Clase 4.0
Análisis

Marcos Rosetti y Luis Pacheco-Cobos

Estadística y Manejo de Datos con R (EMDR) — Virtual

Estadística inferencial

Estadística inferencial

- `wilcox.test()` *dos* muestras, *independientes* o *pareadas*, **datos no paramétricos**.

```
?airquality
head(airquality)
```

```
wilcox.test(Ozone ~ Month, data = airquality %>%
              filter(Month %in% c(5, 8))) # paired = T si son muestras pareadas
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  Ozone by Month
## W = 127.5, p-value = 0.0001208
## alternative hypothesis: true location shift is not equal to 0
```

Estadística inferencial

- `kruskal.test()` *tres o más* muestras *independientes*, **datos no paramétricos**.

```
?PlantGrowth
head(PlantGrowth)
```

```
kruskal.test(weight ~ group, data = PlantGrowth)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  weight by group
## Kruskal-Wallis chi-squared = 7.9882, df = 2, p-value = 0.01842
```

```
# posthocs con pairwise.wilcox.test
pairwise.wilcox.test(PlantGrowth$weight, PlantGrowth$group, p.adjust.method = "BH")
```

```
##
##  Pairwise comparisons using Wilcoxon rank sum test with continuity correction
##
## data:  PlantGrowth$weight and PlantGrowth$group
##
##      ctrl  trt1
## trt1 0.199 -
## trt2 0.095 0.027
##
## P value adjustment method: BH
```

Estadística inferencial

- `friedman.test()` Tres o más muestras *repetidas*, **datos no paramétricos**.

```
data("selfesteem", package = "datarium")
?selfesteem
head(selfesteem)
```

```
# operador barra | marca el elemento que identifica las repeticiones
friedman.test(score ~ time | id, data = selfesteem %>%
                tidyr::gather(time, score, -id))
```

```
##
##  Friedman rank sum test
##
## data:  score and time and id
## Friedman chi-squared = 18.2, df = 2, p-value = 0.0001117
```

Estadística inferencial

- Correlación `cor()` entre dos vectores.

```
cor(airquality$Ozone,airquality$Wind, use = "complete.obs")
```

```
## [1] -0.6015465
```

Estadística inferencial

- Correlación `cor()` de una matriz o df.

```
cor(airquality, use = "complete.obs")
```

```
##                Ozone     Solar.R        Wind       Temp        Month
## Ozone    1.000000000  0.34834169 -0.61249658  0.6985414  0.142885168
## Solar.R  0.348341693  1.00000000 -0.12718345  0.2940876 -0.074066683
## Wind    -0.612496576 -0.12718345  1.00000000 -0.4971897 -0.194495804
## Temp     0.698541410  0.29408764 -0.49718972  1.0000000  0.403971709
## Month    0.142885168 -0.07406668 -0.19449580  0.4039717  1.000000000
## Day     -0.005189769 -0.05775380  0.04987102 -0.0965458 -0.009001079
##                 Day
## Ozone   -0.005189769
## Solar.R -0.057753801
## Wind     0.049871017
## Temp    -0.096545800
## Month   -0.009001079
## Day      1.000000000
```

Estadística inferencial

- `cor.test()` prueba de asociación o correlación entre muestras pareadas.

```
cor.test(women$height, women$weight, method = "pearson") # "spearman" si no paramétricos
```

```
##
##  Pearson's product-moment correlation
##
## data:  women$height and women$weight
## t = 37.855, df = 13, p-value = 1.091e-14
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9860970 0.9985447
## sample estimates:
##       cor
## 0.9954948
```

Estadística inferencial

- Ajuste de modelos lineales `lm()` y `cor.test()`

```
lm1 <- lm(Volume ~ Height, trees)
lm2 <- lm(Girth ~ Height, trees)
cor(lm1$residuals, lm2$residuals) # correlación controlando por Height
```

```
## [1] 0.9586123
```

