HW 8: Inference for β_1, μ_k, and Y_p in SLR

Instructions: Work must be shown to receive full credit. You may work with others on the homework, but you must write and turn in your own copy. **This does not mean that you can simply copy someone else's work!!** Also, make sure your homework is neat, stapled, and all answers are written in complete sentences!! Come and see me if you have any questions.

On problems that require the use of R, PLEASE give me the RELEVANT R code and output to for each problem so I can assess partial credit. I may take off for including unnecessary R output. If one problem refers back to output from another problem, make sure to cite that output in your answer. Incorrect one-sentence answers will get little or no credit.

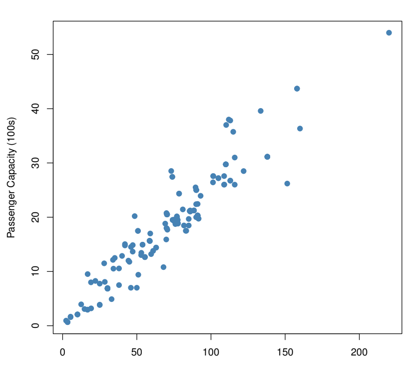
NOTE: If a problem asks you to perform a hypothesis test, make sure to give the hypotheses, test statistic, p-value, and a conclusion in the terms of the problem. Also, if the problem asks you to perform a confidence interval, make sure to interpret the confidence interval.

"By Hand" Problems: For hypothesis tests, you may use R to find the p-value. For confidence intervals, you may use R to find the multiplier.

1. (Continuation of HW 7, Problem 3) Suppose it is of interest to examine the relationship between the size of cruise chips and their passenger capacity. A data from 158 cruise ships was collected based on the following two variables:

   - Size of the ship (`Tonnage`) - the gross tonnage in 1000s of tons
   - Maximum passenger capacity (`Passengers`) - maximum number of passengers in 100s of people

   A regression analysis in R was performed with the partial results given below along with a scatterplot of the data.




```
> cruise.mod = lm(Passengers~Tonnage)
> summary(cruise.mod)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.946302   0.546733   1.731   0.0855 .
Tonnage     0.245650   0.006803  36.109   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 3.173 on 156 degrees of freedom
Multiple R-squared:  0.8931,    Adjusted R-squared:  0.8925
F-statistic:  1304 on 1 and 156 DF,  p-value: < 2.2e-16
```

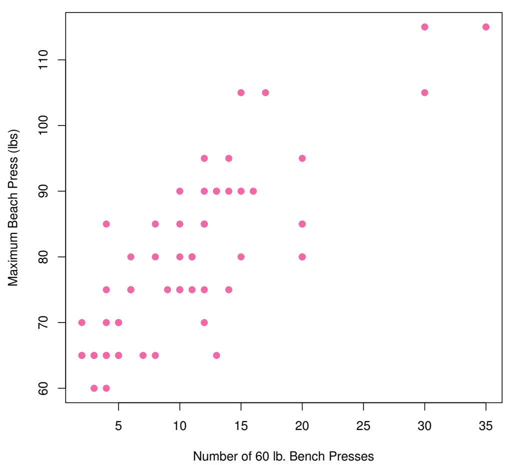
   Use the partial R output above to answer the following questions.

(a) Obtain a point estimate of σ^2 from the output.

(b) Compute and interpret a 95% confidence interval for the slope.

(c) Is there evidence that the size of the ship is linearly related to the number of passengers? Use the `R` output to conduct the appropriate hypothesis test. Show all five steps as usual.

2. (Continuation of HW 7, Problem 4) How can you measure a person's strength? One way is to find the maximum number of pounds that the individual can bench press. However, this technique can be risky for people who are unfamiliar with proper lifting techniques or who are inexperienced in using a bench press. Is there a variable that is easier to measure yet is a good predictor of the maximum bench press? A recent study of 57 high school female athletes in Georgia studied several measures of strength including ones that are easier and safer to assess than maximum bench press but are thought to relate highly with it. One such variable is the number of times she can lift 60 pounds before she becoming too fatigued to lift it again. It is of interest to examine how well the number of times someone can lift 60 pounds predicts the maximum bench press? The following summary information and scatterplot of the data was obtained to aide in the analysis.



