

R Report III

FORMAT

- * Use complete sentences and proper grammar to answer all questions.
- * Use R Markdown to create an html document.
- * Code should not be in the body of the text, so be sure to add `echo = FALSE` in the preface to your R chunks. All code should be included at the end of the homework, as an appendix.

I. On Canvas you will find the file `crime.csv`. It has two columns, one of which is the percentage of individuals in the county with at least a high-school diploma (column `dip`), and the other is the crime rate per 100,000 residents for the counties (column `rate`). Consider Y to be crime rate, and X to be percentage with high school diploma.

Use R to complete the following tasks:

- (a) Plot a scatter plot of Y and X , being sure to label the axes and give a main title.
- (b) Calculate the estimated regression line.
- (c) Interpret the slope and intercept (if appropriate) in terms of the problem.
- (d) Visually, does there appear to be outliers in the plot from (a)? If so, identify them in R (for example, list the pair (X, Y) that are outliers, or equivalently the row). Note: You do not need to calculate the outliers. Just by visual inspection of the Scatterplot, subset the data to identify the point(s).
- (e) Create a QQ plot (normal probability plot) of the residuals. Does it appear that they are normally distributed? Explain.
- (f) Create a plot of the errors vs. the fitted values (\hat{Y}_i 's). Does it appear the variance of the errors is constant? Explain.
- (g) Find the 95% confidence interval for the slope, and interpret it in terms of the problem. Does the interval suggest there is a significant linear relationship? Explain.

II. We have learned a variety of commands in R this past quarter. Those include:

- . Basic calculations, built-in R functions, how to subset data
- . Plotting graphs - Pie chart, boxplot, histogram
- . If-then-else, `lapply`, `sapply`, and the sample functions
- . Creating our own function
- . Simulating random variables from specific distributions
- . Linear regression

Do a little research and determine how to conduct a one-sample t-test to test the claim that the true average percent of people with a high school diploma is 80. Use $\alpha = 0.10$.

- (a) State H_0 and H_A
- (b) Calculate the test statistic.
- (c) Calculate and interpret the p-value.
- (d) Make your decision.
- (e) State your conclusion in terms of the problem.
- (f) Construct and interpret a 90% confidence interval. Does this support your decision?
- (g) What kind of error could you have made in your hypothesis test? Interpret this error.