HW 7: OR and Introduction to Regression

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. A statistics class conducted an experiment to see if the distance of a toss determined the success of a toss. Students tossed a coin in a cup from either a short distance (3ft) or a long distance (6ft). Each attempt was recorded as a make or miss. The data are summarized below.

		Outcome		
		Made it!	Missed it!	Total
Distance	Short	9	9	18
	Long	2	17	19
	Total	11	26	37

   (a) **Calculate and interpret** the odds of making the toss from a short distance.

   (b) **Calculate and interpret** the odds of making the toss from a long distance.

   (c) Calculate the odds ratio for the odds of making it at a short distance versus long distance.

   (d) Construct a 95% confidence interval for the odds ratio.

   (e) Interpret the interval in context.

2. A vaccine to prevent severe rotavirus gastroenteritis (diarrhea) was given to African children withing the first year of life as part of a drug study. The study reported that of the 3298 children randomly assigned the vaccine, 63 got the virus. Of the 1641 children randomly assigned the placebo, 80 got the virus.

   (a) **Calculate and interpret** the odds of NOT getting sick on the vaccine.

(b) **Calculate and interpret** the odds of NOT getting sick on the placebo.

(c) Calculate the odds ratio for the odds of NOT getting sick on the vaccine versus placebo.

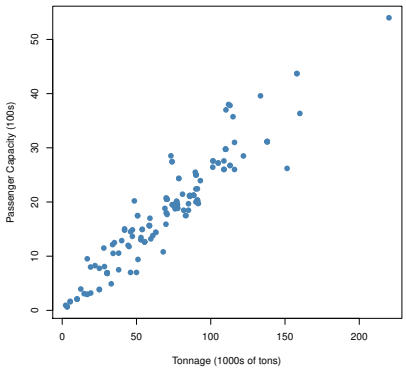
(d) Construct a 95% confidence interval for the odds ratio.

(e) Interpret the interval in context.

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)

             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 ***
```

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

   (a) Identify the response(dependent variable) and explanatory variable(independent).

   (b) Describe the association between size of the ship and number of passengers in a sentence.

   (c) Write down the estimated regression equation from the output.

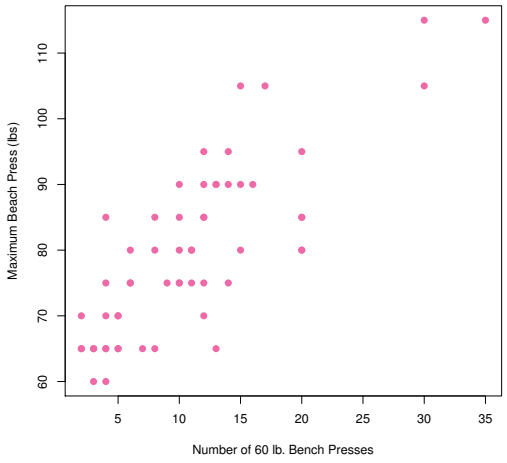
   (d) Interpret the slope in context of the problem.

   (e) Interpret the intercept in the context of the problem. Is this interpretation appropriate?

   (f) Suppose a cruise ship has a gross tonnage of 175,000 tons. What is the estimated passenger capacity for this ship?

   (g) The *Carnival Fantasy* has a gross tonnage of 70,367 tons and has a passenger capacity of 2,056. Find the residual for this value.

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	7.1427

$S_{XY} = 4259.912$ and $S_{XX} = 10403.58$

(a) What are the explanatory and response variables for this study?

(b) Calculate and interpret the slope of estimated regression equation.

(c) Calculate and interpret the intercept (if appropriate) of the estimates regression equation.

(d) Write down the LSR line.

(e) Obtain a point estimate of the mean maximum bench press for all high school female athletes who can lift 60 lbs. a total of 25 times.

(f) A female who can lift 60 lbs. a total of 20 times has a maximum bench of 100lbs. Obtain the residual for this observation.

"R" Problems:

5. Use `R` to verify the interval found in (1). Report your input and output.

6. Use `R` to verify the interval found in (2). Report your input and output.

7. 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) Identify the response and explanatory variables in this very scientific study.

   (b) Using `R`, obtain the LSR line. Report the input, output, and write down the equation of the line.

   (c) Using `R`, construct a well labeled scattterplot of the data and overlay the regression line on the plot.

   (d) Interpret the slope and intercept in context. Is the interpretation of the intercept meaningful? Explain.

   (e) Using `R`, obtain the predicted length for 0, 1, 2, and 3 bites of a cookie.

   (f) What are the first two observations in the data set? Find the residuals for these observations using `R`.