STAT 101 Exam #1 Form A

Name (print): _____ Section:

February 24th, 2020

Instructions

You have 50 minutes to complete this exam. There are a total of 65 points to earn on the exam along with 2 bonus points. Read the questions carefully and completely and give complete answers to all questions. Partial credit will be given where it is deserved.

You are allowed to use only a pencil and a calculator to complete this exam. All necessary information will be provided in the question space. A formula sheet is attached at the back.

Show your work. Draw pictures, write formulas and equations, and don't forget units where needed. **Good luck!**

Problem 1. (2 points) Which of the following are quantitative variables? Circle all correct answers.

a) Education in number of years

b) Age in years

c) Highest level of education achieved

d) Pizza preference

e) Income in terms of Socio-economic class

Problem 2. (2 points) Which of the following are categorical variables? Circle all correct answers.

a) Test Scores

b) Eye Colour

c) Height

d) Political Party Identification

e) Distance in miles

Problem 3. (2 points) What type of variable can be used in mathematical calculations?

a) Categorical Variables

b) Response Variables

c) Ordinal Variables

d) Lurking Variables

e) Quantitative Variables

Problem 4. (2 points) Which type of data visualization is good for comparing a quantitative variable across different categories?

a) Pie Chart

b) Histograms

c) Box Plots

d) Stem-and-Leaf Plot

e) Mosaic Plot

Problem 5. (2 points) What is it called when a summary statistic is not easily effected by outliers.

a) Static

b) Outlier

c) Robust

d) Not Robust

e) Skewed

Problem 6. (2 points) What is it called when data accumulates around smaller values and trails off towards larger values in a histogram?

a) Symmetric

b) Left-Skewed

c) Harmonic

d) Right-Skewed

e) Uniform

Problem 7. (2 points) If the mean is greater than the median for a dataset visualized as a histogram, how would you describe the symmetry of the histogram?

a) Right-Skewed

b) Left-Skewed

c) Unimodal

d) Symmetric

e) Outliers Present

Problem 8. (2 points) The marginal distribution is found in which part of the contingency table?

a) Columns

b) Rows

c) The interior of the table

d) The margins of the table

e) The marginal distribution is not found in a contingency table

Problem 9. (3 points) Label the following components of the box plot below. If it is unclear where you are placing the label on the box plot, no credit will be given.

a) Outlier

b) Q1

c) Median

d) Smallest observation less than Q3 + 1.5*IQR

e) 75th percentile
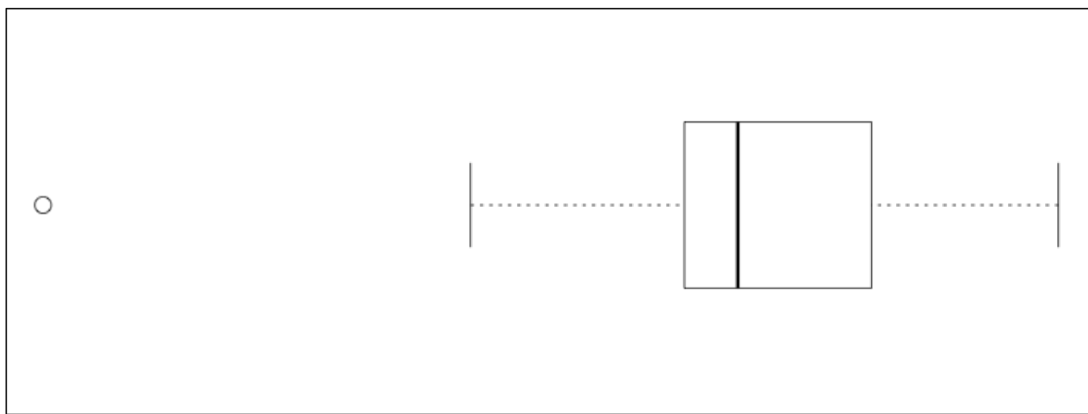
**Problem 10.** (3 points) Which of the following measures of center and spread are affected by outliers.

$$\textbf{I. } \text{Mean} \quad \textbf{II. } \text{Median} \quad \textbf{III. } \text{Range} \quad \textbf{IV. } \text{IQR} \quad \textbf{V. } \text{Standard Deviation}$$

a) I. & II.

b) II., III., and IV.

c) I., V., and IV.

d) All of them.

e) None of them.

