1.

Consider the data set given below

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

And weights given by

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

Give the value of *μ* that minimizes the least squares equation

∑*ni*=1*wi*(*xi*−*μ*)2



0.300



1.077



0.0025



0.1471

2.

Consider the following data set

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 through the origin and get the slope treating y

as the outcome and x as the regressor. (Hint, do not center the data since we want regression through the origin, not through the means of the data.)



0.59915



-1.713



-0.04462



0.8263

Sol:

lm(formula = y ~ x - 1)

Coefficients:

x

0.8263

3.

Do data(mtcars) from the datasets package and fit the regression

model with mpg as the outcome and weight as the predictor. Give

the slope coefficient.



-9.559



30.2851



-5.344



0.5591

Sol:

> lmmt <- lm(mpg ~ wt, data = mtcars)

> coef(lmmt)

(Intercept) wt

37.285126 -5.344472

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 .5. What value would the slope coefficient for the regression model with *Y* as the outcome and *X* as the predictor?



1



3



0.25



4

Sol :

Sd(x) = 0.5 sd(y)

Slope = cor(y,x) \* (sd(y)/sd(x))

Slope = 0.5 \*( sd(y)/0.5 sd(y)) =1

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?



0.6



0.16



1.0



0.4

Sol: 1.5 \* .4= 0.6

6.

Consider the data given by the following

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

What is 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

Sol :

(x-mean(x))/sd(x)

[1] -0.9718658 1.5310215 -0.3993969 0.4393366 -0.5990954

7.

Consider the following data set (used above as well). What is the intercept for fitting the model with x as the predictor and y as the outcome?

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)



1.567



1.252



-1.713



2.105

Sol:

coef(lm(y ~ x))

(Intercept) x

1.567461 -1.712846

8.

You know that both the predictor and response have mean 0. What

can be said about the intercept when you fit a linear regression?



It must be exactly one.



It must be identically 0.



Nothing about the intercept can be said from the information given.



It is undefined as you have to divide by zero.

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?



0.36



0.44



0.8



0.573

> sum((x-.36)^2)

[1] 0.7077

> sum((x-.44)^2)

[1] 0.4309

> sum((x-.8)^2)

[1] 0.7693

> sum((x-.573)^2)

[1] 0.25401

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?



1



*Cor*(*Y*,*X*)



*Var*(*Y*)/*Var*(*X*)



2*SD*(*Y*)/*SD*(*X*)