

Tarea 07 Regresión y Correlación

Heidy Valdelamar Gonzalez

2022-11-17

Cargamos los paquetes que se usarán

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr  0.3.4
## v tibble  3.1.8      v dplyr  1.0.9
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(rcompanion)
```

1. Reiss et al. compararon los ensayos en el punto de atención y en el laboratorio hospitalario estándar para controlar a los pacientes que reciben un único anticoagulante o un régimen consistente en una combinación de anticoagulantes. Cuando se comparan dos técnicas de medición, es bastante habitual utilizar un análisis de regresión en el que una variable se utiliza para predecir otra. En el presente estudio, los investigadores obtuvieron medidas del cociente internacional normalizado (INR) mediante el análisis de muestras de sangre capilar y venosa recogidas de 90 sujetos que tomaban warfarina. El INR, utilizado especialmente cuando los pacientes reciben warfarina, mide la capacidad de coagulación de la sangre. Las pruebas de INR en el punto de atención se realizaron con el producto de ensayo CoaguChek. Las pruebas hospitalarias se realizaron con los ensayos estándar del laboratorio del hospital. Los autores utilizaron el nivel de INR del ensayo hospitalario para predecir el nivel de INR de CoaguChek. Las mediciones se encuentran en el archivo EXR_C09_S03_04.csv.

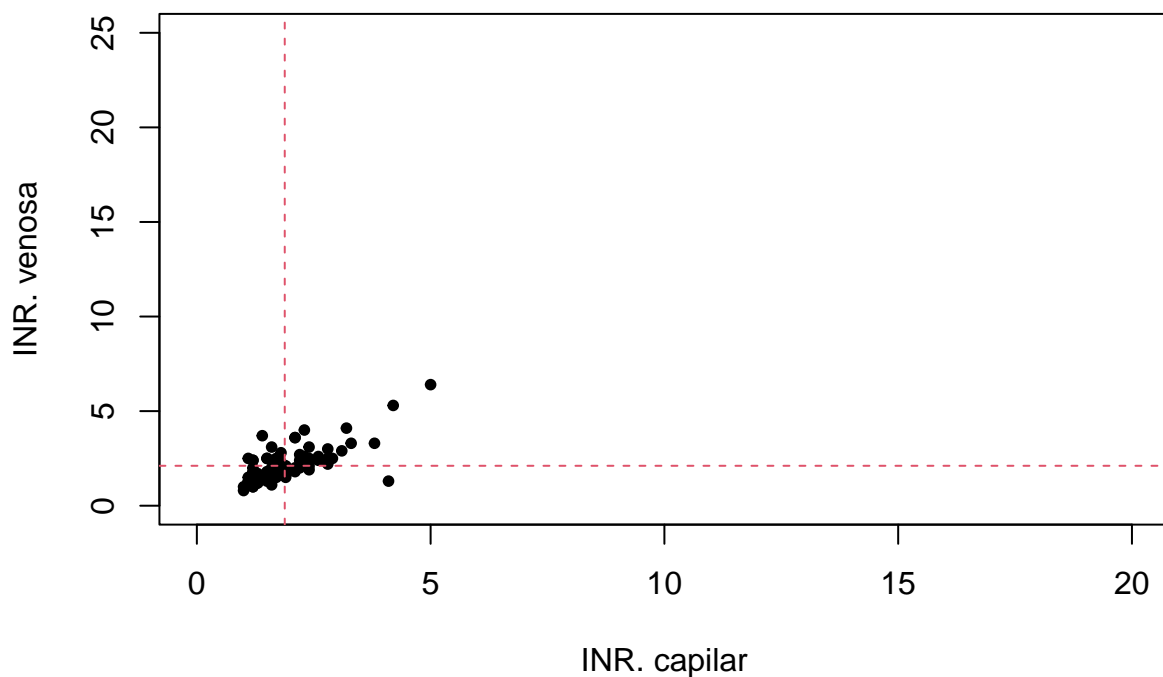
Trace un diagrama de dispersión y obtenga el modelo de regresión y el p . Trazarlo en el diagrama de dispersión.

