BIOS 507 HOMEWORK 2

Due 2/24/2025 by 11:59pm

Directions: Complete all questions. Any R or SAS code used should be attached at the end of the homework. Collaboration is encouraged, but the final product must be your own work.

Problem 1

For each of the situations below, define the predictor variable and the response variable, write out the population-level model, and interpret the slope. Note that since you will not have "numbers" for the slope, you can just interpret everything in terms of "β_1"

1. A researcher is interested in the relationship between salary and number of years of experience in software development jobs. They collect data on 100 developers with a range of 5–20 years of experience, and measure each person's salary in thousands of dollars.

2. A hardware store chain is interested in running a new sales promotion on refrigerators, and they are trying to assess the relationship between the sale price and total refrigerator units sold. The company selects 40 stores across the country and has each of them apply a different discount with values ranging from 100 to 1000 dollars. The sales promotion is continued for 8 weeks, and at the end of the 8 weeks the total number of refrigerators sold is measured.

Problem 2

On Canvas, you have a data set called `solar.txt` containing data collected during a solar energy project at Georgia Tech. The data contain several columns, but for now we are going to focus on heat flux (column labeled Y) measured in kilowatts and radial deflection of the deflected rays (column labeled X_4) measured in milliradians. The researchers are interested in using the radial deflection to predict the heat flux.

1. What exploratory analyses should you do using the data? Conduct these and report your findings as well as any supporting figures.
2. Write out the assumed regression model for Y. What are your assumptions about the model error?
3. Fit the model using R or SAS. Write out the estimated model.
4. Fill out the ANOVA table for this analysis.
5. Interpret the R^2 value for this model.
6. Carry out a hypothesis test to test the null hypothesis that the slope is 0. Be sure to write out the α, the null and alternative hypothesis, the test statistic, critical value, and your final decision. Interpret the result in the context of the study.
7. Find and interpret a 99% confidence interval for the slope.
8. Find and interpret a 95% confidence interval for the mean heat flux when the radial deflection is 16.5 milliradians.
9. The lab would like to predict the heat flux when the radial deflection is 16.5 milliradians for a new measurement. Give a 95% prediction interval on the kilowatts.
10. Which interval is wider? Why?

Problem 3

This example is adapted from "A modern approach to regression with R" by Simon Sheather. The manager of the purchasing department of a large company is interested in developing a regression model to predict the average amount of time it takes to process a given number of invoices. Data were collected over a period of 30 days. For each data point, information was collected on:

- The number of invoices processed (`Invoices` in the dataset)
- The number of hours it took to process the set of invoices (`Time` in the dataset)

The data are provided on Canvas in `invoices.txt`.

1. What exploratory analyses should you do using the data? Conduct these and report your findings as well as any supporting figures.
2. Write out the assumed regression model for Y. What are your assumptions about the model error?
3. Fit the model using R or SAS. Write out the estimated model.
4. Fill out the ANOVA table for this analysis.
5. Interpret the R^2 value for this model.
6. Carry out a hypothesis test to test the null hypothesis that the slope is 0. Be sure to write out the α, the null and alternative hypothesis, the test statistic, critical value, and your final decision. Interpret the result in the context of the study.
7. Find and interpret a 99% confidence interval for the slope.
8. Find and interpret a 95% confidence interval for the amount of time it would take to process a stack of 160 invoices.
9. Find and interpret a 95% prediction interval for the amount of time it would take to process a new stack of 160 invoices.
10. Which interval is wider? Why?