STT 4660
Homework #3

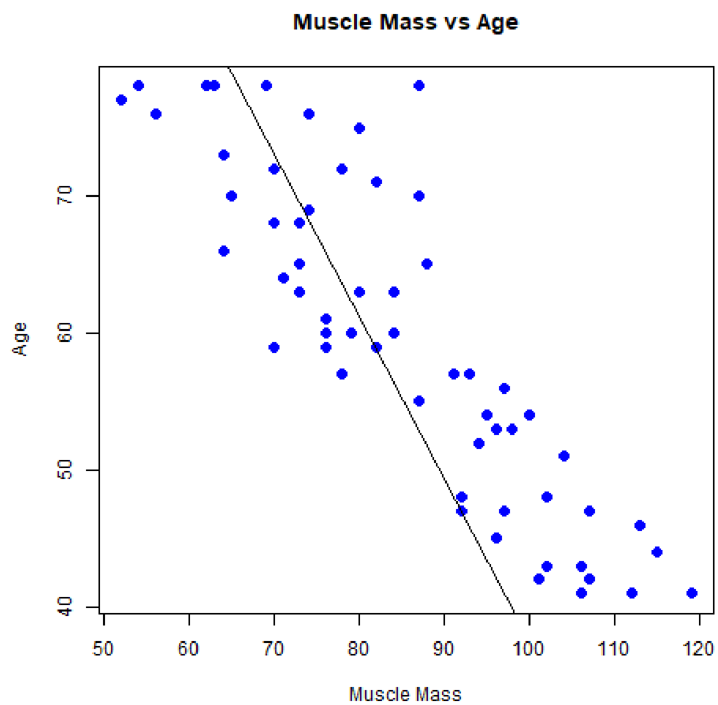
Timothy Stubblefield

September 23, 2020

1.27

a) We need to find the regression function for the Muscle Mass dataset.

   1) By using SAS, I obtained the regression function for Muscle Mass, which is
   $$\hat{y} = 156.34656 - 1.19000x$$
   2) The linear regression function seems to give a decent fit. The $R^2 = 0.7501$ which means only 75 percent of the points are explained by the model. In general, we would want the R^2 to be a little higher, but it isn't noticeably low either.
   3) Overall, the model supports the anticipation that muscle mass decreases with age since it shows that on average, for each year a person increases in age, their muscle mass measure decreases by about 1.21.
   4) Below is the plot for the regression line from R,



5) Upon viewing the regression line for the data, we can visually see that the R^2 value is a bit low and it seems that perhaps another model could fit the data better.

b) Now we want to solve the following problems.

1) Well, we want a point estimate for muscle mass of women differing by one year in age. From the regression model, we calculated that $b_1 = -1.19000$ which can be used to estimate β_1. Thus, if we were to find a point estimate of two women differing in age by one year, that could be represented as such,

$$W_1 : y_x = 156.34656 - 1.19000x$$

$$W_2 : y_{x+1} = 156.34656 - 1.19000(x + 1)$$

Now, compute $W_2 - W_1$, yielding,

$$W_2 - W_1 = 156 - 1.19000x - 1.19000 - (156.34656 - 1.19000x)$$

$$W_2 - W_1 = -1.19000$$

Thus, on average, as a woman's age increases by 1 year, her muscle mass decreases by 1.19000 units.

2) Now we want a point estimate for X = 60. Thus, we use the regression model, that is,

$$\hat{y} = 156.34656 - 1.19000x$$

So,

$$\hat{y}_{60} = 156.34656 - 1.19000(60)$$

$$\hat{y}_{60} = 156.34656 - 71.4$$

$$\hat{y}_{60} = 84.94656$$

3) Now, we want to find the residual for the eighth case. For the eighth case, we have the following data,

$$X_8 = 41 \qquad\qquad Y_8 = 112$$

So, $Y_8 = 112$ occurs when $x = 41$, thus that means,

$$y_8 = y(41) = 112$$

Now, use the regression model to find $\hat{y}_8$, as such,

$$\hat{y}_8 = 156.34656 - 1.19000(41)$$

$$\hat{y}_8 = 156.34656 - 44.69$$

$$\hat{y}_8 = 111.65656$$

Hence, the eighth residual will be $e_8 = y_8 - \hat{y}_8$. Evaluate e_8 as such,

