Homework 11 (47 points + 8.5 points BONUS)

(27 points LaunchPad + 20 points work + 2 points BONUS (LaunchPad) + 6.5 points BONUS Work

due Tuesday Dec. 8:11:55 pm

(1 pts.) 1 (Problem 1.1). In simple linear regression, both the t and F tests can be used as model utility tests.

A. True
B. False

True. F tests are used for model utility tests. However, in simple linear regression the t distribution can also be used if $\beta_{10} = 0$ and it is a two-sided test.

(1 pt.) 2 (Problem 2.1) The sample correlation coefficient is a measure of the strength of a linear relationship between two continuous variables.

A. True
B. False

True. By definition.

(1 pts.) 3 (Problem 3.1) A study reported a correlation r = 0.5 based on a sample size of n = 15; another study reported the same correlation based on a sample size of n = 25.

Without performing either studies, which of the following is correct?

A. This phenomenon of different conclusions for different sample sizes is typical only for r = 0.5.
B. For every two samples, the sample with the smaller sample size has more chances for a higher correlation.
C. With the larger sample size, we are more likely to get a significant result since r estimates r better in large samples.
D. With the larger sample size, we are less likely to get a significant result since r estimates r better in large samples.

C. r is a better estimate in larger samples

(1 pts.) 4 (Problem 4.1) Crimini mushrooms are more common than white mushrooms, and they contain a high amount of copper, which is an essential element according to the U.S. Food and Drug Administration. A study was conducted to determine whether the weight of a mushroom is linearly related to the amount of copper it contains. A random sample of crimini mushrooms was obtained, and the weight (in grams) and the total copper content (in mg) was measured for each.

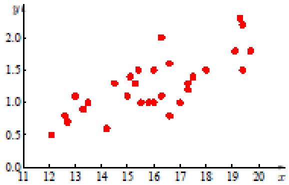
Which of the following is correct concerning which variable is X and which is Y?

A. The response variable is copper content and the explanatory variable is the mushroom weight.
B. The response variable is mushroom weight and the explanatory variable is the copper content.

A. It is easier to measure the total weight than to measure the amount of copper. Therefore, you would like to determine the amount of copper from the total weight.

(1 pts + 2 pts. work) 5 (Problem 4.2) Written. Crimini mushrooms are more common than white mushrooms, and they contain a high amount of copper, which is an essential element according to the U.S. Food and Drug Administration. A study was conducted to determine whether the weight of a mushroom is linearly related to the amount of copper it contains. A random sample of crimini mushrooms was obtained, and the weight (in grams) and the total copper content (in mg) was measured for each. The scatterplot is show below:

Commented [LAF1]: 1 pt.: work for r, this is the equation or the substituted numbers
1 pt.: relationship, each part is 0.5 pts, do not go above 1 pt. moderate, positive, linear, no outliers.



The summary statistics are: S_{XX}=137.48, S_{YY}=5.7787, S_{XY}=21.275
The sample correlation coefficient is _____ . (4 decimal places). Please show your work.

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{21.275}{\sqrt{(137.48)(5.7787)}} = 0.7548$$

On a separate piece of paper, using the scatterplot and the sample correlation coefficient, describe the relationship between crimini mushroom weight and copper content.

The relationship is a moderate positive relationship. The form looks linear and I don't see any outliers.

(1 pts. + 1.5 pts. BONUS) 6 (Problem 5.1) Written. An investigative reporter believes that certain automobile service stations that offer state vehicle inspections routinely charge for unnecessary repair work. Preliminary data suggest that the cost of the repair work may be related to the age of the car. A random sample of automobiles inspected at these stations was obtained, and the age (in years) along with the cost of the repairs (in dollars) were recorded for each vehicle. The summary data is given below:

Commented [LAF2]: Bonus: 0.5 separating the terms, 0.5: realizing that the second term is 0 (explanation is the summation and is worth 0.3 pts), 0.5: the rest of the derivation

n = 15, S_{XX}=82.89, S_{YY}=3848000, S_{XY}=7686, $\bar{x}$ = 4.887, $\bar{y}$ = 823.3

The sample correlation coefficient is ____ (4 decimal places)

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{7686}{\sqrt{(82.89)(3848000)}} = 0.4304$$