Problem 11. (8 points) Match each statistical concept to its mathematical symbol. Each symbol is only used once

$$\textbf{I. } \rho \quad \textbf{II. } b_o \quad \textbf{III. } s_x \quad \textbf{IV. } \bar{x} \quad \textbf{V. } b_1 \quad \textbf{VI. } R^2 \quad \textbf{VII. } e \quad \textbf{VIII. } \hat{y}$$

a) Mean of x _____

b) Standard Deviation of x _____

c) Correlation _____

d) Residual _____

e) Predicted value of y _____

f) Coefficient of Determination _____

g) Slope of Linear Regression _____

h) Y-intercept of Linear Regression _____

Use the following histogram to answer questions 12 and 13.
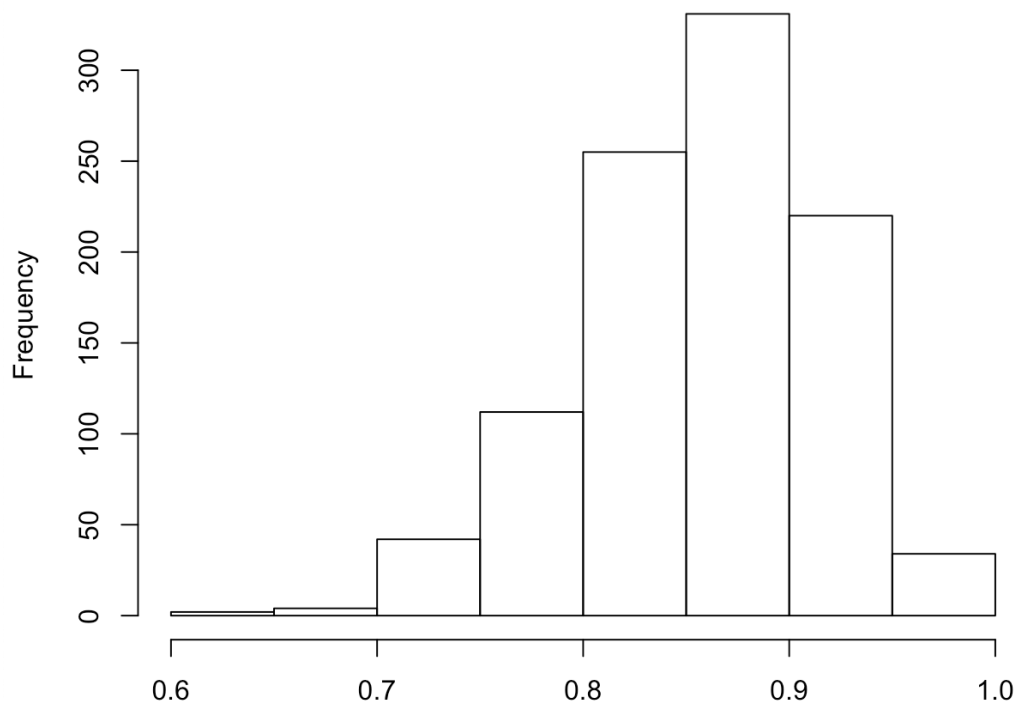
Problem 12. (3 points) Describe the shape of this distribution.

**Problem 13.** (1 points) Which measure of spread is most appropriate for this data.

a) Range

b) Standard Deviation

c) IQR

Problem 14. (11 points) **Flight Etiquette:** A study commissioned by FiveThirtyEight about flying etiquette. Data was collected in 2014 as a survey of 1,040 respondents. In particular we are interested in whether it is considered rude to recline your seat on a flight and how often participants due in fact recline their seats. Use the data table and mosaic plot provided to answer the following questions.

| Is it rude to recline?	How often do you recline? Always	Usually	Half the Time	Once in a while	Never	Total
Yes, very rude	3	3	0	11	54	71
Yes, somewhat rude	9	27	35	129	81	281
No, not rude at all	124	145	82	116	35	502
Total	136	175	117	256	170	854

a) (1 points) Identify the Who of the dataset.

b) (1 points) Identify the What of the dataset.

c) (1 point) What is the explanatory variable?

d) (2 points) What is the proportion of people who think it is somewhat rude to recline and recline more than once in a while?

e) (2 points) Give the conditional distribution "How often do you recline?" given that we are only interested in responses from people who do not think it is rude.

