

Tarea-04

Pruebas visuales y Permutación

Pruebas de permutación

1. La tabla de datos *Phillies2009* contiene información de la temporada de baseball 2009 para el equipo de Philadelphia *Phillies*, en este ejercicio queremos comparar los strikes (**StrikeOuts**) entre los juegos en casa y de visitante:
 - a) Realiza un *lineup* donde cada panel muestre un diagrama de caja y brazos para la variable de interés separando los juegos jugados en casa (home) y los juegos jugados fuera (away). ¿Puedes localizar los datos verdaderos entre los nulos?
 - b) Calcula el promedio de strikes por juego en cada grupo (donde grupo se define por Location).
 - c) Realiza una prueba de permutación para la diferencia de las medias. Grafica la distribución de referencia y calcula el valor *p* de dos colas.

```
# Lectura de datos
phillies <- read_csv("Phillies2009.csv")
```

```
## Error in read_csv("Phillies2009.csv"): could not find function "read_csv"
```

Pruebas pareadas

En este ejemplo buscamos comparar la diferencia entre dos medicinas para dormir. - ID es el identificador de paciente, y medicina_1 y medicina_2 son las horas extras de sueño vs. no usar medicina.
- Examina los datos.

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.1
## v tidyverse 1.1.1    v stringr 1.4.0
## v readr   1.3.1     vforcats 0.5.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()
```

```

dormir <- sleep %>%
  pivot_wider(names_from = group,
              names_prefix = "medicina_",
              values_from = extra)

dormir

## # A tibble: 10 x 3
##   ID    medicina_1 medicina_2
##   <fct>    <dbl>     <dbl>
## 1 1        0.7      1.9
## 2 2       -1.6      0.8
## 3 3       -0.2      1.1
## 4 4       -1.2      0.1
## 5 5       -0.1      -0.1
## 6 6        3.4      4.4
## 7 7        3.7      5.5
## 8 8        0.8      1.6
## 9 9        0        4.6
## 10 10       2       3.4

```

La pregunta de interés es si una medicina es mejor que otra para prolongar el sueño. Nótese que en este caso, no tenemos grupos, sino mediciones repetidas.

- Escribe la hipótesis nula.
- Nuestra estadística de interés es media de las diferencias entre las medicinas. Calcula la diferencia observada.
- Hay variación entre los pacientes. ¿Tenemos evidencia para rechazar que son iguales? ¿Cómo hacemos nuestra distribución de referencia?
- Haz una gráfica de la distribución de referencia y grafica encima el valor observado en los datos originales.
- Calcula el valor p (justifica porque a una o dos colas).

Valores- p

Cuando usamos simulación para pruebas de permutación, el valor- p de una cola se puede calcular como

$$\hat{P} = (X + 1)/(N + 1)$$

donde X es el número de estadísticas de prueba remuestreadas que son al menos tan extremas como la observada. Supongamos que el verdadero valor p (si pudiéramos hacer todas las permutaciones de manera exhaustiva) es p

\hat{P} - ¿Cuál es la varianza de \hat{P} ?

- ¿Cuál es la varianza de \hat{P}_2 para la prueba de dos lados? (suponiendo que p no es cercano a 0.5, donde p es el verdadero valor p para la prueba de una cola).