STAT 8010 Summer 2023 Exam III

Due Date: June 23, 5:00pm via Canvas

Name:_____

Directions

1. Show your work on ALL questions. Unsupported work will NOT receive full credit.
2. Decimal answers should be exact, or to exactly 4 significant digits.
3. Please write legibly. If I cannot read your writing, NO credit will be given.
4. Put your work into a **single file** and upload it to Canvas before 5:00pm.
5. **Please work on these questions independently. Collaborating with another student will be considered a violation of academic integrity.**

Problem	Points Possible	Points Earned
1	35	
2	15	
3	25	
4	25	
Total	100	

Problem 1

(5 points for each question.) Consider the Maximum Heart Rate (`MHR`) vs. `Age` example discussed in class. Suppose a researcher conducted another study and would like to perform a simple linear regression analysis. Use the data set, `2020_Exam3_Q5.txt`, to answer the following questions:

(a) Write down the linear regression equation and compute the fitted value when Age is 35.

(b) Construct a 95% confidence interval for β_1, the regression slope.

(c) Use the resulting CI from (b) to perform the following test at $\alpha = 0.05$:

$$H_0 : \beta_1 = -1 \text{ vs. } H_a : \beta_1 \neq -1$$

(d) Construct a 95% confidence interval for E[MHR|Age = 35].

(e) Construct a 95% prediction interval for a new patient given that his/her age is 35-year-old.

(f) Compute and interpret R^2 for the simple linear regression mode and r between `MHR` and `Age`.

(g) Would it be a good idea to predict `MHR` given that `Age` = 105 using the fitted simple linear regression model? Explain your answer.

Problem 2

(5 points for each question.) A new genetic treatment of 200 patients with a particular type of cancer resulted in 60 patients surviving at least 5 years after treatment.

(a) Estimate the proportion p of all patients who would survive at least 5 years after being administered this treatment and construct a 95% CI for p.

(b) Suppose in another study, the same genetic treatment of 500 patients with a particular type of cancer resulted in 150 patients surviving at least 5 years after treatment. Would it change the point estimate for p? Would the width of 95% CI increase or decrease?

(c) Determine the minimum sample size required such that the width of 95% CI for p (assuming $p = 0.3$) is 0.05. Would the required sample size increase or decrease if we assume $p = 0.5$?

Problem 3

(5 points for each question.) Use the data set `Q2.csv` to study the relationship between speeding violations (Yes/NO) and texting while driving (Yes/NO)

(a) Construct a contingency table to summarize the data

(b) Let $p_1 = \mathbb{P}(\text{speeding violation}|\text{texting while driving})$ and $p_2 = \mathbb{P}(\text{speeding violation}|\text{No texting while driving})$. Perform a hypothesis test with $H_0 : p_1 = p_2$ vs. $H_a : p_1 \neq p_2$ with $\alpha = 0.01$.

(c) Compute the expected frequency for each cell (e.g., speeding violation and texting while driving) under the assumption that there is no association between these two variables.

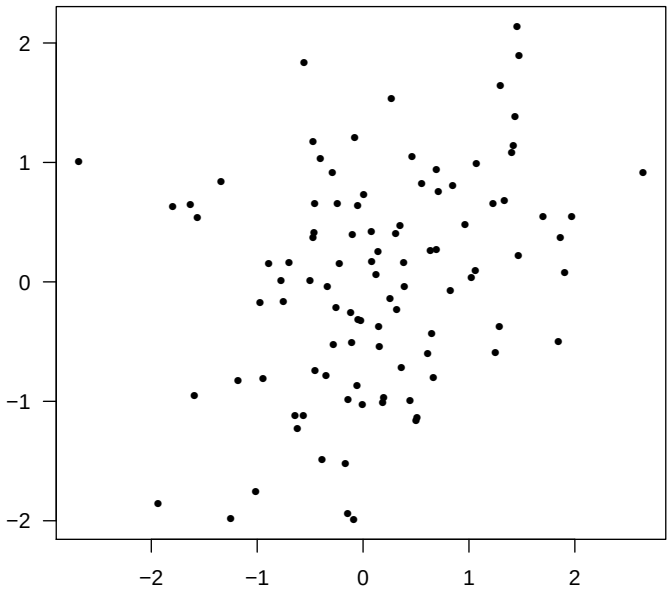
(d) Conduct a χ^2 test for independence using $\alpha = .01$.

