Regression Models: Quiz 2

Marcos Gestal

Question 1

Consider the following data with x as the predictor and y as as the outcome.

```r
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.

```r
fit <- lm(y ~ x)
summary(fit)
```

Solution

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.27636 -0.18807  0.01364  0.16595  0.27143
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.1885     0.2061   0.914    0.391
## x             0.7224     0.3107   2.325    0.053 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.223 on 7 degrees of freedom
## Multiple R-squared:  0.4358, Adjusted R-squared:  0.3552
## F-statistic: 5.408 on 1 and 7 DF,  p-value: 0.05296
```

```r
# Answer: 0.05296
```

Question 2

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

```r
# Answer: 0.223
```

Solution

Question 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?

```r
data(mtcars)

fit <- lm(mpg ~ wt, data=mtcars)
summary(fit)
```

Solution

```
##
## Call:
## lm(formula = mpg ~ wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5432 -2.3647 -0.1252  1.4096  6.8727
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.2851     1.8776  19.858  < 2e-16 ***
## wt           -5.3445     0.5591  -9.559 1.29e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.046 on 30 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7446
## F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

```r
avgWeight <- data.frame(wt = mean(mtcars$wt))
predict(fit, avgWeight, interval='confidence')
```

```
##        fit      lwr      upr
## 1 20.09062 18.99098 21.19027
```

```r
# Answer: 18.99098
```

Question 4

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

Solution

- It can't be interpreted without further information

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

Question 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?

```r
data(mtcars)
fit <- lm(mpg ~ wt, data=mtcars)
summary(fit)
```

Solution

```
##
## Call:
## lm(formula = mpg ~ wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5432 -2.3647 -0.1252  1.4096  6.8727
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.2851     1.8776  19.858  < 2e-16 ***
## wt           -5.3445     0.5591  -9.559 1.29e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.046 on 30 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7446
## F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

```r
newCar_weight <- 3000/1000
predict(fit, data.frame(wt = newCar_weight), interval='prediction')
```

```
##        fit      lwr      upr
## 1 21.25171 14.92987 27.57355
```

```r
# Answer: 27.57355
```

Question 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.

```r
data(mtcars)
shortTon <- mtcars$wt / 2
fit <- lm(mtcars$mpg ~ shortTon)
summary(fit)
```

Solution

```
##
## Call:
## lm(formula = mtcars$mpg ~ shortTon)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5432 -2.3647 -0.1252  1.4096  6.8727
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   37.285      1.878  19.858  < 2e-16 ***
## shortTon     -10.689      1.118  -9.559 1.29e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.046 on 30 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7446
## F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

```r
confint(fit)
```

```
##                 2.5 %   97.5 %
## (Intercept)  33.45050 41.11975
## shortTon    -12.97262 -8.40527
```

```r
# Answer: -12.97262
```

Question 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?

Solution

- It would get divided by 100
- It would get multiplied by 10
- It would get divided by 10
- It would get multiplied by 100

Question 8

I have an outcome, Y, and a predictor, X and fit a linear regression model with $Y = \beta_0 + \beta_1 \mathring{u} X + \epsilon$ 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?

Solution

- The new slope would be c$\hat{\beta}_1$

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

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

Question 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)?

```r
data(mtcars)

fitInterceptor <- lm(mpg ~ wt + 1, data=mtcars)
fitWithoutInterceptor <- lm(mpg ~1, data=mtcars)

ratio_of_residuals <- sum(resid(fitInterceptor)^2) / sum(resid(fitWithoutInterceptor)^2)
ratio_of_residuals
```

Solution

```
## [1] 0.2471672
```

Question 10

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

Solution

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

- The residuals must always sum to zero.

- If an intercept is included, the residuals most likely won't sum to zero.

- The residuals never sum to zero.