Inference for Linear Regression

Mini-Assignment - MTH 361 A/B - Spring 2023

Instructions:

- Please provide complete solutions for each problem. If it involves mathematical computations, explanations, or analysis, please provide your reasoning or detailed solutions.
- Note that some problems have multiple solutions or ways to solve it. Make sure that your solutions are clear enough to showcase your work and understanding of the material.
- Creativity and collaborations are encouraged. Use all of the resources you have and what you need to complete the mini-assignment. Each student must take personal responsibility and submit their work individually. Please abide by the University of Portland Academic Honor Principle.
- **Please save your work as one pdf file, don't put your name in any part of the document, and submit it to the Teams Assignments for this course. Your document upload will correspond to your name automatically in Teams.**
- If you have questions or concerns, please feel free to ask the instructor.

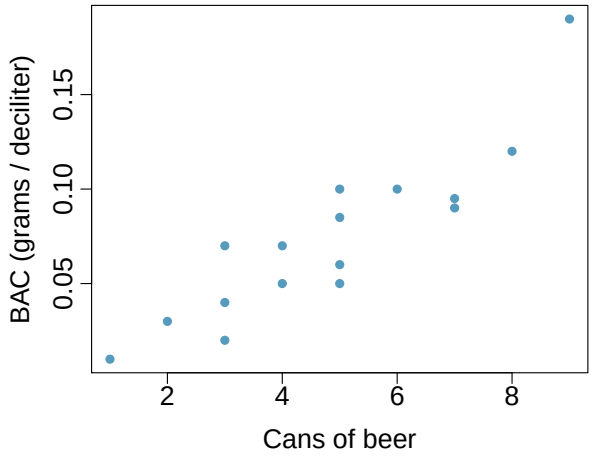
I. Infering Relationships Between Numerical Variables

Materials

The exercises below are derived from the textbook OpenIntro Statistics (4th edition) by David Diez, Mine Cetinkaya-Rundel, and Christopher Barr.

Exercises

1. **Beer and blood alcohol content.** Many people believe that gender, weight, drinking habits, and many other factors are much more important in predicting blood alcohol content (BAC) than simply considering the number of drinks a person consumed. Here we examine data from sixteen student volunteers at Ohio State University who each drank a randomly assigned number of cans of beer. These students were evenly divided between men and women, and they differed in weight and drinking habits. Thirty minutes later, a police officer measured their blood alcohol content (BAC) in grams of alcohol per deciliter of blood. (Malkevitch & Lesser, 2008) The scatterplot and regression table summarize the findings.



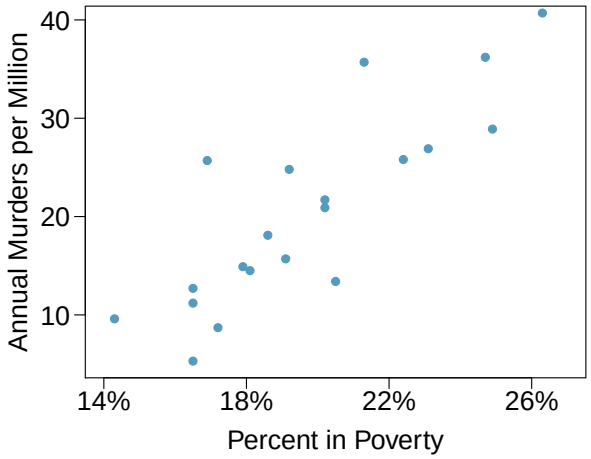
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.0127	0.0126	-1.00	0.3320
beers	0.0180	0.0024	7.48	0.0000

   a. Describe the relationship between the number of cans of beer and BAC.
   b. Write the equation of the regression line. Interpret the slope and intercept in context.
   c. Do the data provide strong evidence that drinking more cans of beer is associated with an increase in blood alcohol? State the null and alternative hypotheses, report the p-value, and state your conclusion.
   d. The correlation coefficient for number of cans of beer and BAC is 0.89. Calculate R^2 and interpret it in context.
   e. Suppose we visit a bar, ask people how many drinks they have had, and also take their BAC. Do you think the relationship between number of drinks and BAC would be as strong as the relationship found in the Ohio State study?

2. **Murders and poverty.** The following regression output is for predicting annual murders per million from percentage living in poverty in a random sample of 20 metropolitan areas.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-29.901	7.789	-3.839	0.001
poverty%	2.559	0.390	6.562	0.000

$s = 5.512 \quad R^2 = 0.7052 \quad R^2_{adj} = 0.6889$



a. Write out the linear model. Interpret the intercept, the slope, and R^2.
b. What are the hypotheses for evaluating whether poverty percentage is a significant predictor of murder rate?
c. State the conclusion of the hypothesis test from part (a) in context of the data.
d. Calculate a 95% confidence interval for the slope of poverty percentage, and interpret it in context of the data.
e. Do your results from the hypothesis test and the confidence interval agree? Explain.

3. (Outstanding Question) **Babies.** Is the gestational age (time between conception and birth) of a low birth-weight baby useful in predicting head circumference at birth? Twenty-five low birth-weight babies were studied at a Harvard teaching hospital; the investigators calculated the regression of head circumference (measured in centimeters) against gestational age (measured in weeks). The estimated regression line is

$$\widehat{head\ circumference} = 3.91 + 0.78 \times gestational\ age$$

   a. What is the predicted head circumference for a baby whose gestational age is 28 weeks?
   b. The standard error for the coefficient of gestational age is 0.35, which is associated with $df = 23$. Does the model provide strong evidence that gestational age is significantly associated with head circumference?

References

Malkevitch, J., & Lesser, L. M. (2008). *For All Practical Purposes: Mathematical Literacy in Today's World.* WH Freeman & Co.