	Mean	Std. Dev.
Maximum Bench (lbs)	79.9123	13.2790
60 lb. Bench Number	10.9825	13.63

$S_{XY} = 4259.912$ and $\sum_{i=1}^{n} e_i^2 = 64.05103$

(a) Obtain a point estimate for the variance of the residuals, $\hat{\sigma^2}$.

(b) Obtain the point estimate for the slope and the standard error of the estimate (i.e $\hat{\beta}$ and $SE_{\hat{\beta}}$).

(c) Construct and interpret a 90% confidence interval for the slope.

(d) Based on the interval, is there evidence of a linear relationship between reps at 60 lbs and the maximum bench press number? Explain using the interval only.

(e) Suppose we wish to predict the maximum bench for Suzy, a new friend at the gym. She can bench 60 lbs a total of 20 times. What kind of interval is most appropriate here (confidence or prediction)? Explain.

(f) Construct and interpret the indicated interval at 90% confidence.

3. Dr. Fribance teaches MCSI classes at CCU. One of her classes in particular involves collecting data on the boat. It is well known that increasing temperatures allow for increased conductivity in water. It is important to study this in different settings and types of water (ex. Hayashi, 2004). Using the student collected data consisting of 55 observations, what kind of conclusions can we draw on the effects of temperature on water conductivity?

	Mean	Std. Dev.
X (Temperature in Celsius)	21.87	0.3249
Y (Conductivity in s/m)	49.48	0.4188

$S_{XY} = 7.24$ and $\sum e_i^2 = 0.26$

   (a) Obtain a point estimate for the variance of the residuals, $\hat{\sigma}^2$.

   (b) Obtain the LSR line.

   (c) We wish to estimate the average conductivity when the water is 18^{o}C. What type of interval would be most appropriate (confidence or prediction)? Explain.

   (d) Construct and interpret the indicated interval with 95% confidence.

   (e) Suppose we wished estimate the average conductivity when the water is 22^{o}C. Without constructing the interval, explain why it would be wider or narrower than the interval in (d).

   (f) Suppose we wished to estimate the conductivity in a new water source that is 18^{o}C. Would the interval be wider or narrower than the interval in (d). Explain without constructing the interval.

"R" Problems:

4. On the class period before spring break, a couple of statistics classes decided to change things up. Each student chose a Girl Scout Cookie and measured its length (cm). Each student then proceeded to nibble the cookie and measure the length (cm) after each nibble. The data are provided on Moodle in the `cookie.xlxs` file. The question of interest is "How does the size of the cookie change with each bite of the cookie?"

   (a) Using `R`, obtain the LSR line. Report the input, complete output, and write down the equation of the line and the estimate $\hat{\sigma}^2$.

   (b) Using the `R` output in (a), write up a complete hypothesis test for to determine if length of the cookie is linearly related to the number of bites.

   (c) Using `R`, obtain a confidence interval for the slope and interpret in context.

   (d) Using `R`, obtain 95% intervals for the mean length of cookies when 0, 1, and 2 bites have been taken. Report all input and output. Interpret the interval for 2 bites.

5. Dr. Fribance teaches MCSI classes at CCU. One of her classes in particular involves collecting data on the boat. It is well known that increasing temperatures allow for increased conductivity in water. It is important to study this in different settings and types of water (ex. Hayashi, 2004). Using the student collected data consisting of 55 observations, what kind of conclusions can we draw on the effects of temperature on water conductivity? From the Moodle page, load the `FribanceStation1.csv` data set collected from Station 1 (south of Georgetown) in 2017.

   (a) Using `R`, obtain the LSR line. Report the input, complete output, write down the equation of the line, and the estimate $\hat{\sigma}^2$.

   (b) Using `R`, obtain a confidence interval for the slope and interpret in context.

   (c) Using `R`, obtain a 95% interval for the conductivity of an individual location when the temperature is 18^oC. (Note: this should verify your answer in 3f).

   (d) Interpret the interval in (c).