

IOWA STATE UNIVERSITY

STAT 101 EXAM #1 FORM B

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) Breeds of Cats
- b) Income in terms of Socio-economic class
- c) Highest level of education achieved
- d) Spiciness of pepper in scovilles units
- e) Weight of birds in ounces

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

- a) Miles per Gallon
- b) Height
- c) Types of cake
- d) Level of education measured in years
- e) Race or Ethnicity

Problem 3. (2 points) What type of variable has labels for values?

- 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 two categorical variables?

- 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 an outlier decreases the Coefficient of Determination, or makes the linear model less accurate? (Circle all that apply)

- a) High Residual
- b) Low Leverage
- c) High Leverage
- d) Influential
- e) Non-Influential

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) If we were to view a residual plot and detect a megaphone pattern in the residuals, would a linear model be appropriate and why or why not?

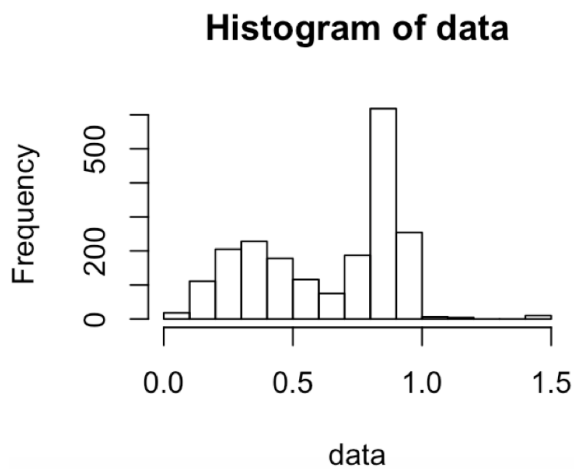
- a) No, non-constant variance is an issue
- b) No, non-linear pattern detected
- c) Yes, but caution is needed, constant non-constant variance is an issue
- d) Yes, but caution is needed, outliers may be influential
- e) Yes, no caution needed

Problem 9. (8 points) Match each statistical concept to its mathematical symbol. Each symbol is only used once

I. ρ II. b_o III. s_x IV. \bar{x} V. b_1 VI. R^2 VII. e 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 10 and 11.



Problem 10. (3 points) Describe the shape of this distribution.

Problem 11. (1 points) Which measure of spread is most appropriate for this data.

- a) Range
- b) Standard Deviation
- c) IQR

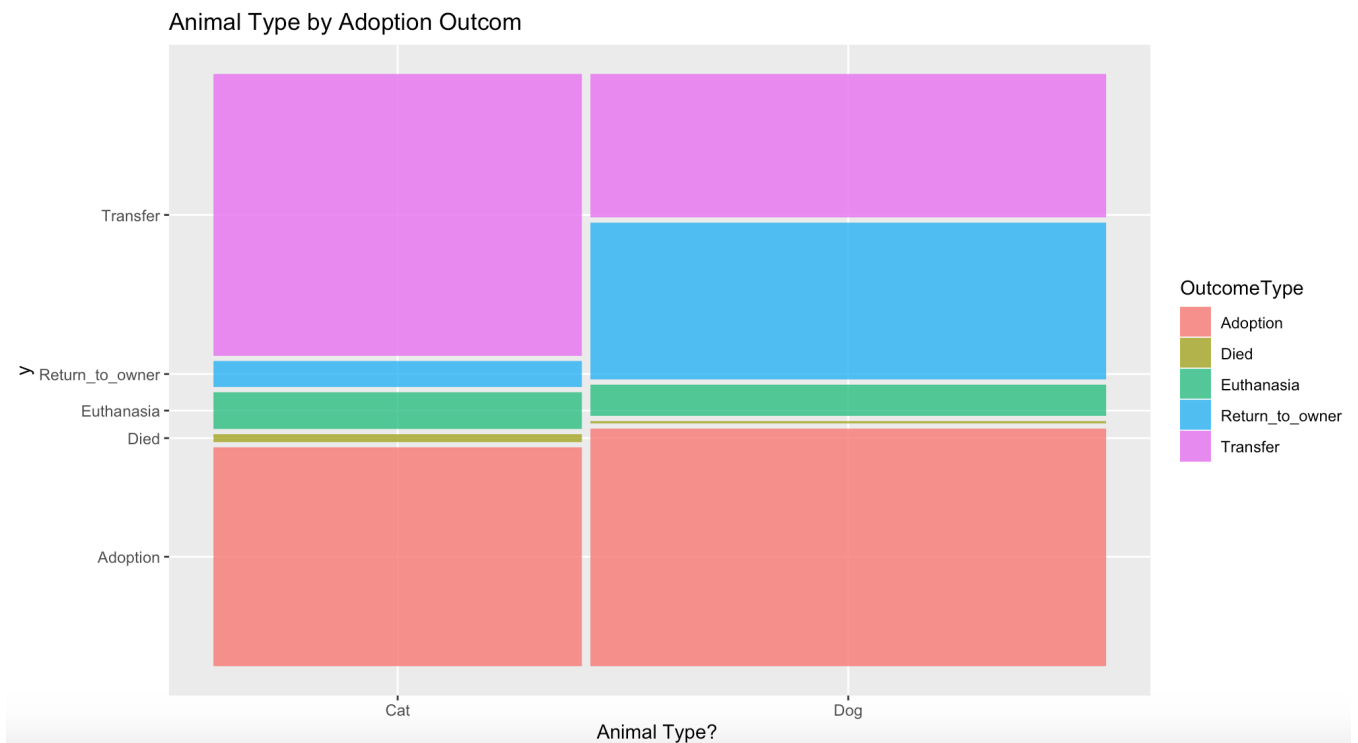
Problem 12. (15 points) **Pet Adoption** In March 2016, data was released by a humane society about pet adoptions and adoption outcomes. Data was collected from 26,729 animal outcomes, and information was recorded including adoption outcome, age at adoption, animal type, colour, and name. Use the data table and the mosaic plot provided to answer the following questions.