```
coagu <- read_csv("EXR_C09_S03_04.csv", show_col_types = FALSE)
coagu
```

```
## # A tibble: 90 x 2
##       Y       X
##   <dbl> <dbl>
## 1  1.8   1.6
## 2  1.6   1.9
## 3  2.5   2.8
```

```
## 4 1.9 2.4
## 5 1.3 1.5
## 6 2.3 1.8
## 7 1.2 1.3
## 8 2.3 2.4
## 9 2 2.1
## 10 1.5 1.5
## # ... with 80 more rows
## # i Use 'print(n = ...)' to see more rows
```

```
# X capilar, Y venosa
plot(Y ~ X, data = coagu, pch = 20, xlab = "INR. capilar", ylab = "INR. venosa", xlim = c(0, 20), ylim = c(0, 25))
abline(v = mean(coagu$X), col = 2, lty = 2)
abline(h = mean(coagu$Y), col = 2, lty = 2)
```



```
# MinSqr
coagu_lm <- lm(Y ~ X, data = coagu)
summary(coagu_lm)
```

```
##
## Call:
## lm(formula = Y ~ X, data = coagu)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.7248 -0.3357 -0.1341  0.1306  2.0040
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.48848    0.18167   2.689  0.00858 **
## X             0.86251    0.08972   9.613 2.24e-15 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.64 on 88 degrees of freedom
## Multiple R-squared:  0.5122, Adjusted R-squared:  0.5067
## F-statistic: 92.41 on 1 and 88 DF,  p-value: 2.244e-15
```

2. Otra variable de interés en el estudio de Reiss et al. (ejercicio 1) fue la tromboplastina parcial (aPTT), la prueba estándar utilizada para controlar la anticoagulación con heparina. Utilice los datos del archivo EXR_C09_S07_02.csv para examinar la correlación entre los niveles de aPTT medidos por el ensayo de punto de atención CoaguCheck y el ensayo estándar del laboratorio del hospital en 90 sujetos que recibieron heparina sola, heparina con warfarina, y warfarina y exoenoxaparina.

- Prepara un diagrama de dispersión.
- Calcule el coeficiente de correlación de la muestra.
- Prueba $H_0: \rho=0$ al nivel de significación del 0,05 y exponga sus conclusiones.
- Determine el valor p de la prueba.
- Construya el intervalo de confianza del 95 por ciento para p.

```
niveles <- read_csv("EXR_C09_S07_02.csv", show_col_types = FALSE)
niveles
```

```
## # A tibble: 90 x 2
##   COAGU  HOSP
##   <dbl> <dbl>
## 1  49.3  71.4
## 2  57.9  86.4
## 3   59   75.6
## 4  77.3  54.5
## 5  42.3  57.7
## 6  44.3  59.5
## 7   90   77.2
## 8  55.4  63.3
## 9  20.3  27.6
## 10 28.7  52.6
## # ... with 80 more rows
## # i Use 'print(n = ...)' to see more rows
```

```
plot(COAGU ~ HOSP, data = niveles, pch = 20, xlab = "aPTT medidos", ylab = "niveles de coagulación")
```

```
cor.test(~ HOSP + COAGU, data = niveles, alternative = c("two.sided"), method = c("pearson"), conf.level = 0.95)
```