f) (3 points) Give the marginal distribution of "How often do you recline?". Compare this distribution to the distribution in (e)
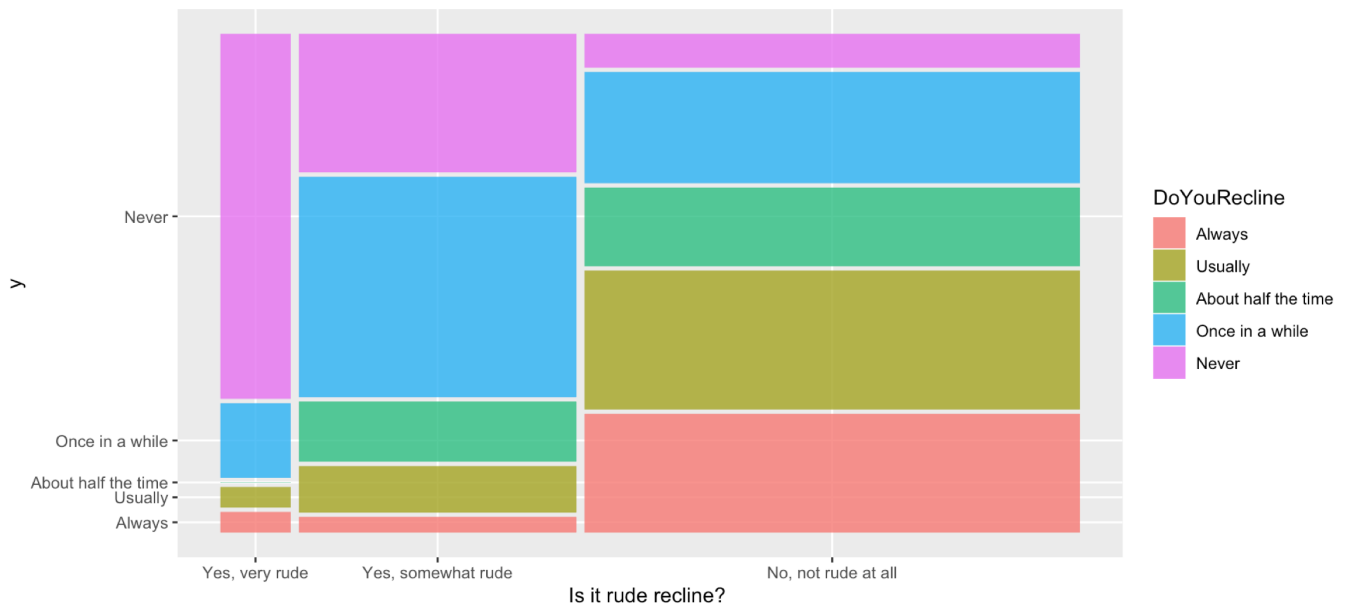
g) (1 points) Based on the mosaic plot given, is there an association between thinking reclining on an airplane is rude and the frequency at which you recline your seat?

Problem 15. (20 points) Data was extracted from the 1974 Motor Trend US magazine and comprisesfuel consumpation and 10 aspects of automobile design and performance for 32 different types of automobiles. Suppose we are interested in studying the relationship between weight (in thousands of pounds) and mpg(miles per gallon).
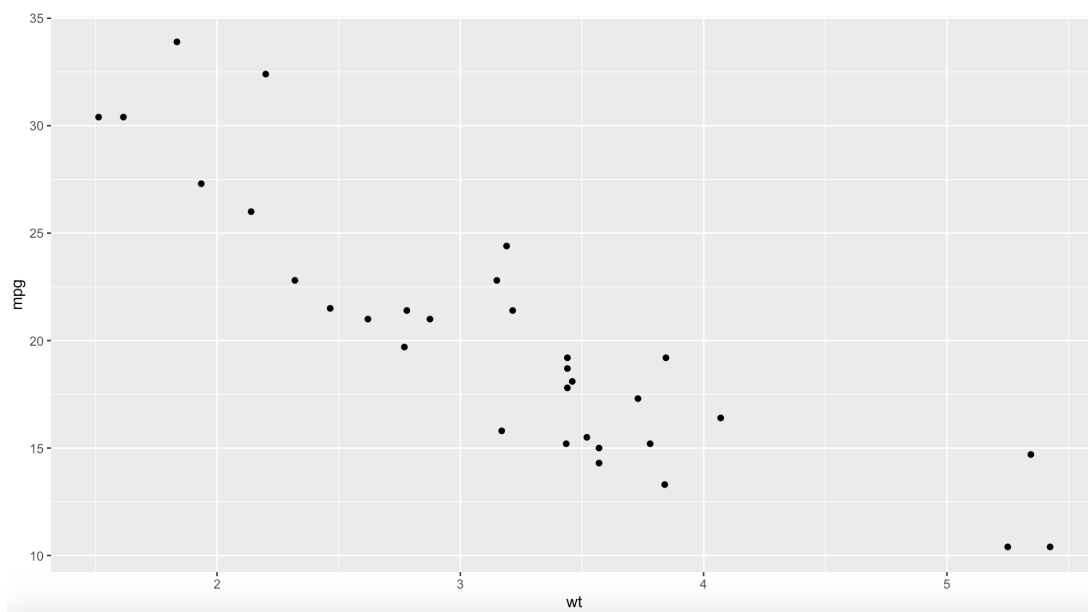