BONUS: Show that $\sqrt{\sum a_i^2} = \frac{1}{\sqrt{S_{XX}}}$ (Findsen's SLIDE 54, Chapter 12

$$b_1 = \frac{\sum[(x_i-\bar{x})(y_i-\bar{y})]}{\sum(x_i-\bar{x})^2} = \frac{\sum[(x_i-\bar{x})y_i]}{\sum(x_i-\bar{x})^2} - \frac{\sum[(x_i-\bar{x})\bar{y}]}{\sum(x_i-\bar{x})^2} = \frac{\sum[(x_i-\bar{x})y_i]}{\sum(x_i-\bar{x})^2}$$

$$\sum[(x_i-\bar{x})\bar{y}] = 0\ because\ \bar{y}\ is\ a\ constant\ and \sum(x_i-\bar{x}) = 0$$

$$Note\text{: }This\ means\ that\ a_i = \frac{(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}$$

$$Therefore, a_i^2 = \frac{(x_i-\bar{x})^2}{(\sum(x_i-\bar{x})^2)^2}$$

$$\sum a_i^2 = \sum \frac{(x_i-\bar{x})^2}{(\sum(x_i-\bar{x})^2)^2} = \frac{\sum(x_i-\bar{x})^2}{(\sum(x_i-\bar{x})^2)^2} = \frac{1}{\sum(x_i-\bar{x})^2} = \frac{1}{S_{xx}}$$

$$Therefore, \sqrt{\sum a_i^2} = \frac{1}{\sqrt{S_{xx}}}$$

(1 pts. + 2 pts. work) 7 (Problem 5.2) Written. An investigative reporter believes that certain automobile service stations that offer state vehicle inspections routinely charge for unnecessary repair work. Preliminary data suggest that the cost of the repair work may be related to the age of the car. A random sample of automobiles inspected at these stations was obtained, and the age (in years) along with the cost of the repairs (in dollars) were recorded for each vehicle. The summary data is given below:

Commented [LAF3]: 1 pt.: slope, equation or values
1 pt.: y-intercept: equation or values
-0.5 if the equation of the line is not explicitly stated

n = 15, S_{XX}=82.89, S_{YY}=3848000, S_{XY}=7686, $\bar{x}$ = 4.887, $\bar{y}$ = 823.3

The equation for the lines is: y^ = ___ (1 decimal places) + ____ (2 decimal places). Please show your work

$$b_1 = \frac{S_{XY}}{S_{XX}} = \frac{7683}{82.89} = 92.69$$

$$b_0 = \bar{y} - b_1\bar{x} = 823.3 - (92.69)4.887 = 370.3$$

$$\hat{y} = 370.3 + 92.69x$$

(1 pts. + 3 pts. work) 8 (Problem 5.3) Written. An investigative reporter believes that certain automobile service stations that offer state vehicle inspections routinely charge for unnecessary repair work. Preliminary data suggest that the cost of the repair work may be related to the age of the car. A random sample of automobiles inspected at these stations was obtained, and the age (in years) along with the cost of the repairs (in dollars) were recorded for each vehicle. The summary data is given below:

Commented [LAF4]: 1 pt.: Calculation of MSE: work required for SSR, SST, SSE, dfe, MSE. Give 0.2 pts for each part.
1 pt. calculation of the CI, 0.4: correct formula, 0.4: correct critical value, 0.2: answer
1 pt. Interpretation: 0.3 pts each (do not go over 1 pt.) 95% confident, population, slope, between values.

n = 15, S_{XX}=82.89, S_{YY}=3848000, S_{XY}=7686, $\bar{x}$ = 4.887, $\bar{y}$ = 823.3

The 95% confidence interval for the slope is (___, _____) (2 decimal places). Show your work. On a separate piece of paper, please write the interpretation of the interval.
Note: Even though the ANOVA table is not asked for, please calculate it and keep the table for the next part

To determine the standard error of the slope, we need the value of MSE, hence the ANOVA table needs to be calculated.

