MTH 324, Final Exam, Fall 2022 Name _____

Instructions: Answer each question thoroughly. For questions in Part 1, use the work you did at home to answer the questions. Be sure to answer each part of each question. In Part 2, report exact answers unless directed to round.

Part I:
Use the work you did at home to answer these questions about tax paid and the neighborhoods in our dataset.

1. Based on the data from sheet 1 on ad preferences, is the data distributed normally or approximately so? Explain.

2. Report the results of your paired test of the ad preference data. Which test did you use and why? Clearly state your hypotheses, the conclusion in the context of the problem, and explain why you came to that conclusion.

3. Based on the graphs of second data set, which of the numerical variables was most symmetric or normal?

4. Based on the graphs of the second data set, which of the categorical variables had the biggest difference in frequency?

5. Based on the numerical summary of Income, provide the 5-number summary.

6. Describe the results of your test of Independence for Dwell Type and Neighborhood. State your hypotheses, your conclusion in the context of the problem and explain your reasoning.

7. Describe the results of your ANOVA test for Dwell Type and Age. State your hypotheses, your conclusion in the context of the problem and explain your reasoning. Does your result agree with your boxplot?

8. Give your confidence interval from your bootstrap sample of Credit Card Debt. Give your confidence interval (95%) for the one-sample t- or z-interval. How do they compare?

9. Describe the results of your test of two proportions for Live Alone by Gender. State your hypotheses, your conclusion in the context of the problem and explain your reasoning.

10. Based on the table in problem #3 on the at-home portion, describe the results of your goodness-of-fit test. State your hypotheses, your conclusion in the context of the problem and explain your reasoning.

Part II:

11. If you needed to create a stratified sample in R (let's say on Gender), explain how you would go about doing that. (I don't need the code, but explain your steps in words.)

12. A two-way table of Dwell Type and Neighborhood is shown below. Use it to answer the following questions.

	East	South	West	Grand Total
Apt	69	50	94	213
Condo	80	57	102	239
Home	128	100	176	404
Grand Total	277	207	372	856

a. What is the probability that a random person selected from this data set is from the East neighborhood?

b. What is the probability that a random person selected from this data set lives in a Condo?

c. What is the probability that a random person selected from this data set is from the East neighborhood and lives in a Condo?

d. What is the probability that a random person selected from this data set is from the East neighborhood or lives in a Condo?

e. What is the probability that a random person selected from this data set is from the East neighborhood given that they live in a condo?

f. Are the variables Neighborhood and Dwell Type independent or dependent? Does your answer differ if you consider only the descriptive properties of the table, or if you infer the answer from the hypothesis you conducted in Part 1? If it does, explain why.

13. The proportion of women in the sample on sheet 2 of the data set from the at-home portion of the exam is 0.465. If we were to randomly select 15 subjects from that data set, answer the following questions about the probability of possible outcomes.
    a. What is the probability of getting exactly 8 women in the sample?

    b. What is the probability of having fewer than 3 women in the sample?

    c. What is the expected number of women in the sample?

14. Suppose that 1% of people have a certain genetic defect. Further suppose that 90% of tests for the gene detect the defect (true positives), and 9.6% of the tests are false positives. If a person gets a positive test result, what is the probability they actually have the genetic defect?

15. Consider the probability density function $f(x) = \frac{x^3}{5000}(10 - x), 0 \leq x \leq 10$ (it is equal to 0 everywhere else). Use this information to answer the questions that follow.
    a. Verify that this function represents a valid probability distribution.

    b. Find $P(1 \leq X \leq 4)$

    c. Find the mean (expected value) of the distribution.