a) (3 points) Describe the relationship between weight and mpg.

b) (4 points)Based on the following summary statistics, calculate the linear regression between the weight of automobiles in thousands of pounds and the mpg.

$$\bar{x} = 3.217 \quad \bar{y} = 20.091 \quad s_x = 0.978 \quad s_y = 6.027 \quad \rho = -0.868$$

c) (3 points) Calculate the residual for a car that weights 4,500 pounds and has an actual mpg of 21. Is this extrapolation?

d) (2 points)Comment on the residual plot below. Is a linear regression model appropriate for this data?
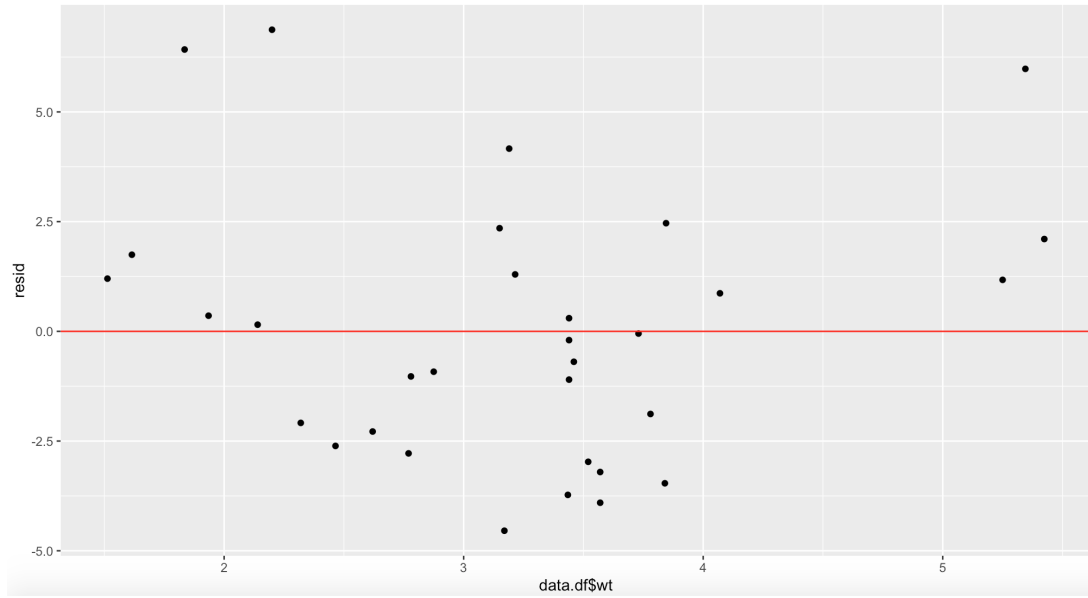
e) (2 points) Regardless of your answer to (c) interpret the slope of the regression line in the context of this problem.

f) (2 points) Regardless of your answer to (c) interpret the y-intercept of the regression line in the context of this problem.

g) (4 points) Calculate the R^2 value for this linear regression and interpret it in the context of this problem.

Bonus: (2 points) What is one word to describe how you feel about statistics?

Formulas

$$\text{IQR} = Q3 - Q1 \qquad \text{Range} = \max - \min$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad s = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n - 1}}$$

$$\text{Boxplot outlier fences:} \qquad Q1 - 1.5 \times \text{IQR} \qquad Q3 + 1.5 \times \text{IQR}$$

$$\hat{y} = b_0 + b_1 x$$

$$b_1 = r \left(\frac{s_y}{s_x} \right) \qquad b_0 = \bar{y} - b_1 \bar{x}$$

$$e = y - \hat{y}$$

$$r = \pm\sqrt{R^2}$$

$$r = \frac{1}{n - 1} \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}$$