$$e_8 = 112 - 111.65656$$

$$e_8 = 0.34344$$

4) Finally, let's find a point estimate for σ^2. To find a point estimate for σ^2, we will use MSE. The formula for MSE is,

$$MSE = \hat{\sigma}^2 = \frac{SSE}{n-2}$$

Previously, we used SAS to generate the regression model. Well, SAS also computed MSE along with other metrics. Thus, our point estimate for σ^2 is,

$$MSE = \hat{\sigma}^2 = 66.80082$$

2.2

1. Well, in the hypothesis test, the analyst concluded $H_0 : \beta_1 \leq 0$ for the regression model $\hat{y} = \beta_0 + \beta_1 x$
2. In order for their to be no linear relationship between X and Y, the analyst would have to conclude that $\beta_1 = 0$
3. If $\beta_1 = 0$, that reduces the linear regression function to
$$\hat{y} = \beta_0$$
4. However, since the analyst concluded $H_0 : \beta_1 \leq 0$, that means that either $\beta_1 < 0$ or $\beta_1 = 0$.
5. Therefore, if $\beta_1 = 0$, then there would be no relationship between X and Y, but if $\beta_1 < 0$ then X and Y have a negative relationship.
6. Thus, with the concluded hypothesis, the relationship between X and Y is inconclusive since it could be negative or none depending on the value of β_1.

2.3

1. So the student's estimated regression function is,
$$\hat{Y} = 350.7 - 0.18X$$
2. However, the student was studying the relationship between advertising costs(X) and sales(Y), which means we would not expect a negative relationship since usually the more money you spend on advertising, the more sales you have.
3. We also know that the Two sided p-value for the estimated slope is 0.91.
4. Based on the results for the regression line, the hypothesis test must have been,
$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 = -0.18$$
5. Now, since p-value = 0.91, that means that,
$$p - value > \alpha$$
since we would never choose an $\alpha > 0.91$ because that would be a 9 percent confidence interval or less, which is useless.
6. Thus, the student should not reject H_0, however they incorrectly rejected H_0 and concluded H_a. As a result, the student is claiming that more advertising reduces sales, which is not a correct conclusion.

2.6

a) We want to estimate β_1 with a 95 percent confidence interval for Airfreight Breakage.

1) Using SAS (and from Problem 1.21), I generated the regression function for Airfreight Breakage. The regression function is,

$$\hat{y} = 10.2 + 4x$$

So, $b_0 = 10.2$ and $b_1 = 4$

2) From the notes, we know that the 100(1-α) confidence interval for b_1 is,

$$b_1 \pm t_{\alpha/2}(n-2) * S_{b_1}$$

3) However, first we must calculate S_{b_1} with the following formula,

$$S_{b_1}^2 = \frac{\hat{\sigma}^2}{S_{xx}}$$

where $\hat{\sigma}^2 = MSE$

4) From Problem 1.21, we have the the following results,

$$MSE = \hat{\sigma}^2 = 2.2$$

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 = 10$$

5) Now, we can use the results from (4) to find S_{b_1} like so,

$$S_{b_1}^2 = \frac{2.2}{10}$$

$$S_{b_1}^2 = 0.22$$

6) And, now we can compute S_{b_1} as such,

$$S_{b_1} = \sqrt{S_{b_1}^2} = \sqrt{0.22}$$

$$S_{b_1} = 0.469041576$$

7) Now, we need to calculate $t_{\alpha/2}(n-2)$

8) Since we want the 95 percent Confidence Interval, so

$$\alpha = 0.05 \qquad \frac{\alpha}{2} = 0.025 \qquad n = 10$$

9) Now, use α to find the t-value (in a two sided test),

$$t_{\alpha/2}(n-2) = t_{0.025}(10-2)$$

$$t_{\alpha/2}(n-2) = t_{0.025}(8)$$

10) Using R, I calcuated the t-value,

$$t_{0.025}(8) = 2.306004$$

11) So, substitute in values to the formula find the confidence interval,

$$b_1 \pm t_{\alpha/2}(n-2) * S_{b_1}$$
$$4 \pm t_{0.025}(8) * S_{b_1}$$
$$4 \pm 2.306004 * 0.469041576$$
$$4 \pm 1.08161175$$

