MTH 324, Exam #1, At Home, Fall 2022 Name KEY

Instructions: This exam is in two parts: Part I is to be completed partly at home using the materials posted in the course for the at-home portion and you will answer questions about that work during the in-class portion of the exam; 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.

At home, prepare for questions in Part I using R. Open the data file entitled **324exam1data.xlsx** posted in Blackboard. Complete the calculations noted below. You will be asked for additional analysis and interpretation of this data in the in-class portion of the test. Print out the results of your analysis and code, and bring the pages with you to the exam. You will submit all this work along with the in-class exam.

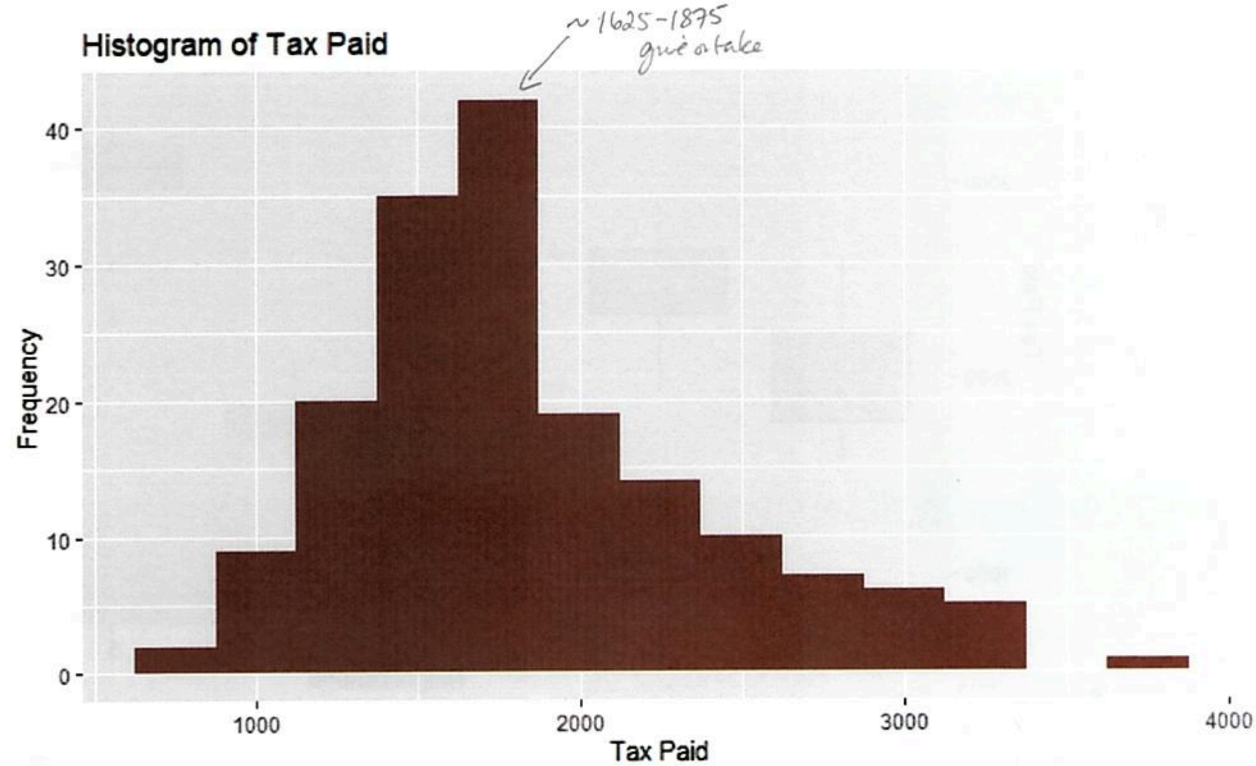
The data represents taxes paid in various neighborhoods in a community. Create the following graphs:

1. A histogram of "Tax Paid". Your histogram should have a bin width of 200-300. Label the graph appropriately.
2. Create a bar graph of neighborhoods (counts). Label the graph appropriately.
3. Create a comparative boxplot of tax paid by neighborhood. Label the graph appropriately.
4. Calculate a set of descriptive statistics for Tax Paid. Be sure to have enough information to identify any extreme values.
5. Create a frequency table of Neighborhood.
6. Find the indicated probabilities.
   a. A particular assembly line produces working computers 99% of the time and computers with malfunctions 1% of time. A sample of 10 computers is sent to quality control. What is the probability of having a sample with no malfunctions?