(e) Comment the relationship between the tests in (b) and (d).

Problem 4

(5 points for each question.)

(a) What is the possible value of correlation in the figure below?



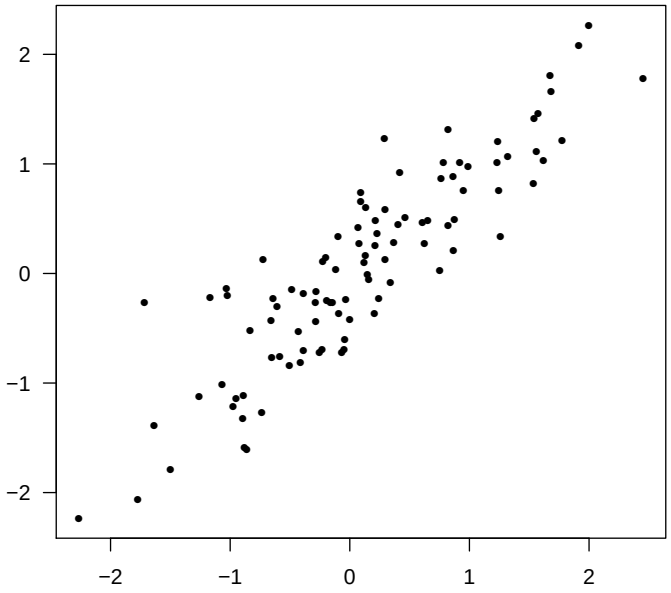
A : 0.89

B : 1.5

C : -0.7

D : 0.30

(b) What is the possible value of correlation in the figure below?



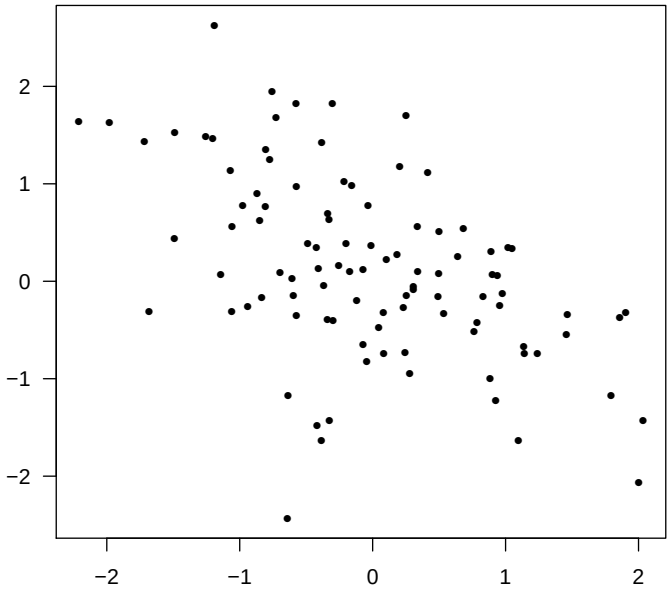
A : 1.1

B : 0.89

C : -0.7

D : -1.2

(c) What is the possible value of correlation in the figure below?



A : -1.0

B : 0.50

C : -0.50

D : 0.89

(d) Compute R^2 of a simple linear regression: $y = \beta_0 + \beta_1 x + \varepsilon$ based on the ANOVA table below

```
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
x          1 115.859 115.859  215.32 < 2.2e-16 ***
Residuals 98  52.732   0.538
```

(e) Compute r, the correlation between x and y based on the ANOVA table above and the scatterplot below.