Data Science - Regression Models - Quiz 1 - Coursera

Quiz 1

This is Quiz 1 from Coursera's Regression Models class within the Data Science Specialization. This publication is intended as a learning resource, all answers are documented and explained. Datasets are available in R packages.

---

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

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

- **0.1471**

Explanation:

Least squares is minimized by the mean

```
mean(w*x)/mean(w)
```

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

---

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.)

- **0.8263**

Explanation:

Subtract 1 from the regressor in ln to fit line through origin.

```
summary(lm(y~x-1))
```

```
##
## Call:
## lm(formula = y ~ x - 1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0692 -0.2536  0.5303  0.8592  1.1286
##
## Coefficients:
##   Estimate Std. Error t value Pr(>|t|)
## x   0.8263     0.5817   1.421    0.189
##
## Residual standard error: 1.094 on 9 degrees of freedom
## Multiple R-squared:  0.1831, Adjusted R-squared:  0.09238
## F-statistic: 2.018 on 1 and 9 DF,  p-value: 0.1892
```

---

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.

- **-5.344**

Explanation:

Predicting with the lower and upper bounds of the confidence intervals

```
fit <- lm(mpg~wt,mtcars)
summary(fit)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 37.285126   1.877627 19.857575 8.241799e-19
```

```
## wt          -5.344472   0.559101 -9.559044 1.293959e-10
```

---

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**

Explanation:

Correlation(XY)* SDy/SDx

```
.5*(1/.5)
```

```
## [1] 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**

Explanation:

Since the distributions are normalized, the slope coefficient is equal to the correlation.

```
1.5*.4
```

```
## [1] 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)?

- **-0.9719**

Explanation:

To normalize, we subtract the mean and divide by the standard deviation.

```
(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**

Explanation:

The slope coefficient represents the change in the outcome per unit change in regressor. (outcome/regressor) So if you divide the regressor (m -> cm) you are effectively multiplying the outcome by shrinking the units. If you multiply the regressor it will have the opposite effect. The actual change is not effected, only how it is expressed relative to the units of the regressor.

```
summary(lm(y~x))
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1640 -0.5818  0.2010  0.6669  1.1928
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.567      1.252   1.252    0.246
## x             -1.713      2.105  -0.814    0.439
##
```

```
## Residual standard error: 1.061 on 8 degrees of freedom
## Multiple R-squared:  0.07642,    Adjusted R-squared:  -0.03903
## F-statistic: 0.662 on 1 and 8 DF,  p-value: 0.4394
```

---

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 identically 0.**

Explanation:

---

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)
```

- **0.573**

Explanation:

For LS, the mean minimizes

```
mean(x)
## [1] 0.573
```

---

10. Let the slope having fit Y as the outcome and X as the predictor be denoted as β1. Let the slope from fitting X as the outcome and Y as the predictor be denoted as γ1. Suppose that you divide β1 by γ1; in other words consider β1/γ1. What is this ratio always equal to?

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

Explanation:

Beta = cor(X,Y) * SDy/SDx….Since Cor(X,Y)=Cor(Y,X) they will cancel out leaving the standard deviations squared (variance.)