Quiz1

jbellier

04 décembre 2016

Question 1

Consider the dataset given below :

```
x <- c(0.18, -1.54, 0.42, 0.95)
```

```
w <- c(2, 1, 3, 1)
```

Give the value of μ tha minimizes the least squares equation :

$$\Sigma_{i=1}^n w_i * (x_i - \mu)^2$$

- 0.300
- 0.0025
- 1.077
- 0.1471

```
mu_hat <- sum(w*x)/sum(w)
mu_hat
```

```
## [1] 0.1471429
```

Answer : The p-value that we are looking for is then : ***0.1471***.

Question 2

Consider the following dataset :

```
x <- c(0.8, 0.47, 0.51, 0.73,0.36,0.58,0.57,0.85,0.44,0.42)
y <- c(1.39, 0.72, 1.55, 0.48, 1.19,-1.59,1.23,-0.65,1.49,0.05)
```

Fit the regression to the origin and get the slope of treating Y as the outcome and x as the regressor. (Hint, do not center the data since we want regression through the origin and not through the means of data).

- 0.8263
- -0.04462
- 0.59915
- -1.713

```
fit2 <- lm(y~x)
fit2$coefficients
```

```
## (Intercept)           x
##    1.567461    -1.712846
```

Answer : The expected slope is the coefficient on x, i.e. ***-1.713***.

Question 3

Do data(mtcars) from the dataset package and fit the regression model with mpg as the outcome and weight as the predictor. Give the slope coefficient.

- -9.559
- -5.344
- 0.5591
- 30.2851

```
fit3 <- lm(mpg~wt,mtcars)
slope <- fit3$coefficients[2]
slope
```

```
##        wt
## -5.344472
```

Answer : The expected slope is the coefficient on x, i.e. ***-5.344***.

Question 4

Consider data with an outcome Y and a predictor X. The standard deviation of the predictor is one half that of the outcome. The correlation between the two variables is 0.5. What value would the slope coefficient for the regression model with Y as the outcome and X as the regressor ?

- 3
- 0.25
- 4
- 1

Answer : We know that the slope is defined as follows :

$$\beta = Corr(X, Y) * \frac{sd(Y)}{sd(X)}$$

So the slope coefficient is :

$$0.5 \frac{sd(Y)}{0.5 sd(Y)}$$

, i.e. ***1***.

Question 5

Students were given two hard tests and scores were normalized to have empirical mean 0 and variance 1. The correlation between the scores on the two tests was 0.4. What would be the expected score on quiz 2 for a student who had a normalized score of 1.5 on quiz 1 ?

- 1
- 0.4
- 0.16
- 0.6

Answer : From the previous question, as th scores are normalized, we just need to multiply the correlation with the score of quiz1 (regressor) to obtain the result on quiz2 (outcome). The result is then 0.4.1.5 = ***0.6***.

Question 6

Consider the data given by the following :

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

What would be the value of the first measurement if x were normalized (to have mean 0 and variance 1) ?

- 8.58
- -0.9719
- 8.86
- 9.31

Answer : Let us normalize x as follows :

```
y <- (x-mean(x))/sd(x)
y[1]
```

```
## [1] -0.9718658
```

The value of the first measurement of the normalized vector is then y_1, i.e. ***-0.9719***.

Question 7

Consider the following dataset (used above as well).

```
x <- c(0.8, 0.47, 0.51, 0.73,0.36,0.58,0.57,0.85,0.44,0.42)
y <- c(1.39, 0.72, 1.55, 0.48, 1.19,-1.59,1.23,-0.65,1.49,0.05)
```

What is the intercept for fitting the model with x as the predictor and y as the outcome ?

- 1.567
- -1.713
- 1.252
- 2.105

Answer : The intercept ios the first coefficient returned by the model :

```
fit7 <- lm(y~x)
fit7$coefficients[1]
```

```
## (Intercept)
##    1.567461
```

The value is then ***1.567***.

Question 8

You know that both the predictor and the response have mean 0. What can be said about the intercept when you fit a linear regression ?

- it is undefined as you have to divide by zero
- it must be exactly one
- it must be identically zero
- nothing about the intercept can be said from the information given

Answer : Let us call X the predictor and Y the outcome. The intercept is defined by

$$\hat{\beta}_0 = \bar{Y} - \beta_1.\bar{X}$$

As $\bar{X}$ and $\bar{Y}$ are both equal to zero, ***the intercept is also 0***.

Question 9

Consider the data given by :

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

What value minimizes the sum of the squared distances between these points and itself ?

- 0.8
- 0.44
- 0.573
- 0.36

Answer : We want to know the value of $\bar{\mu}$ that minimizes the equation :

$$\Sigma_i (X_i - \mu)^2$$

which is known as $\bar{X}$.

```
mean(x)
```

```
## [1] 0.573
```

Question 10

Let the slope having fit Y as the outcome and X as the predictor be denoted as β_1. Let the slope from fitting X as the outcome and Y as the predictor be denoted as γ_1. Suppose that you divide β_1 by γ_1; in other words consider β_1/γ_1. What is this ratio always equal to ?

- Var(Y)/Var(X)
- 2Sd(Y)/Sd(X)
- Cor(Y,X)
- 1

Answer : By construction we have :

$$\beta_1 = Cor(Y, X).\frac{Sd(Y)}{Sd(X)}$$

and

$$\gamma_1 = Cor(X, Y).\frac{Sd(X)}{Sd(Y)}$$

So

$$\frac{\beta_1}{\gamma_1} = \frac{Cor(Y, X)}{Cor(X, Y)}.\frac{\frac{Sd(Y)}{Sd(X)}}{\frac{Sd(X)}{Sd(Y)}} = \frac{Sd(Y)^2}{Sd(X)^2} = \frac{Var(Y)}{Var(X)}$$