Source of variation	Sum of squares	Degrees of freedom	Mean square	F	P-value
Regression	712,415.34	1	712,415.34	2.954	0.1094
Error	3,135,584.66	13	241,198.82		
Total	3,848,000	14			

$SSR = b_1S_{xy} = 92.69(7686) = 712{,}415.34$
$SST = S_{yy} = 3{,}848{,}000$
$SSE = SST - SSR = 3{,}848{,}000 - 712{,}415.34 = 3{,}135{,}584.66$
dfr = 1
dfe = n – 2 = 15 – 2 = 13
dft = n – 1 = 15 – 1 = 14 = dfr + dfe
$MSE = \frac{SSE}{dfe} = \frac{3{,}135{,}84.66}{13} = 241{,}198.82$

$$F = \frac{MSR}{MSE} = \frac{712{,}415.34}{241{,}198.82} = 2.95$$

P = P(F > 2.95, df1 = 1, df2 = 13) = 0.1095

$$b_1 \pm t_{0.025,13}\sqrt{\frac{MSE}{S_{XX}}} = 92.69 \pm 2.1604\sqrt{\frac{241{,}198.82}{82.89}} = 92.69 \pm 116.539 \Longrightarrow (-23.85{,}209.23)$$

We are 95% confident that the population slope is between -23.85 and 209.23.

(1 pt. + 3 pts. work) 9 (Problem 5.4). Written. An investigative reporter believes that certain automobile service stations that offer state vehicle inspections routinely charge for unnecessary repair work. Preliminary data suggest that the cost of the repair work may be related to the age of the car. A random sample of automobiles inspected at these stations was obtained, and the age (in years) along with the cost of the repairs (in dollars) were recorded for each vehicle. The summary data is given below:

Commented [LAF5]: 0.5: NOT including step 1
0.5: hypothesis
0.5 pt.: calculation of F and df2. Additional work, 0.2: SSR or MSR, 0.3: the calculation of F, 0.1: writing down df1, df2
0.5 pts: P
0.5: decision (0.3) and why (0.2)
0.5: statement in context: 0.2 pts each (do not go over 0.5)
Not, P value, significance linear relationship, context

n = 15, S_{XX}=82.89, S_{YY}=3848000, S_{XY}=7686, $\bar{x}$ = 4.887, $\bar{y}$ = 823.3

On a separate piece of paper, please perform the 4 step hypothesis test for the F test. Remember, you need to show your work except for the values that you have used in previous parts. You have to determine the exact P-value using computer software.

The value of the test statistic is ____ (3 decimal places).

Step 1:
not required

Step 2:
H_0: There is no significant linear relationship
H_a: There is a significant linear relationship

Step 3:
F = 2.954 (see above)
df1 = 1, df2 = 13
p-value = 0.1094

Step 4:
Fail to reject H_0 since 0.1094 > 0.05.

The data does not provide sufficient support (P = 0.1094) to the claim that there is a significant linear relationship between the age of the car and the cost of repairs.

(1 pt. + 1 pt. work) 10 (Problem 5.5). Written. An investigative reporter believes that certain automobile service stations that offer state vehicle inspections routinely charge for unnecessary repair work. Preliminary data suggest that the cost of the repair work may be related to the age of the car. A random sample of automobiles inspected at these stations was obtained, and the age (in years) along with the cost of the repairs (in dollars) were recorded for each vehicle. The summary data is given below:

Commented [LAF6]: 1 pt: 0.2 pts: correlation, 0.4 pts: 0 is in the interval, 0.4: fail to reject

n = 15, S_{XX}=82.89, S_{YY}=3848000, S_{XY}=7686, $\bar{x}$ = 4.887, $\bar{y}$ = 823.3

Using the results of the sample correlation coefficient, the confidence interval and the hypothesis test, do you believe the reporter's claim? On a separate piece of paper, explain your answer using the three items mentioned above.

A. Yes
B. No

No. The correlation coefficient is moderately strong. However, the confidence interval and hypothesis test both show that the slope of the line could be 0 which means that there is not a significant linear relationship between the age of the car and the cost of repairs.