```
##
## Pearson's product-moment correlation
##
## data:  HOSP and COAGU
## t = 10.17, df = 88, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6227350 0.8176615
## sample estimates:
##      cor
## 0.735034
```

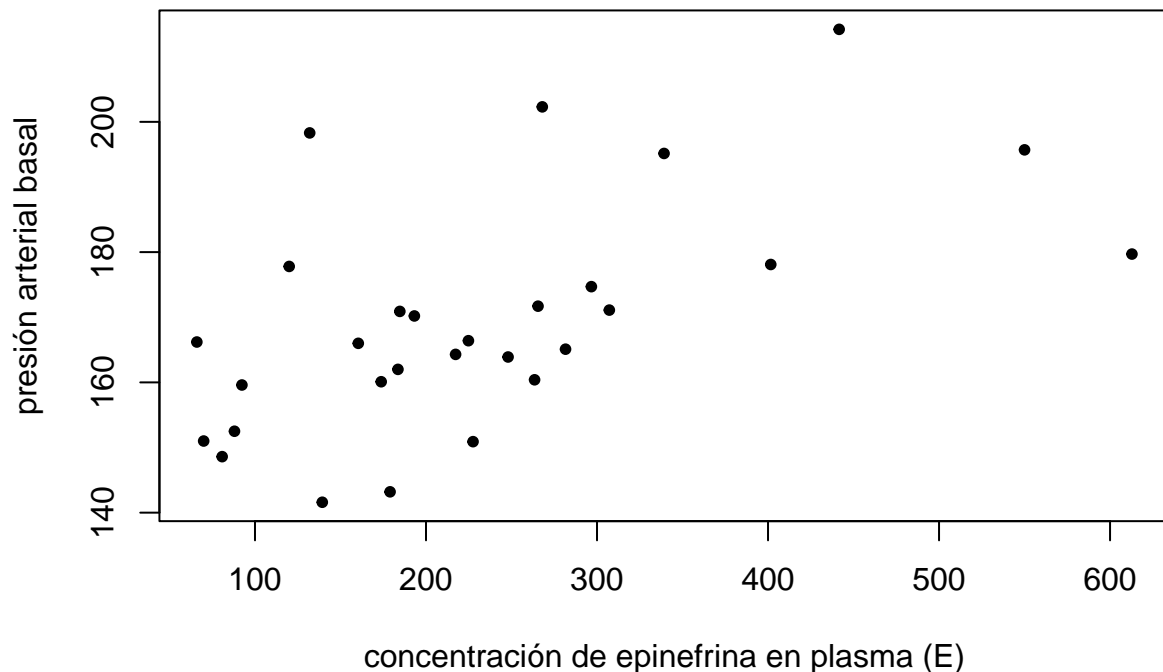
3. Un estudio de Scrogin et al. fue diseñado para evaluar los efectos de las manipulaciones simultáneas de NaCl y calcio en la dieta sobre la presión arterial, así como la presión arterial y las respuestas de las catecolaminas al estrés. Los sujetos eran ratas macho espontáneamente hipertensas y sensibles a la sal. Entre los análisis realizados por los investigadores figuraba una correlación entre la presión arterial basal y la concentración de epinefrina en plasma (E). Se recogieron los datos del archivo REV_C09_16.csv sobre estas dos variables. Sea $\alpha = .01$.

- Realice el análisis necesario y pruebe las hipótesis a los niveles de significación indicados. Calcule el valor p para el modelo lineal.
- Indique todos los supuestos necesarios para validar su análisis. Describa las poblaciones sobre las cuales cree que las inferencias basadas en su análisis serían aplicables.

```
diet <- read_csv("REV_C09_16.csv", show_col_types = FALSE)
diet
```

```
## # A tibble: 29 x 2
##       BP PLASMAE
##   <dbl> <dbl>
## 1  164.   248
## 2  195.   339.
## 3  170.   193.
## 4  171.   307.
## 5  149.    80.8
## 6  196.   550
## 7  151.    70
## 8  166.    66
## 9  178.   120
## 10 165.   282.
## # ... with 19 more rows
## # i Use 'print(n = ...)' to see more rows
```

```
plot(BP ~ PLASMAE, data = diet, pch = 20, xlab = "concentración de epinefrina en plasma (E)", ylab = "presión arterial basal")
```



```
cor.test(~ BP + PLASMAE, data = diet, alternative = c("two.sided"), method = c("pearson"), conf.level =

