

Data Science - Regression Models - Quiz 2 - Coursera

Quiz 2

This is Quiz 2 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 following data with x as the predictor and y as as the outcome.

```
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)
```

Give a P-value for the two sided hypothesis test of whether β_1 from a linear regression model is 0 or not.

• **0.05296**

Explanation:

P-value on beta1 coefficient given by summary of the linear model

```
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)

summary(lm(y~x))$coef[8]

## [1] 0.05296439
```

2. Consider the previous problem, give the estimate of the residual standard deviation.

• **0.223**

Explanation:

Residual standard deviation is given by the square root of the sum of the squared residuals of degrees of freedom.

```
sum(resid(lm(y~x))^2)
## [1] 0.348097

sqrt(sum(resid(lm(y~x))^2))
## [1] 0.5899974

sqrt(sum(resid(lm(y~x))^2)/7)
## [1] 0.2229981
```

3. In the `mtcars` data set, fit a linear regression model of weight (predictor) on mpg (outcome). Get a 95% confidence interval for the expected mpg at the average weight. What is the lower endpoint?

-
- **18.991**

Explanation:

Predicting with the lower and upper bounds of the confidence intervals

```
dat <- mean(mtcars$wt)
fit <- lm(mpg~wt,mtcars)

predict(fit, data.frame(wt = dat), interval = "confidence")
##           fit          lwr          upr
## 1 20.09062 18.99098 21.19027
```

4. Refer to the previous question. Read the help file for `mtcars`. What is the weight coefficient interpreted as?

-
- **The estimated expected change in mpg per 1,000 lb increase in weight.**

Explanation:

Mtcars reports the weight in units of 1000 lbs. `[, 6] wt` Weight (1000 lbs)

5. Consider again the `mtcars` data set and a linear regression model with `mpg` as predicted by weight (1,000 lbs). A new car is coming weighing 3000 pounds. Construct a 95% prediction interval for its `mpg`. What is the upper endpoint?

- **27.57**

Explanation:

Using same fit, changing predictor to 3 (in 1000lbs units).

```
predict(fit, data.frame(wt = 3.0), interval = "prediction")
##           fit          lwr          upr
## 1 21.25171 14.92987 27.57355
```

6. Consider again the `mtcars` data set and a linear regression model with `mpg` as predicted by weight (in 1,000 lbs). A “short” ton is defined as 2,000 lbs. Construct a 95% confidence interval for the expected change in `mpg` per 1 short ton increase in weight. Give the lower endpoint.

- **-12.973**

Explanation:

Multiplying the estimated change per 1000lbs (beta coef) by 2 and adding the error term*2.

```
summary(fit)$coef[2]*2+2*c(-1,1)*qt(.975,30)*summary(fit)$coef[4]
## [1] -12.97262 -8.40527
```

7. If my X from a linear regression is measured in centimeters and I convert it to meters what would happen to the slope coefficient?

- **It would get multiplied by 100.**

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.

8. I have an outcome, Y, and a predictor, X and fit a linear regression model with $Y = \beta_0 + \beta_1 X + \epsilon$ to obtain $\hat{\beta}_0$ and $\hat{\beta}_1$. What would be the consequence to the subsequent slope and intercept if I were to refit the model with a new regressor, $X+c$ for some constant, c?

- **The new intercept would be $\hat{\beta}_0 - c\hat{\beta}_1$**
-

Explanation:

This is a consequence of the least squares criteria.

9. Refer back to the mtcars data set with mpg as an outcome and weight (wt) as the predictor. About what is the ratio of the the sum of the squared errors, $\sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ when comparing a model with just an intercept (denominator) to the model with the intercept and slope (numerator)?

- **0.25**
-

Explanation:

Fitting a model with just an intercept will always predict at the mean.

```
fit2 <- lm(mpg ~ 1,mtcars)
fit <- lm(mpg~wt,mtcars)

sum((predict(fit)-mtcars$mpg)^2)/sum((predict(fit2)-mtcars$mpg)^2)

## [1] 0.2471672
```

10. Do the residuals always have to sum to 0 in linear regression?

- **If an intercept is included, then they will sum to 0.**
-

Explanation:

Least squares effectively minimizes the sum of the squared residuals. By not including an intercept, the mean of x times beta MUST equal the mean of y . This effectively weights the coefficients so that the line passes through zero.
