

Pauta del Primer Control de Laboratorio

- (1) Considere el siguiente vector de datos y el vector de ponderadores

```
x <- c(0.18, -1.54, 0.42, 0.95)
w <- c(2, 1, 3, 1)
```

¿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(0.18, -1.54, 0.42, 0.95)
> w <- c(2, 1, 3, 1)
> weighted.mean(x, w)
[1] 0.1471429
```

- (2) 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
```

- (3) Cargue la base de datos mtcars y estime una regresión que explique la variable mpg en función de la variable wt. ¿Cuál es el coeficiente de la pendiente?

Desarrollo.

```
> data(mtcars)
> lm(mpg ~ wt, mtcars)

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

Coefficients:
(Intercept)          wt
   37.285         -5.344
```

- (4) Considere una regresión hipotética en la cual Y es el output y X es el input. La desviación típica del input es la mitad de la del output. La correlación entre ambas variables es 0,5. ¿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

$$\beta = \text{cor}(X, Y) \cdot \frac{\text{sd}(Y)}{\text{sd}(X)} = 1$$

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

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

- (5) 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)
> mean <- mean(x)
> sd <- sd(x)
> (x-mean)/sd
[1] -0.9718658  1.5310215 -0.3993969  0.4393366 -0.5990954
```

- (6) 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.

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

- (7) Considere un modelo lineal con datos hipotéticos en el que el input y el output tienen media 0. ¿Qué se puede afirmar respecto del intercepto si se estima una regresión?

Desarrollo. En el modelo sin constante la regresión pasa por $(0, 0)$. En el modelo con constante la regresión pasa por (\bar{X}, \bar{Y}) . Luego el único caso posible a partir de los datos es que el intercepto sea cero en todos los casos.

- (8) 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 al 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
```

- (9) 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)
```

¿Cuál es el valor p de test de dos colas para determinar si acaso el β_1 de la regresión lineal es 0 o no?

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)
> fit <- lm(y ~ x)
> summary(fit)
```

```
Call:
lm(formula = y ~ x)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.27636 -0.18807  0.01364  0.16595  0.27143
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.1885     0.2061   0.914   0.391
x              0.7224     0.3107   2.325   0.053 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.223 on 7 degrees of freedom
Multiple R-squared:  0.4358,    Adjusted R-squared:  0.3552
F-statistic: 5.408 on 1 and 7 DF,  p-value: 0.05296
```

- (10) ¿Cuál es la desviación típica residual en el problema anterior?

Desarrollo. Idéntico al desarrollo anterior.

Tiempo: 20 minutos