##
## Pearson's product-moment correlation
##
## data: BP and PLASMAE
## t = 3.6729, df = 27, p-value = 0.001044
## alternative hypothesis: true correlation is not equal to 0
## 99 percent confidence interval:
## 0.1519265 0.8221563
## sample estimates:
## cor
## 0.5772133
```

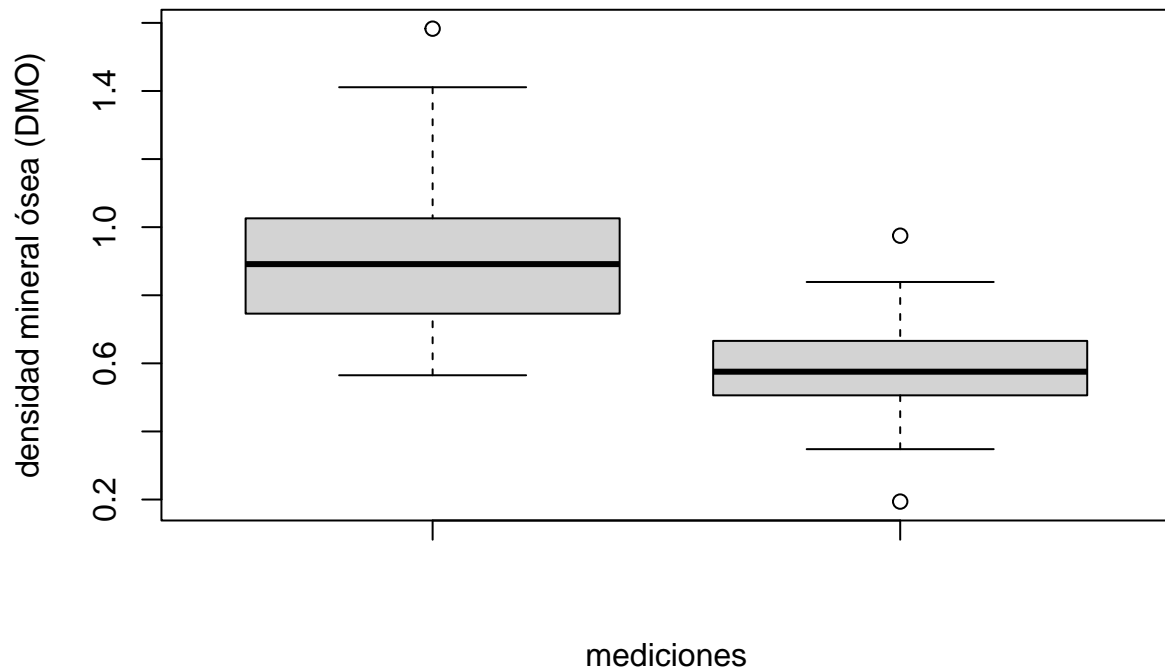
4. El objetivo de un estudio realizado por Peacock et al. era investigar si la osteoartritis espinal es responsable de que la densidad mineral ósea (DMO) de la columna lumbar sea mayor cuando se mide en el plano anteroposterior que cuando se mide en el plano lateral. Se estudiaron las radiografías laterales de la columna vertebral de mujeres (con un rango de edad de 34 a 87 años) que acudieron a un servicio ambulatorio de un hospital para la medición de la densidad ósea y se sometieron a una radiografía de la columna lumbar. Entre los datos recogidos (archivo REV_C09_41.csv) se encontraban las mediciones de la DMO anteroposterior (A) y lateral (L) (g/cm²).

- Construya los gráficos que considere útiles para ilustrar las relaciones entre las variables.
- Realice un análisis estadístico de los datos (que incluya la comprobación de hipótesis y la construcción de intervalos de confianza) que, en su opinión, aporte información útil para los investigadores.
- Utilice las técnicas aprendidas en otros capítulos, como las pruebas de hipótesis y la estimación de intervalos en relación con las medias y las proporciones.
- Determine los valores p para cada estadística de prueba calculada.

