

Instructions: This exam is in two parts: Part I is to be completed partly at home using the materials posted on Blackboard for Part I and you will answer questions about that work in class below; Part II is to be completed entirely in class. You may not use cell phones, and you may only access internet resources you are specifically directed to use. You may access your data file for Part I of the exam in Blackboard. You may access the data files posted to Blackboard for the Exam part II. Be sure you are using the data file that matches the exam version you are given.

Part I:

1. Describe what you see in the comparative box plots. Are they about the same? Do any seem different than the others? Is the spread about the same? (15 points)

The means/medians and spread are similar.
0 children has the highest median and 2 children the smallest spread (IQR).

2. Consider the scatterplot of money spent vs. number of children. Write the equation and the R^2 value. (8 points)

$$y = -203.27x + 1406.6$$

$$R^2 = .0494$$

3. Which of three scatterplots appears to show the strongest linear relationship? Why? (8 points)

Salary vs. Amount Spent
 R^2 is highest for this model

4. Using the pivot table you created does homeownership or marital status appear to affect the average amount of money spent? Explain. (15 points)

yes. unmarried people (0) spend about half of what married (1) people do.

Calculations in Excel: (1) 25 points, (2) 28 points, (3) 12 points.

Part II:

5. Create a Pivot Table from the data in the Excel file comparing Pay Type and Dwell Type. Use it to answer the following questions about a randomly selected person from the dataset:
- What is the probability the person lives in a condo? (6 points)

$$239/856 = 27.92\%$$

- What is the probability that the person is paid hourly? (6 points)

$$375/856 = 43.80\%$$

- What is the probability that the person lives in a condo given that they are paid hourly? (6 points)

$$97/375 = 25.87\%$$

- What is the probability that the person both lives in a condo and is paid hourly? (6 points)

$$97/856 = 11.33\%$$

- What is the probability that the person either lives in a condo or is paid hourly? (6 points)

$$(375 + 239 - 97) / 856 = 60.40\%$$

- What is the probability that someone does not live in a condo? (6 points)

$$1 - 239/856 = 72.08\%$$

- Are the variables Dwell Type and Pay Type independent? Why or why not? (10 points)

they are dependent since $P(\text{condo})$ and $P(\text{condo} | \text{hourly})$ are not equal

- h. Create a stacked column graph of the data. What do you notice overall? (10 points)

hourly workers are more likely to own homes

answers may vary

6. Create a scatterplot of the data in the Excel file to compare the Federal Funds rate vs. Prime Rate. Explain the meaning of the slope of the equation of the regression line you find in context. (10 points)

for each increase of 1% in the federal funds rate, one can expect the prime rate to go up by 0.9%.

7. Using the same scatterplot, explain what the R^2 value means in context. (8 points)

$$R^2 = 0.9797$$

roughly 98% of the change in the prime rate can be explained by the relationship to the federal funds rate

8. What is the value of the correlation (r)? (5 points)

$$r = 0.989791$$

9. Create a Pivot Table comparing Pay Type vs. Live Alone. In the body of the table, report Average amount of Credit Card Debt. Create a graph of the data. What do you see? Be thorough. (12 points)

hourly workers have a bit less CC debt (\$160) but living alone has very little effect on debt levels.

10. Explain the difference between a classical (theoretical) probability and an experimental (observational) probability. (8 points)

classical probability assumes all outcomes are equally likely and calculates probability from portions of all possible events. Experimental probability is based on repeated observations of the same event.

11. A probability distribution is provided in the Excel file. Calculate the following:
- Find the expected value of the probability distribution. (6 points)

4.54

- Calculate the variance of the probability distribution. (8 points)

4.94

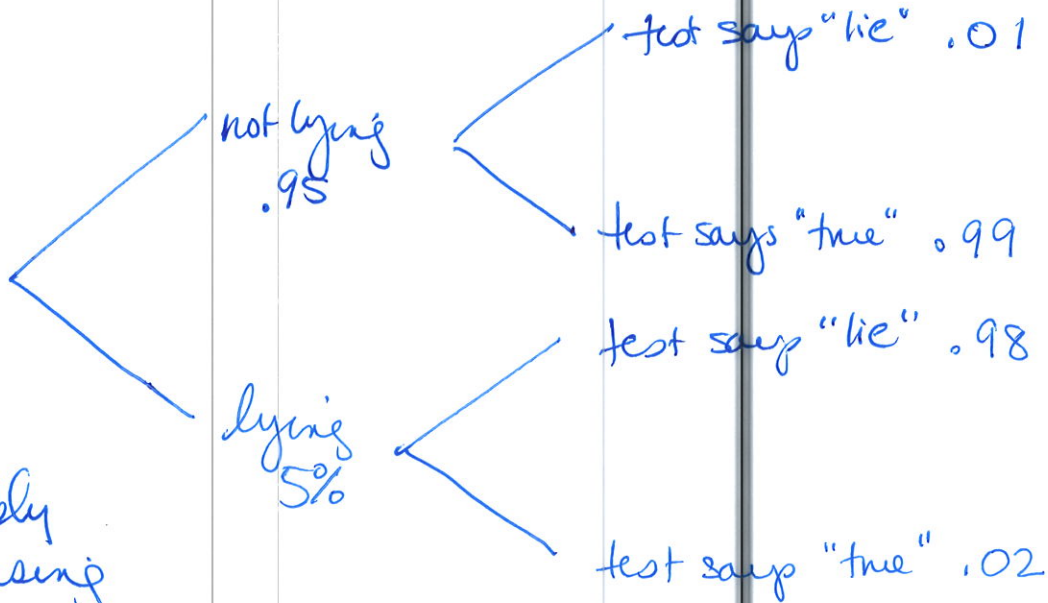
- What is the standard deviation of the distribution? (4 points)

2.22

12. Create a simulation in Excel using the Rand or Randbetween function to model of die roll for 500 rolls. Use the COUNTIF Function to construct a probability distribution for the simulation? How close is it to fair dice model? (20 points)

it stays pretty close to $\approx 16-17\%$ can
can vary by a couple percent

13. A particular model of lie detector test has a 98% probability of correctly detecting someone who is lying, and 99% probability of correctly detecting someone who is not lying. A human resources office interviews using the lie detector, hoping to catch the 5% of interviewees they believe to be lying. If someone tests positively on the lie detector for lying, what is the probability that the person is actually lying? Construct a tree diagram to model the situation. Should the company continue to use the lie detector? (12 points)



$$\frac{.05 * .98}{.05 * .98 + .95 * .01} = 83.76\%$$

they probably can keep using it since most of the people thought to be lying actually are.

Upload your completed Excel files to the Exam #2 submission box in Blackboard, and submit your completed paper exam to your instructor. You may not modify anything once the exam is submitted.

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

$$\mu = E(X) = \sum x_i p(x_i)$$

$$\sigma^2 = \text{Var}(X) = \sum (x_i - \mu)^2 p(x_i)$$