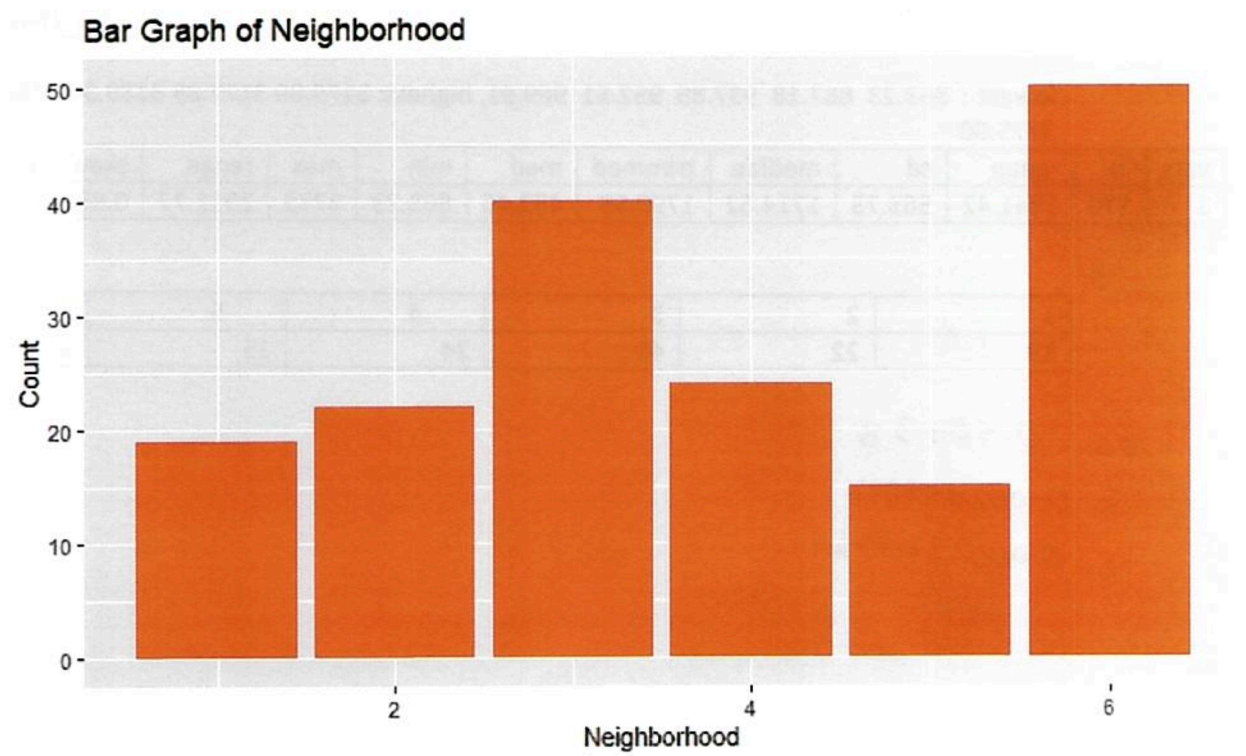
   b. A security check line at a particular airport sees 100 travelers pass through during a particular hour of the day. Determine the probability that the check line will see 30 or more passengers in the next 10 minutes?

   c. The weight of a particular colony of feral cats has a mean of 7.8 pounds and a standard deviation of 0.6 pounds. What is the probability that a cat in the colony will weigh more than 10 pounds?

