Pauta Control de Laboratorio N°1 - Sección 2

(1) Considere los siguientes vectores con x como input e y como output

```
x <- c(0.61, 0.93, 0.83, 0.35, 0.54, 0.16, 0.91, 0.62, 0.62)
y <- c(0.67, 0.84, 0.6, 0.18, 0.85, 0.47, 1.1, 0.65, 0.36)
```

Desarrollo.

```
> x <- c(0.61, 0.93, 0.83, 0.35, 0.54, 0.16, 0.91, 0.62, 0.62)
> y <- c(0.67, 0.84, 0.6, 0.18, 0.85, 0.47, 1.1, 0.65, 0.36)
> summary(lm(y ~ x -1))

Call:
lm(formula = y ~ x - 1)

Residuals:
     Min        1Q   Median        3Q       Max
-0.25215 -0.16557  0.03785  0.20153  0.31684

Coefficients:
  Estimate Std. Error t value Pr(>|t|)
x   0.9873     0.1109   8.905    2e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2207 on 8 degrees of freedom
Multiple R-squared:  0.9084,    Adjusted R-squared:  0.8969
F-statistic:  79.3 on 1 and 8 DF,  p-value: 2.003e-05
```

(2) ¿Cuál es la diferencia entre R^2 y $\overline{R}^2$ en el problema anterior?

Desarrollo.

```
> 0.9084 - 0.8969
[1] 0.0115
```

(3) Considere el siguiente vector de datos y el vector de ponderadores

```
x <- c(0.26, -1.54, 0.42, 0.95)
w <- c(2, 1, 5, 3)
```

¿Cuál es el valor de μ que minimiza la ecuación de mínimos cuadrados $\sum_{i=1}^{n} w_i(x_i - \mu)^2$?

Desarrollo.

```
> x <- c(0.26, -1.54, 0.42, 0.95)
> w <- c(2, 1, 5, 3)
> weighted.mean(x,w)
[1] 0.3572727
```

(4) Considere los siguientes datos

```
x <- c(0.8, 0.47, 0.51, 0.73, 0.36, 0.58, 0.57, 0.85, 0.44)
y <- c(1.39, 0.72, 1.55, 0.48, -1.59, 1.23, -0.65, 1.49, 0.05)
```

Estime una regresión que pase por el origen considerando x como input e y como output. (Hint: no centre los datos ya que se pide una regresión que pase por el origen y no por las medias de los datos.) ¿Cuál es el valor de la pendiente?

Desarrollo.

```
> x <- c(0.8, 0.47, 0.51, 0.73, 0.36, 0.58, 0.57, 0.85, 0.44)
> y <- c(1.39, 0.72, 1.55, 0.48, -1.59, 1.23, -0.65, 1.49, 0.05)
> lm(y ~ x -1)

Call:
lm(formula = y ~ x - 1)

Coefficients:
    x
1.086
```

(5) Cargue la base da datos `mtcars` y estime una regresión que explique la variable `mpg` en función de la variable `hp`. ¿Cuál es el coeficiente del intercepto?

Desarrollo.

```
> data("mtcars")
> lm(mpg ~ hp, mtcars)

Call:
lm(formula = mpg ~ hp, data = mtcars)

Coefficients:
(Intercept)           hp
   30.09886     -0.06823
```

(6) Considere el vector

```
x <- c(8.58, 10.46, 9.01, 9.64, 8.86)
```

Si se normalizan los datos ¿Cuál es el valor de la primera observación?

Desarrollo.

```
> x <- c(8.58, 10.46, 9.01, 9.64, 8.86)
> (x - mean(x))/sd(x)
[1] -0.9718658  1.5310215 -0.3993969  0.4393366 -0.5990954
```

(7) Sea β_1 la pendiente de un modelo en el que Y es el output y X es el input. Sea γ_1 la pendiente de un modelo en el que X es el output e Y es el input. ¿Cuál es el valor de β_1/γ_1?

Desarrollo.

$$\frac{\beta}{\gamma} = \frac{\operatorname{cor}(X,Y)\cdot\dfrac{sd(Y)}{sd(X)}}{\operatorname{cor}(X,Y)\cdot\dfrac{\operatorname{sd}(X)}{\operatorname{sd}(Y)}} = \frac{\operatorname{sd}^2(Y)}{\operatorname{sd}^2(X)} = \frac{\operatorname{var}(Y)}{\operatorname{var}(X)}$$

(8) Considere una regresión hipotética en la cual Y es el output y X es el input. La desviación típica del input es igual a la del output. La correlación entre ambas variables es $0,25$. ¿Cuál es el valor de la pendiente de un modelo en el que se invierte el input y el output?

Desarrollo. De acuerdo a lo visto en clase

$$\beta = \operatorname{cor}(X,Y)\cdot\frac{\operatorname{sd}(Y)}{\operatorname{sd}(X)} = 0{,}25$$

Si se invierte el output y el input el coeficiente de la pendiente es

$$\gamma = \operatorname{cor}(X,Y)\cdot\frac{\operatorname{sd}(X)}{\operatorname{sd}(Y)} = 0{,}25$$

(9) Considere un modelo lineal con datos hipotéticos en el que el input y el output tienen medias iguales a μ_1 y μ_2 respectivamente. ¿Qué se puede afirmar respecto del intercepto si se estima una regresión con constante?

Desarrollo. En el modelo con constante la regresión pasa por $(\overline{X},\overline{Y})$. Entonces el intercepto es el punto (μ_1,μ_2).

(10) Considere el vector

```
x <- c(0.8, 0.47, 0.51, 0.73, 0.36, 0.58, 0.57, 0.85, 0.44, 0.42)
```

¿Cuál es el valor que minimiza la suma de las distancias el cuadrado entre estos puntos y el valor mismo?

Desarrollo. Como se vio en clases el vector pedido es el que sus coordenadas son iguales a la media de los datos que se calcula de la siguiente manera:

```
> x <- c(0.8, 0.47, 0.51, 0.73, 0.36, 0.58, 0.57, 0.85, 0.44, 0.42)
> mean(x)
[1] 0.573
```