12) Thus, the 95 percent confidence interval for β_1 is,

$$(2.91838825, 5.08161175)$$

13) Therefore, we a 95 percent confident that β_1 lies in the range of (2.91838825, 5.08161175).

b) Now, we want to do a t-test along with hypothesis testing to determine if there is a relationship between X and Y.

1) So, first since we want to test whether or not X and Y are related, our hypothesis test situation will be,

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

2) If we conclude H_0, then there is evidence to show that there is not likely a relationship between X and Y, but it we conclude H_a, then the evidence shows that there is likely a relationship between X and Y.

3) Now, we calculate the t-value with the formula,

$$t = \frac{b_1 - \beta_{10}}{S_{b_1}}$$

4) From Problem 2.6 (b), we know that,

$$b_1 = 4, \qquad S_{b_1} = 0.469041576$$

And, we are testing against to see if $\beta_1 = 0$, so $\beta_{10} = 0$

5) Substitute in the above values to calculate t-value,

$$t_{obs} = \frac{4 - 0}{0.469041576}$$
$$t_{obs} = \frac{4}{0.469041576}$$
$$t_{obs} = 8.528028654$$

6) From R, I computed the p-value using the following code,

```
t_obs <- 8.528028654
#df <- n-1
df <- 9
2*pt(-abs(t), df=9)
```

7) The computed p-value is 0.04653987

8) Compare the p-value to α,

$$p - value = 0.04653987 < 0.05 = \alpha$$

Thus, reject H_0 and conclude H_a

9) Hence, there is strong evidence to suggest that there is a relationship between X and Y.

c) Finally, we want to find a 95 percent confidence interval for β_0 and interpret it.

1) From the notes, we have that a $100(1-\alpha)$ percent confidence level for β_0 is,

$$b_0 \pm t_{\alpha/2}(n-2) * S_{b_0}$$

2) Now, we need to compute S_{b_0}, so we use the formula,

$$S_{b_0}^2 = \hat{\sigma}^2(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}})$$

3) Recall that $\hat{\sigma}^2 = MSE = \frac{SSE}{n-2}$ from the notes.

4) From Problem 1.25 (b) in Homework 2, I calculated the MSE, that is,

$$\hat{\sigma}^2 = MSE = 2.2$$

5) From the same Homework, I calculated S_{xx}, that is,

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 = 10$$

And n = 10, $\bar{x} = 1$, and $b_0 = 10.2$

6) Now, we just substitute each of these values into the formula for $S_{b_0}^2$, yielding,

$$S_{b_0}^2 = 2.2(\frac{1}{10} + \frac{1^2}{10})$$

$$S_{b_0}^2 = 2.2(\frac{1}{10} + \frac{1}{10})$$

$$S_{b_0}^2 = 2.2(\frac{1}{5})$$

$$S_{b_0}^2 = 0.44$$

7) Now, take the square root of $S_{b_0}^2$ and get,

$$S_{b_0} = \sqrt{0.44} = 0.6633249581$$

8) Since we are using a 95 percent confidence interval, with n = 10, we have the same t-value as the confidence interval from 2.6 (a),

$$t_{\alpha/2}(n-2) = t_{0.025}(8) = 2.306004$$

9) Finally, we have the necessary values to compute the confidence interval of β_0, as such,

$$b_0 \pm t_{\alpha/2}(n-2) * S_{b_0}$$

$$10.2 \pm 2.306004 * 0.6633249581$$

$$10.2 \pm 1.529630007$$

10) Thus, the 95 percent confidence interval for β_0 is,

$$(8.670369993, 11.72963001)$$

11) Therefore, we are 95 percent confident that β_0 lie in the range of (8.670369993, 11.72963001).