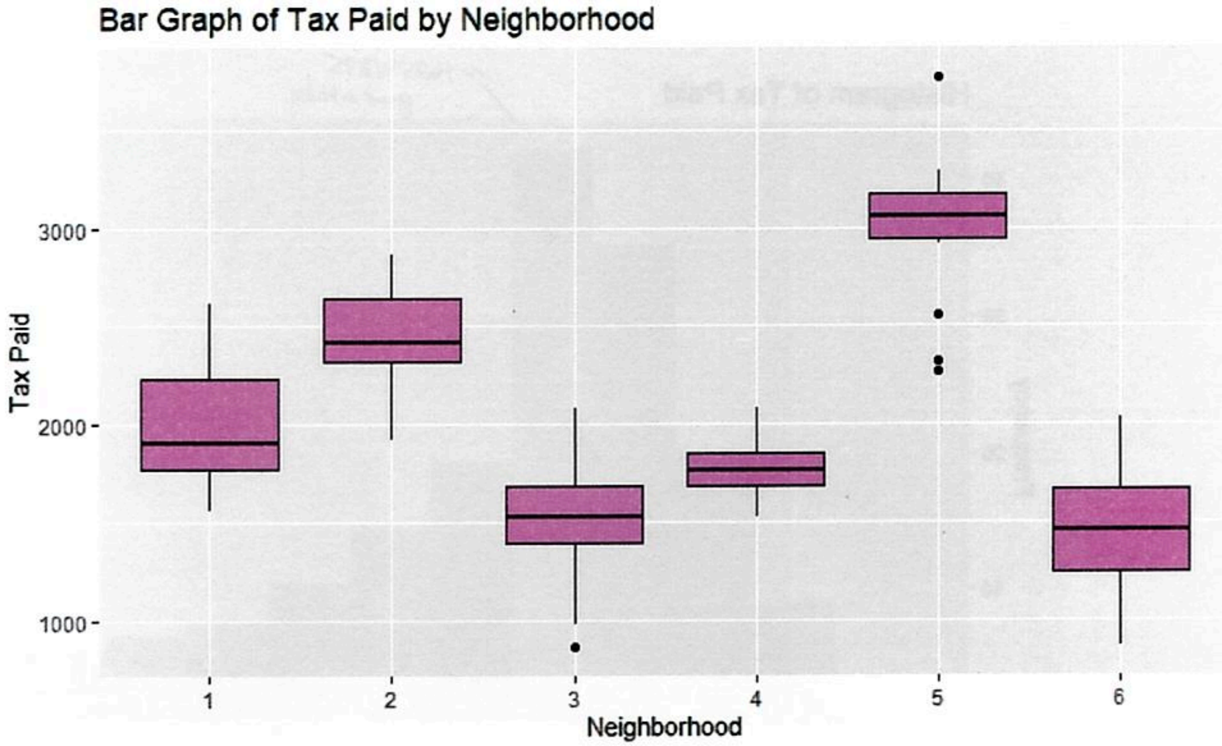
Output of Analyses for MTH 324 Exam #1 At-home portion.



1.



2.



3.

4. IQR= 632.3225

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
170	0	170	1	1841	618.1	1074	1189	1499	1715	2131	2649	3025

lowest : 863.23 867.18 937.85 952.51 969.61, highest: 3175.09 3177.85 3189.34 3288.01 3775.00

vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
1	170	1841.42	563.73	1714.62	1790.58	432.46	863.23	3775	2911.77	0.85	0.45	43.24

5.

1	2	3	4	5	6
19	22	40	24	15	50

6a. 0.90438

b. 0.002058046

c. 0.0001228664

MTH 324, Exam #1, Fall 2022 Name KEY

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. What is the modal class on your histogram? (3)

   approx 1625–1875

   depending on bin width, answers may vary slightly

2. What is the shape of the Tax Paid histogram? Roughly symmetric, right-skewed, left-skewed? (3)

   right-skewed

3. Which neighborhood has the fewest members in the sample? How many homes is that? (3)

   #5 ~ 15 homes

4. Based on your comparative box plot, which neighborhood appears to be the most different from the other neighborhoods? Explain why you think so. (3)

   neighborhood 5 – highest median, outliers on both ends

5. Calculate the upper and lower fences for the Tax Paid data, and the extreme upper and lower fences. Are there any outliers in the data? Are there any extreme outliers in the data? (4)