Estadística inferencial

- `chisq.test()` prueba de independencia con frecuencias.

```
data(survey, package = "MASS")
?survey
head(survey)
```

```
tbl <- table(survey$Smoke, survey$Exer)
chisq.test(tbl)
```

```
##
##  Pearson's Chi-squared test
##
## data:  tbl
## X-squared = 5.4885, df = 6, p-value = 0.4828
```

Estadística inferencial

- `t.test()` para una muestra, compara contra la media poblacional.

```
t.test(scale(PlantGrowth$weight, center = F), mu = 1)
```

```
##
##  One Sample t-test
##
## data:  scale(PlantGrowth$weight, center = F)
## t = -1.0479, df = 29, p-value = 0.3033
## alternative hypothesis: true mean is not equal to 1
## 95 percent confidence interval:
##  0.9239545 1.0245198
## sample estimates:
## mean of x
## 0.9742371
```

Estadística inferencial

- `t.test()` para dos muestras independientes.

```
t.test(weight ~ group, data = PlantGrowth %>% filter(group != "trt2"))
```

```
##
##  Welch Two Sample t-test
##
## data:  weight by group
## t = 1.1913, df = 16.524, p-value = 0.2504
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.2875162  1.0295162
## sample estimates:
## mean in group ctrl mean in group trt1
##              5.032              4.661
```

Estadística inferencial

- `t.test(, paired=T)` para dos muestras pareadas.

```
t.test(weight ~ group, data = PlantGrowth %>% filter(group != "trt2"), paired = T)
```

```
##
##  Paired t-test
##
## data:  weight by group
## t = 0.99384, df = 9, p-value = 0.3463
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.4734609  1.2154609
## sample estimates:
## mean of the differences
##                   0.371
```

Estadística inferencial

- `aov()` para ANOVA o comparación de tres o más muestras.

```
aov(len ~ supp + as.factor(dose), ToothGrowth)
```

```
## Call:
##    aov(formula = len ~ supp + as.factor(dose), data = ToothGrowth)
##
## Terms:
##                       supp as.factor(dose) Residuals
## Sum of Squares   205.350        2426.434   820.425
## Deg. of Freedom        1               2        56
##
## Residual standard error: 3.82759
## Estimated effects may be unbalanced
```

Estadística inferencial

- ANOVA con interacción.

```
aov(len ~ supp * as.factor(dose), ToothGrowth)
```

```
## Call:
##    aov(formula = len ~ supp * as.factor(dose), data = ToothGrowth)
##
## Terms:
##                       supp as.factor(dose) supp:as.factor(dose) Residuals
## Sum of Squares   205.350        2426.434              108.319   712.106
## Deg. of Freedom        1               2                    2        54
##
## Residual standard error: 3.631411
## Estimated effects may be unbalanced
```

Estadística inferencial

- `summary()` también devuelve el resumen de un objeto, el ANOVA aquí.

```
tg.aov <- aov(len ~ supp * as.factor(dose), ToothGrowth)
summary(tg.aov)
```

```
##                      Df Sum Sq Mean Sq F value   Pr(>F)
## supp                  1  205.4   205.4  15.572 0.000231 ***
## as.factor(dose)       2 2426.4  1213.2  92.000  < 2e-16 ***
## supp:as.factor(dose)  2  108.3    54.2   4.107 0.021860 *
## Residuals            54  712.1    13.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Estadística inferencial

- Comparaciones *post hoc*.

```
tg.aov <- aov(len ~ supp * as.factor(dose), ToothGrowth)
tg.Tukey <- TukeyHSD(tg.aov)
tg.Tukey$supp
```

```
##       diff       lwr       upr        p adj
## VC-OJ -3.7 -5.579828 -1.820172 0.0002311828
```

```
tg.Tukey$`as.factor(dose)`
```

```
##         diff       lwr       upr        p adj
## 1-0.5  9.130  6.362488 11.897512 3.553066e-10
## 2-0.5 15.495 12.727488 18.262512 4.384271e-13
## 2-1    6.365  3.597488  9.132512 2.707572e-06
```

Licencia CC BY