(1 pt.) 11 (Problem 6.1). Many factors affect the length of a professional football game. A study was conducted to determine the relationship between the total number of penalty yards (x) and the time required to complete a game (y, in hours). The following is the summary data:

n = 9, S_{XX}=26,256, S_{YY}=3.956, S_{XY}=244.8, MSE = 0.2390

The expected value of the slope is _____ . (6 decimal places)

$$b_1 = \frac{S_{XY}}{S_{XX}} = \frac{244.8}{26.256} = 0.009324$$

(1 pt.) 12 (Problem 6.2). Many factors affect the length of a professional football game. A study was conducted to determine the relationship between the total number of penalty yards (x) and the time required to complete a game (y, in hours). The following is the summary data:

n = 9, S_{XX}=26,256, S_{YY}=3.956, S_{XY}=244.8, MSE = 0.2390

The 95% confidence interval for the slope is (____ , _____) (6 decimal places)

$$b_1 \pm t_{0.025,7}\sqrt{\frac{MSE}{S_{XX}}} = 0.009324 \pm 2.3646\sqrt{\frac{0.2390}{26{,}256}} = 0.009324 \pm 0.007134$$
$$\Longrightarrow (0.002190, 0.016458)$$

(1 pt. + 3 pts work) 13 (Problem 6.3). Written. Many factors affect the length of a professional football game. A study was conducted to determine the relationship between the total number of penalty yards (x) and the time required to complete a game (y, in hours). The following is the summary data:

Commented [LAF7]: 0.5: beta_1 (0.2), population (0.2), slope (0.1)
0.5: hypothesis
0.5 pt.: calculation of t
0.5 pts: df. P
0.5: decision (0.3) and why (0.2)
0.5: statement in context: 0.2 pts each (do not go over 0.5) positive, P value, significance linear relationship, context

n = 9, S_{XX}=26,256, S_{YY}=3.956, S_{XY}=244.8, MSE = 0.2390

On a separate piece of paper, please perform the 4 step hypothesis test for the t test. Remember, you need to show all of your work. You have to determine the exact P-value using computer software.

The value of the test statistic is _____ (3 decimal places)

Step 1:
β_1 is the slope of the population regression line.

Step 2:
H_0: $\beta_1 = 0$
H_a: $\beta_1 \neq 0$

Step 3:

$$T = \frac{b_1 - \beta_{10}}{\sqrt{\frac{MSE}{S_{XX}}}} = \frac{0.009324 - 0}{\sqrt{\frac{0.2390}{26,256}}} = 3.090$$

df = 7
P = 2P(T > 3.090) = 0.01757

Step 4:
Reject H_0 since 0.01757 < 0.05.

The data does provide sufficient support (P = 0.01757) to the claim that there is a significant linear relationship between the number of penalty yards and the length of a professional football game..

(1 pt. + 1 pt. work) 14 (Problem 6.4). Written. Many factors affect the length of a professional football game. A study was conducted to determine the relationship between the total number of penalty yards (x) and the time required to complete a game (y, in hours). The following is the summary data:

Commented [LAF8]: 1 pt: 0.5 realizing that you have to square r,
0.5: work for r

n = 9, S_{XX}=26,256, S_{YY}=3.956, S_{XY}=244.8, MSE = 0.2390

The proportion of the observed variance in the time required to complete a professional football game due to the total number of penalty yards is ____ (4 decimal places). Please show your work

$$r^2 = \left(\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}\right)^2 = \left(\frac{244.8}{\sqrt{(26,256)(3.956)}}\right)^2 = 0.5769$$

(1 pt. + 1 pts. work) 15 (Problem 7.1). **Written**. **The temperature of the upper layer of ocean water is affected by sunlight and wind. There is often a very sharp difference in temperature between the surface zone and the more stationary deep zone. The thermocline layer marks the abrupt drop-off in temperature. The following data were obtained in a study of temperature (x, measured in °C) versus depth (y, measured in meters) above the thermocline layer in the Mediterranean Sea.**
The ANOVA table from the data is:

Commented [LAF9]: 1 pt.: 0.3 pts negative sign, 0.4 pts sqrt of r2, 0.3 pts work for r2.

source	SS	df	MS
Regression	108.54	1	108.54
Error	78.06	6	13.01
Total	186.6	7	

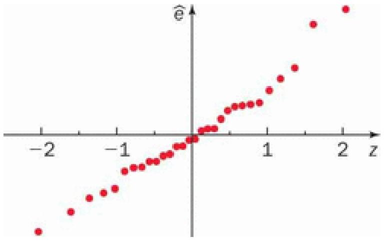
The equation of the line is: ŷ = 23.091 - 0.084 x

The correlation coefficient is _____ (4 decimal places). Please show your work

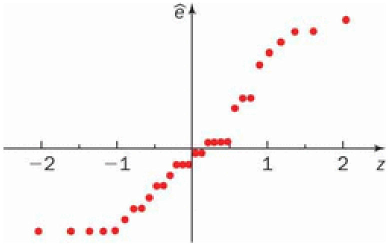
$$r = -\sqrt{r^2} = -\sqrt{\frac{SSR}{SST}} = -\sqrt{\frac{108.54}{186.6}} = -0.7627$$

(1 pt.) 16 (Problem 8.1) The following are four QQ -plots of the residuals from different data sets. For which of these plots is the normality assumption valid?