   Q1 = 1499 IQR = 632
   Q3 = 2131

   there are outliers, none on the low end, but at least 5 on the high end.
   no extreme outliers

   lower fence = 1499 − 1.5(632) = 551

   upper fence = 2131 + 1.5(632) = 3079

   extreme lower fence = 1499 − 3(632) = −397 not possible

   extreme upper fence = 2131 + 3(632) = 4027

6. Find the indicated probabilities.
   a. A particular assembly line produces working computers 99% of the time and computers with malfunctions 1% of time. A sample of 10 computers is sent to quality control. What is the probability of having a sample with no malfunctions? (3)

      0.90438

   b. A security check line at a particular airport sees 100 travelers pass through during a particular hour of the day. Determine the probability that the check line will see 30 or more passengers in the next 10 minutes? (3)

      0.002058046

   c. The weight of a particular colony of feral cats has a mean of 7.8 pounds and a standard deviation of 0.6 pounds. What is the probability that a cat in the colony will weigh more than 10 pounds? (3)

      0.0001228664

Part II:

7. Describe the procedure for calculating a cluster sample, and a stratified sample. Highlight how they differ from each other. Describe a situation in which each method is used. (4)

A cluster sample divides the population into many small groups and groups are randomly selected. A stratified sample divides the population into a small # of groups and each group has a random selection drawn from it. Cluster sampling often done for geographic regions; stratified done to ensure all demographic groups are included.

8. What is the purpose of doing a block design in an experiment? (3)

to control for a confounding variable - for instance to understand how gender might influence the effect of a drug.

9. Why are medical studies often double blind? (3)

So that those administering medication or placebo do not give away which the patient is receiving through unconscious body language

10. For each of the following variables, identify whether the variable is i) categorical or numerical, ii) it's level of measurement (nominal, ordinal, interval or ratio), and if it is numerical iii) whether it is discrete or continuous (write NA if it does not apply).
   a. Favorite color (3)

   categorical, nominal, NA

   b. Level of pain on a scale of 1 to 10 (3)

   ordinal, categorical, NA

   c. Football jersey number

   categorical, nominal, NA (3)

   d. Number of books in a library

   numerical, ratio, discrete (3)

   e. SAT score

   numerical, interval, discrete (will accept continuous) (3)

11. Use the contingency table below to answer the probability questions that follow.

		Sport Preference			
		Archery	Boxing	Cycling	
Gender	Female	35	15	50	**100**
	Male	10	30	60	**100**
		45	**45**	**110**	**200**

   a. What is the probability that someone selected randomly from this sample prefers boxing? (3)

   $45/200 = 9/40$

   b. What is the probability that someone selected randomly from this sample is female? (3)

   $100/200 = 1/2$

c. What is the probability that someone selected randomly from this sample is a female and prefers boxing? (3)

$$15/200 = 3/40$$

d. What is the probability that someone selected randomly from this sample is a female or prefer boxing? (3)

$$\frac{45+100-15}{200} = \frac{130}{200} = \frac{13}{20}$$

e. What is the probability that someone selected randomly from this sample is a female given that they prefer boxing? (3)

$$15/45 = \frac{1}{3}$$

f. What is the probability that someone selected randomly from this sample does not prefer boxing? (3)

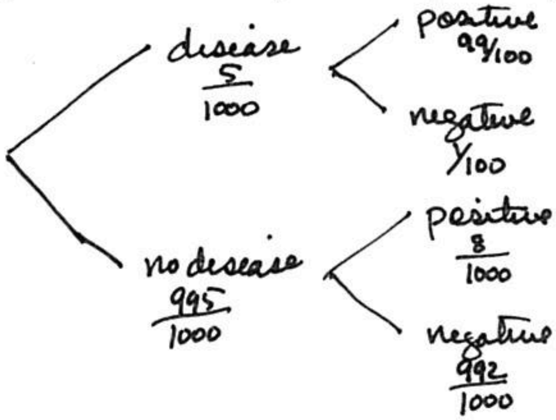
$$1 - \frac{45}{200} = \frac{31}{40}$$

g. Are sports preference and gender independent events? Explain your reasoning. Show math to support your conclusion. (3)

