

# Coursera Regression Models Quiz 1

*Cheng-Han Yu*

*August 11, 2015*

## Question 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  $\mu$  that minimizes the least squares equation  $\sum_{i=1}^n w_i(x_i - \mu)^2$ .

**Solution:**

The minimizer is  $\frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$ .

```
x <- c(0.18, -1.54, 0.42, 0.95)
w <- c(2, 1, 3, 1)
sum(w * x) / sum(w)
## [1] 0.1471429
```

## Question 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.)

**Solution:**

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

```
##          x
## 0.8263
```

### Question 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.

**Solution:**

```
data(mtcars)
lm(mtcars$mpg ~ mtcars$wt)
##
## Call:
## lm(formula = mtcars$mpg ~ mtcars$wt)
##
## Coefficients:
## (Intercept)      mtcars$wt
##      37.285         -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 .5. What value would the slope coefficient for the regression model with Y as the outcome and X as the predictor?

**Solution:**

$\hat{\beta}_1 = \text{Corr}(Y, X) \frac{\text{sd}(Y)}{\text{sd}(X)} = 0.5 \frac{\text{sd}(Y)}{0.5 \text{sd}(Y)} = 1.$

### 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?

**Solution:**

$q_1 = q_2 \hat{\beta}$ .  $\hat{\beta} = \text{Corr}(q_1, q_2) \frac{\text{sd}(q_1)}{\text{sd}(q_2)} = 0.4$ . And so  $q_1 = (0.4)(1.5) = 0.6$ .

### Question 6

Consider the data given by the following

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

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

**Solution:**

```
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
```

## Question 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)
```

**Solution:**

```
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)
lm(y ~ x)
##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)          x
##      1.567      -1.713
```

## Question 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?

**Solution:**

It must be identically 0.

```
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)
y <- (y - mean(y)); x <- (x - mean(x))
lm(y ~ x)
##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)          x
##  1.003e-16  -1.713e+00
```

## 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?

```
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
```

## Question 10

Let the slope having fit Y as the outcome and X as the predictor be denoted as  $\beta_1$ . Let the slope from fitting X as the outcome and Y as the predictor be denoted as  $\gamma_1$ . Suppose that you divide  $\beta_1$  by  $\gamma_1$ ; in other words consider  $\beta_1/\gamma_1$ . What is this ratio always equal to?

**Solution:**

$\hat{\beta}_1 = \text{Corr}(Y, X) \text{sd}(Y) / \text{sd}(X)$  and  $\hat{\gamma}_1 = \text{Corr}(X, Y) \text{sd}(X) / \text{sd}(Y)$ , and hence  
 $\hat{\beta}_1 / \hat{\gamma}_1 = \text{sd}(Y)^2 / \text{sd}(X)^2 = \text{Var}(Y) / \text{Var}(X)$ .