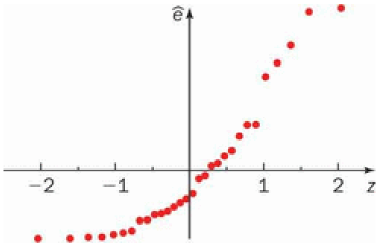
A.



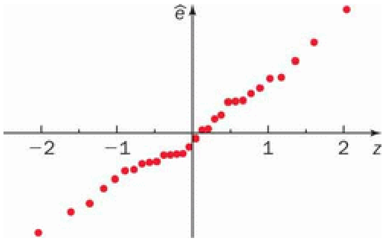
B.



C.

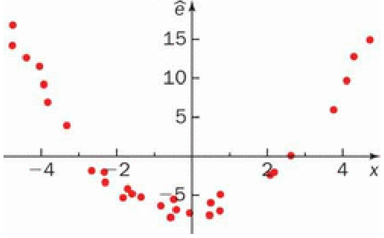


D.



A and D are the only plots that have points relatively close to the line. B is symmetric but not normal and C is right skewed.

(1 pt.) 17 (Problem 9.1). The following is a residual plot (residuals versus predictor variable) for a particular data set.

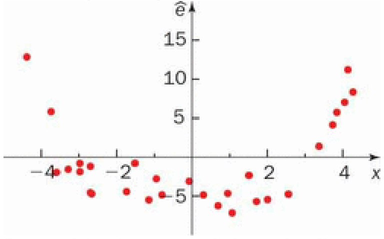


Does this data set violate the linear assumption and/or the constant variance assumption?

a. The data violates only the constant variance assumption
b. The data violates neither assumption.
c. The data violates only the linear assumption.
d. The data violates both assumptions.

c. The data violates only the linear assumption.

(1 pt.) 18 (Problem 9.2) The following is a residual plot (residuals versus predictor variable) for a particular data set.

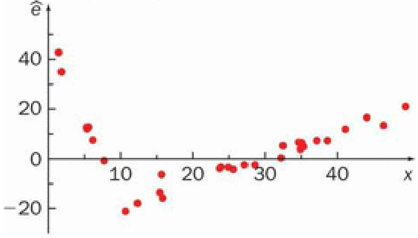


Does this data set violate the linear assumption and/or the constant variance assumption?

a. The data violates only the constant variance assumption
b. The data violates neither assumption.
c. The data violates only the linear assumption.
d. The data violates both assumptions.

d. The data violates both assumptions.

(1 pt) 19 (Problem 9.3). The following is a residual plot (residuals versus predictor variable) for a particular data set.

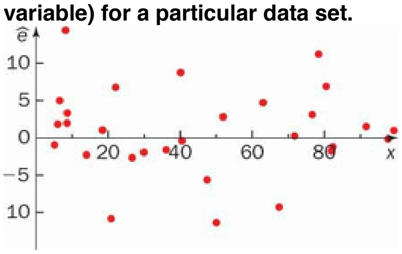


Does this data set violate the linear assumption and/or the constant variance assumption?

a. The data violates only the constant variance assumption
b. The data violates neither assumption.
c. The data violates only the linear assumption.
d. The data violates both assumptions.

c. The data violates only the linear assumption.

(1 pt.) 20 (Problem 9.4). The following is a residual plot (residuals versus predictor variable) for a particular data set.



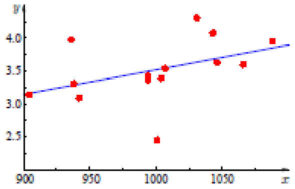
Does this data set violate the linear assumption and/or the constant variance assumption?

a. The data violates only the constant variance assumption
b. The data violates neither assumption.
c. The data violates only the linear assumption.
d. The data violates both assumptions.

a. The data violates constant variance. The variance of the points at the end are less than the variance of the points at the beginning.