They are dependent

$P(F) = 1/2$ $\quad P(F|B) = 1/3$ they are not equal

12. A particular rare disease occurs in just 5 out of 1000 people in population. A test for that disease correctly identifies those with the disease 99% of the time. For people without the disease, the test correctly identifies that they do not have the disease 99.2% of the time. If a patient tests positive for the disease, what is the probability that they actually have the disease? (6)



$$\frac{\frac{5}{1000} * \frac{99}{100}}{\frac{5}{1000} * \frac{99}{100} + \frac{995}{1000} * \frac{8}{1000}} = 0.3834$$

13. Consider the probability distribution given by $\int_{-1}^{1} K(x^4)dx$.

(3)

a. Find the value of K that makes this a valid probability distribution.

$$\int_{-1}^{1} Kx^4 dx = \frac{K}{5}x^5\Big|_{-1}^{1} = \frac{K}{5}(1^5 - (-1)^5) = \frac{K}{5}(2) = 1 \quad K = \frac{5}{2}$$

b. Find the probability that $P(0.5 \le X \le 1)$.

$$\int_{0.5}^{1} \frac{5}{2}x^4\,dx = \frac{1}{2}x^5\Big|_{0.5}^{1} = \frac{1}{2}\left(1^5 - \frac{1}{32}\right) = \frac{1}{2}\left(\frac{31}{32}\right) = \frac{31}{64}$$ (3)

$= 0.484$

c. Find the mean of the distribution.

$$\int_{-1}^{1} x\,\frac{5}{2}x^4\,dx = \int_{-1}^{1} \frac{5}{2}x^5\,dx = 0$$ (odd function) (3)

$= E(x)$

d. Find the variance of the distribution.

$$E(x^2) = \int_{-1}^{1} x^2\left(\frac{5}{2}x^4\right)dx = \int_{-1}^{1}\frac{5}{2}x^6\,dx = \frac{5}{14}x^7\Big|_{-1}^{1} = \frac{5}{14}(1-(-1)) = \frac{5}{7}$$ (3)

$$\frac{5}{7} - 0^2 = \frac{5}{7} = V(x)$$

e. What value of X represents the 90th percentile?

$$0.9 = \int_{-1}^{t}\frac{5}{2}x^4\,dx = \frac{1}{2}(t^5 - (-1)^5) = \frac{1}{2}(t^5+1) = .9$$

(3)

$$t^5 + 1 = 1.8 \rightarrow t^5 = 0.8 \qquad t = 0.95635$$

14. Use the joint probability table below to answer the questions that follow.

		x_2			
		0	1	2	3
	0	.08	.07	.04	.00
	1	.06	.15	.05	.04
x_1	2	.05	.04	.10	.06
	3	.00	.03	.04	.07
	4	.00	.01	.05	.06

(3)

a. What is $P(X_1 = 2, X_2 = 3)$.

0.06

b. What is $P(X_1 > X_2)$ (4)

0.34

c. Find the marginal probability distributions of X_1 and X_2. (4)

X_1	0	1	2	3	4
$P(X_1)$	0.19	0.3	.25	0.14	0.12

X_2	0	1	2	3
$P(X_2)$	.19	0.3	.28	.23

d. Find the mean of X_2. (3)

$0(0.19) + 1(0.3) + 2(.28) + 3(0.23) = 1.55$

e. Find $P(X_1 = 2 | X_2 = 1)$ (3)

$$\frac{P(X_1 = 2 \text{ AND } X_2 = 1)}{P(X_2 = 1)} = \frac{0.04}{0.3} = 0.1\overline{3} = 2/15$$