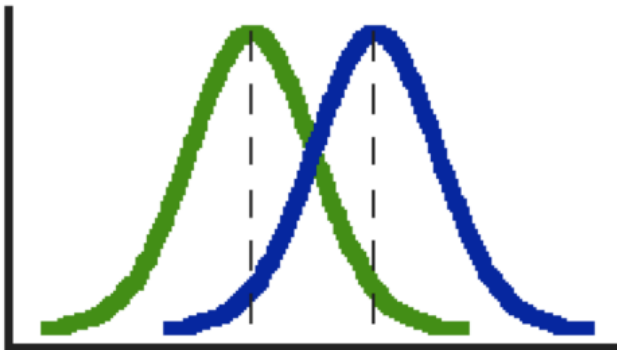
```
## 0.4552586 0.5447396
```

```
## sample estimates:
```

```
## cor
```

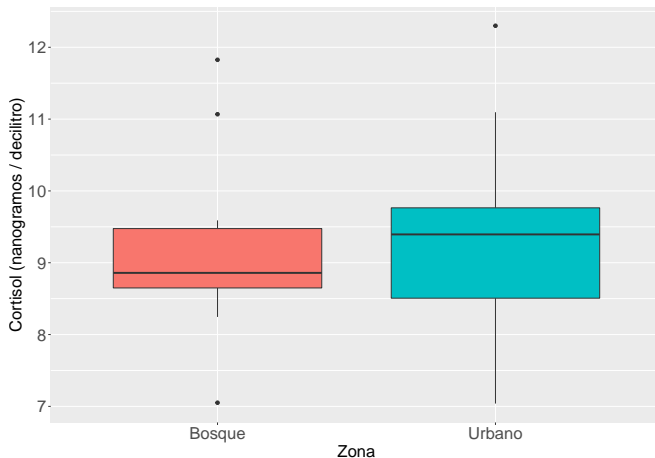
```
## 0.5013383
```

# PRUEBA DE COMPARACIÓN DE MEDIAS





# ESTUDIO DE CASO: CORTIZOL



**Figure 2:** Nivel de cortisol luego de caminar en zona urbana v/s zona boscosa.

Adaptado de Kobayashi, et al 2019

# HIPÓTESIS COMPARACIÓN DE MEDIAS

## Hipótesis

$$H_0 : \mu_1 = \mu_2.$$

$$H_1 : \mu_1 \neq \mu_2$$

## Supuestos

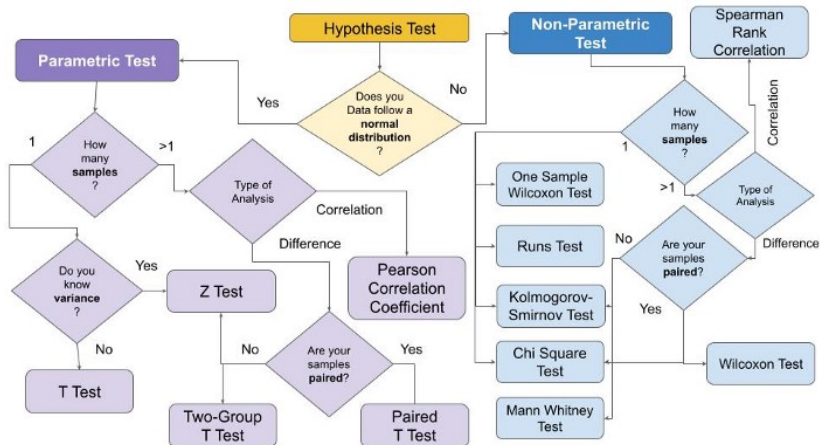
- 1) Las variables X es continua.
- 2) Distribución normal.

# PRUEBA DE T PARA DOS MUESTRAS INDEPENDIENTES

```
t.test(Cortisol ~ Zona, dat, alternative = c("two.sided"),  
       var.equal=TRUE)
```

```
##  
## Two Sample t-test  
##  
## data: Cortisol by Zona  
## t = -0.27153, df = 18, p-value = 0.7891  
## alternative hypothesis: true difference in means is not  
## 95 percent confidence interval:  
## -1.533583 1.182547  
## sample estimates:  
## mean in group Bosque mean in group Urbano  
## 9.202381 9.377899
```

# PRÁCTICA ANÁLISIS DE DATOS



# RESUMEN DE LA CLASE

1. **Conceptos básicos de inferencia estadística**
2. **Conceptos básicos de pruebas de hipótesis**
  - ▶ Hipótesis nula, hipótesis alternativa.
3. **Realizar pruebas de hipótesis**
  - ▶ Test de correlación.
  - ▶ Test de comparación de medias para 2 muestras independientes.
4. **Realizar gráficas avanzadas con ggplot2.**