(1 pt. + 1 pt. work) 21 (Problem 10.1) Written. Farmers in northern Sweden sell one of the most expensive cheeses in the world, which is made from moose milk. Suppose a random sample of female moose was obtained, and the weight of each was measured (x , in kilograms). The amount of milk produced by each moose in one day was also measured (y , in liters). The following is the scatterplot made from the data:

Commented [LAF10]: 1 pt. data is linear. I put in the second statement because of the other parts.



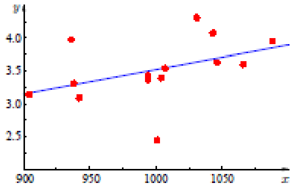
What it be appropriate to use this line to predict the amount of milk produced from a female moose who weighs 914 pounds? Please explain your answer.

A. Yes.
B. No

Yes. The data is approximately linear. We would expect the linear regression line to be a moderately good predictor of the data and 914 is within the range of the data.

(1 pt. + 1 pts. work) 22 (Problem 10.2). Written. Farmers in northern Sweden sell one of the most expensive cheeses in the world, which is made from moose milk. Suppose a random sample of female moose was obtained, and the weight of each was measured (x , in kilograms). The amount of milk produced by each moose in one day was also measured (y , in liters). The following is the scatterplot made from the data:

Commented [LAF11]: 1 pt.: this is an extrapolation so no



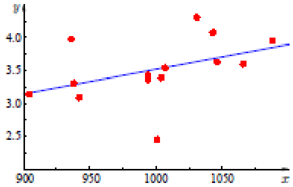
What it be appropriate to use this line to predict the amount of milk produced from a female moose who weighs 800 pounds? Please explain your answer.

A. Yes.
B. No

No. Our data only reflects moose who weigh between 900 and 1100 pounds. 800 pounds is outside of the interval that we have data from.

(1 pt. + 1 pts. work) 23 (Problem 10.3). Written. Farmers in northern Sweden sell one of the most expensive cheeses in the world, which is made from moose milk. Suppose a random sample of female moose was obtained, and the weight of each was measured (x , in kilograms). The amount of milk produced by each moose in one day was also measured (y , in liters). The following is the scatterplot made from the data:

Commented [LAF12]: 1 pt. this is an extrapolation so no



What it be appropriate to use this line to predict the amount of milk produced from a female moose who weighs 1200 pounds? Please explain your answer.

A. Yes.
B. No

No. Our data only reflects moose who weigh between 900 and 1100 pounds. 1200 pounds is outside of the interval that we have data from.

(1 pt.) 24 (Problem 11.1) For x = x* and a fixed confidence level, a prediction interval for an observed value Y is wider than a confidence interval for the mean value of Y.

A. True
B. False

True. The prediction interval also has a term for the variance of the point as well as the terms for the variance of the mean of the point.

(1 pt.) 25 (Problem 12.1) For a fixed confidence level, the width of a confidence interval for the mean value of Y is the same for any value of x*.

A. True
B. False

False. The formula for the confidence interval depends on the value of x*, specifically $(x - x^*)^2$. Thus, it will be different for a different value of x*.

(1 pt.) 26 (Problem 13.1) For x = x*, a confidence interval for the mean value of Y and a prediction interval for an observed value of Y are centered at the same value.

A. True
B. False

True. Both will be centered at $\hat{y}_{x^*}$

(1 pt. + 1 pts work) 27 (Problem 14.1) Written. A new solar collector is being tested for use in charging batteries that can provide electricity for an entire home. A random sample of days was selected and the amount of solar radiation was measured (x, in langleys) for each. The total battery charge was measured as a proportion (y, between 0 and 1). The summary statistics are given.

Commented [LAF13]: 1 pt.: 0.3 pts each (do not go over 1 pt.) 95% confident, population, slope, between values

$$\hat{\beta}_0 = 0.2007 \qquad n = 21 \qquad \text{MSE} = 0.06135$$
$$\hat{\beta}_1 = 0.00446 \qquad \bar{x} = 103.095 \qquad S_{xx} = 12335.8$$

Fill in the blanks. (Give your answer to five decimal places.)