Age at outcome	Adoption Outcome					
	Adoption	Died	Euthanasia	Return to Owner	Transfer	Total
Baby	0	55	76	29	2,088	2,248
Young	5,740	83	293	484	3,020	9,620
Adult	5,029	59	1,184	4,273	4,298	14,843
Total	10,769	197	1,553	4,786	9,406	26,711

- a) (1 points) Identify the Who of the dataset.
- b) (1 points) Identify the What of the dataset.
- c) (1 point) In the above table, what is the explanatory variable?
- d) (2 points) What proportion of young or baby animals are adopted?

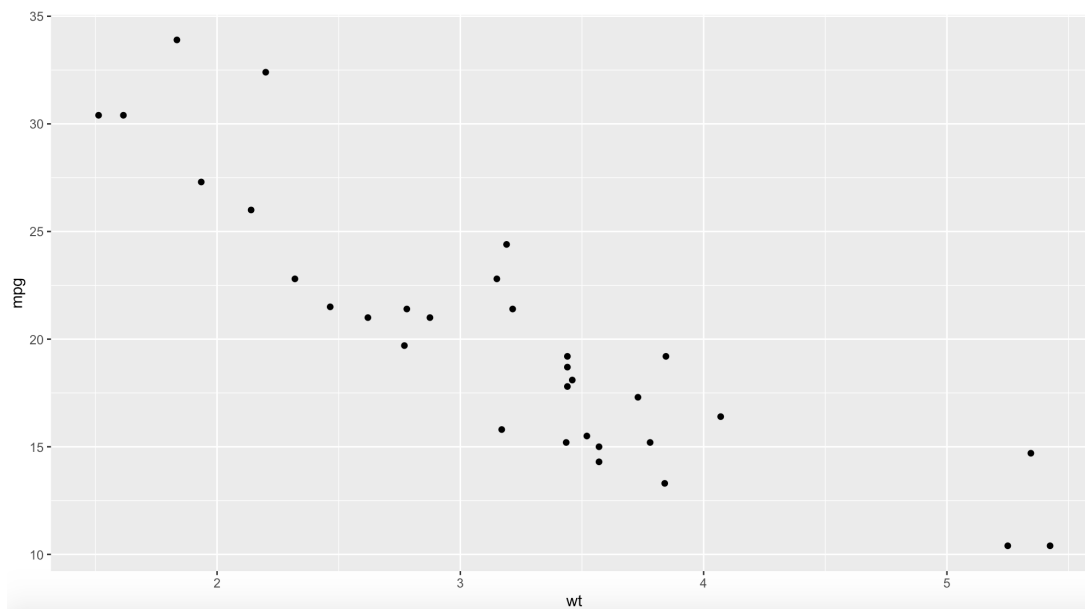
e) (2 points) Give the conditional distribution Animal Outcomes given that we are only interested in adult animals.

f) (3 points) Give the marginal distribution of Animal Outcomes. Compare this distribution to the distribution in (e)



g) (1 points) Based on the mosaic plot given, is there an association between animal type and adoption outcome?

Problem 13. (20 points) Data was extracted from the 1974 Motor Trend US magazine and comprises fuel consumption 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).



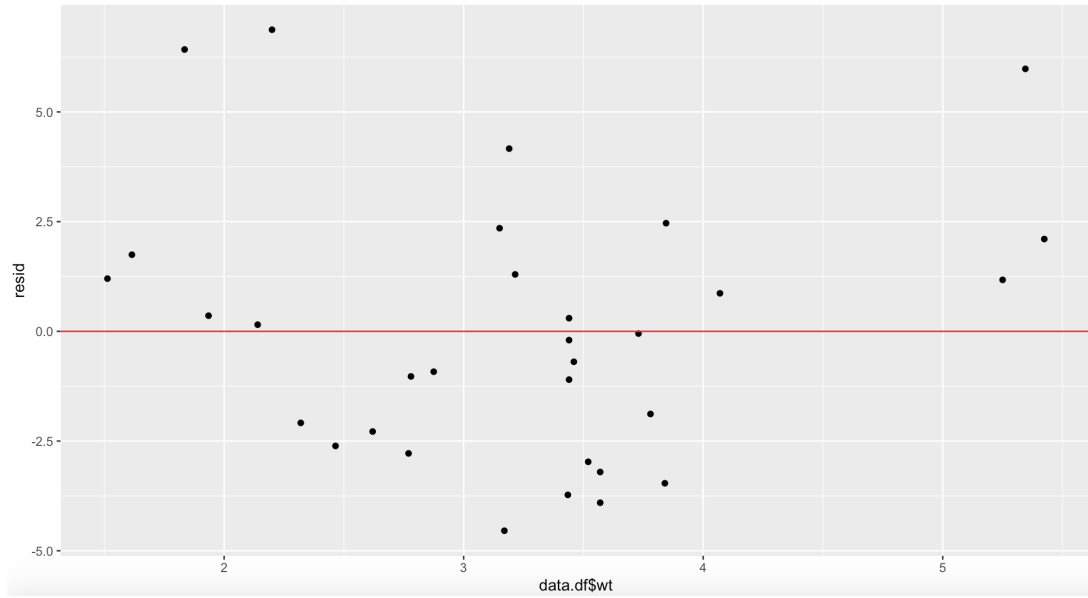
a) (3 points) Describe the relationship between weight and mpg.

- 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 weighs 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

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

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

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

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

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

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

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

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