```
DMO <- read_csv("REV_C09_41.csv", show_col_types = FALSE)
DMO
```

```
## # A tibble: 66 x 2
##   ABMD LBMD
##   <dbl> <dbl>
## 1 0.879 0.577
## 2 0.824 0.622
## 3 0.974 0.643
## 4 0.909 0.664
## 5 0.872 0.559
## 6 0.93 0.663
## 7 0.912 0.71
## 8 0.758 0.592
## 9 1.07 0.702
## 10 0.847 0.655
## # ... with 56 more rows
## # i Use 'print(n = ...)' to see more rows
```

```
boxplot(DMO$ABMD, DMO$LBMD, xlab = "mediciones", ylab = "densidad mineral ósea (DMO)")
```



```
# Para hacer una prueba t usando notación de fórmula DMO ~ Condition
# Se necesitan alargar los datos
```

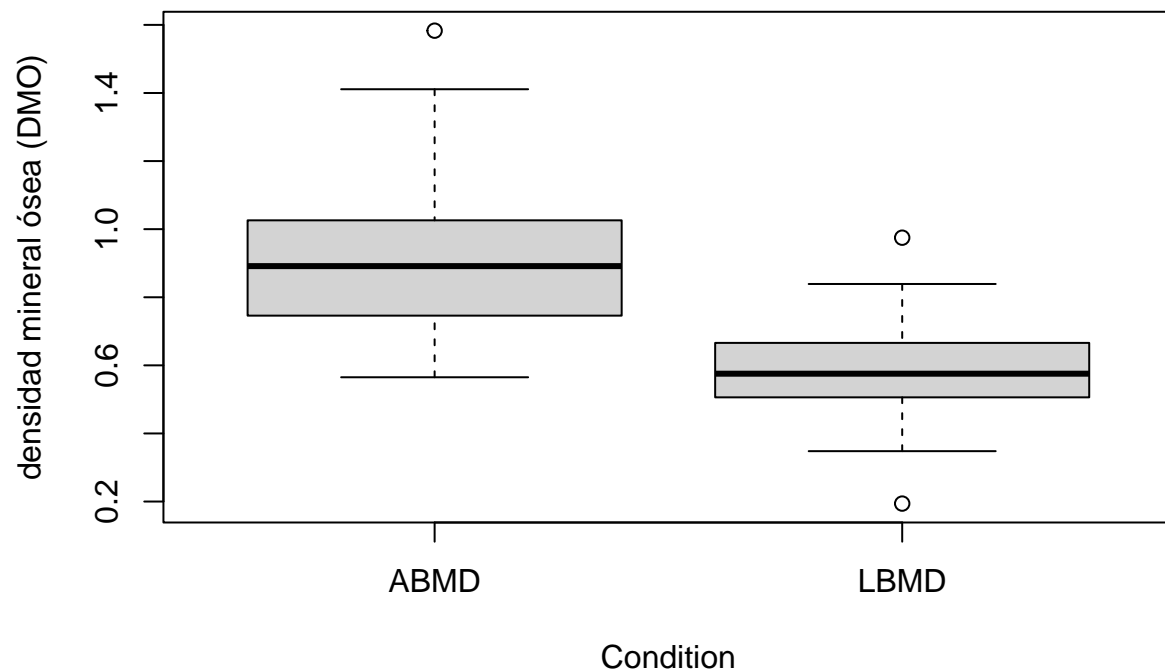
```
DMO_long <- DMO %>% pivot_longer(cols = everything(), names_to = "Condition", values_to = "DMO")
```

```
# y estar seguros que Condition es un factor en el orden adecuado
```

```
DMO_long <- DMO_long %>% mutate( Condition = Condition %>% fct_relevel("DMO anteroposterior (A)", "DMO lateral (L)"))
```

```
## Warning: Unknown levels in 'f': DMO anteroposterior (A), DMO lateral (L)
```

```
boxplot(DMO ~ Condition, data = DMO_long, ylab = "densidad mineral ósea (DMO)")
```



```
cor.test(~ ABMD + LBMD, data = DMO, alternative = c("two.sided"), method = c("pearson"), conf.level = 0
```

```
##
## Pearson's product-moment correlation
##
## data: ABMD and LBMD
## t = 8.6028, df = 64, p-value = 2.804e-12
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.5958788 0.8276398
## sample estimates:
## cor
## 0.7322979
```