The 95% confidence interval for the slope is (____ , _____). Please interpret your answer..

$$b_1 \pm t_{0.025,19}\sqrt{\frac{MSE}{S_{XX}}} = 0.00446 \pm 2.0930\sqrt{\frac{0.06135}{12335.8}} = 0.00446 \pm 0.004668 \Longrightarrow (-0.00021, 0.00913)$$

We are 95% confident that the true slope of the population regression line is between -0.00021 and 0.00913

(1 pt. + 2 pts. BONUS) 28 (Problem 14.2) BONUS Written. A new solar collector is being tested for use in charging batteries that can provide electricity for an entire home. A random sample of days was selected and the amount of solar radiation was measured (x, in langleys) for each. The total battery charge was measured as a proportion (y, between 0 and 1). The summary statistics are given.

Commented [LAF14]: 1 pt. work: 0.2: calculation of predicted value, 0.6: correct SE, 0.2: rest of work.
1 pt. interpretation: (do not go over 1 pt.) 95% confident (0.1), population or true (0.3) , mean (0.3), value at 130 (0.3), between values (0.2)

$$\hat{\beta}_0 = 0.2007 \qquad n = 21 \qquad \text{MSE} = 0.06135$$
$$\hat{\beta}_1 = 0.00446 \qquad \bar{x} = 103.095 \qquad S_{xx} = 12335.8$$

Fill in the blanks. (Give your answer to four decimal places.)

The 95% confidence interval for the mean value at the amount of solar radiation of 130 langleys is (_____ , ______). Please show your work and interpret your answer.

$\hat{y}_{130}$ = 0.2007 + 0.00446 (130) = 0.7805

$$\mu^*_{130} \pm t_{0.025,19}\sqrt{MSE\left(\frac{1}{n}+\frac{(x^*-\bar{x})^2}{S_{XX}}\right)} = 0.7805 \pm 2.0930\sqrt{(0.06135)\left(\frac{1}{21}+\frac{(130-103.095)^2}{12335.8}\right)}$$
$$= 0.7805 \pm 0.1690 \Longrightarrow (0.6114, 0.9495)$$

We are 95% confident that the true mean value of total battery charge at 130 langleys is between 0.6114 and 0.0.9495.

(1 pt. + 3 pts. Bonus) 29. (Problem 11.1) BONUS Written. A new solar collector is being tested for use in charging batteries that can provide electricity for an entire home. A random sample of days was selected and the amount of solar radiation was measured (x, in langleys) for each. The total battery charge was measured as a proportion (y, between 0 and 1). The summary statistics are given.

Commented [LAF15]: 1 pt. work: give full credit if they use the correct SE
1 pt. interpretation: (do not go over 1 pt.) 95% confident (0.1), population or true (0.3) , next value (0.3), value at 130 (0.3), between values (0.2)
1 pt different: 27: slope, 28/29 prediction. (0.3 pts.) 28 prediction for the mean value, 29: prediction for the next value, (0.3 pts.) prediction for next value always larger because of variance at point. (0.4 pts.)

$$\widehat{\beta}_0 = 0.2007 \qquad n = 21 \qquad \text{MSE} = 0.06135$$

$$\widehat{\beta}_1 = 0.00446 \qquad \overline{x} = 103.095 \qquad S_{xx} = 12335.8$$

Fill in the blanks. (Give your answer to four decimal places.)
The 95% confidence interval for the observed value at the amount of solar radiation of 130 langleys is (____ , _____). Please show your work and interpret your answer.

$$\mu_{130}^{*} \pm t_{0.025,19}\sqrt{MSE\left(1+\frac{1}{n}+\frac{(x^{*}-\bar{x})^2}{S_{XX}}\right)}$$

$$= 0.7805 \pm 2.0930\sqrt{(0.06135)\left(1+\frac{1}{21}+\frac{(130-103.095)^2}{12335.8}\right)} = 0.7805 \pm 0.5453$$

$$\Longrightarrow (0.2352, 1.3258)$$

We are 95% confident that the next observed value of total battery charge at 130 langleys is between 0.2352 and 1.3258.

On a separate piece of paper, explain the difference between this problem and the previous two pars. In your explanation, please explain why the widths are different and why the order of the sizes of the widths is always the same.

This problem is different from the previous one because in the previous one we are looking at the interval of the mean value at 130 langleys and in this problem, we are looking at the interval for the next observation. Because a prediction interval also has to consider the variance of the point itself, it will always have a larger spread.

The problem before the last one is concerning the slope parameter which is used in association rather than what is used in prediction.