MTH 324, Exam #1, At Home, Fall 2023 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.

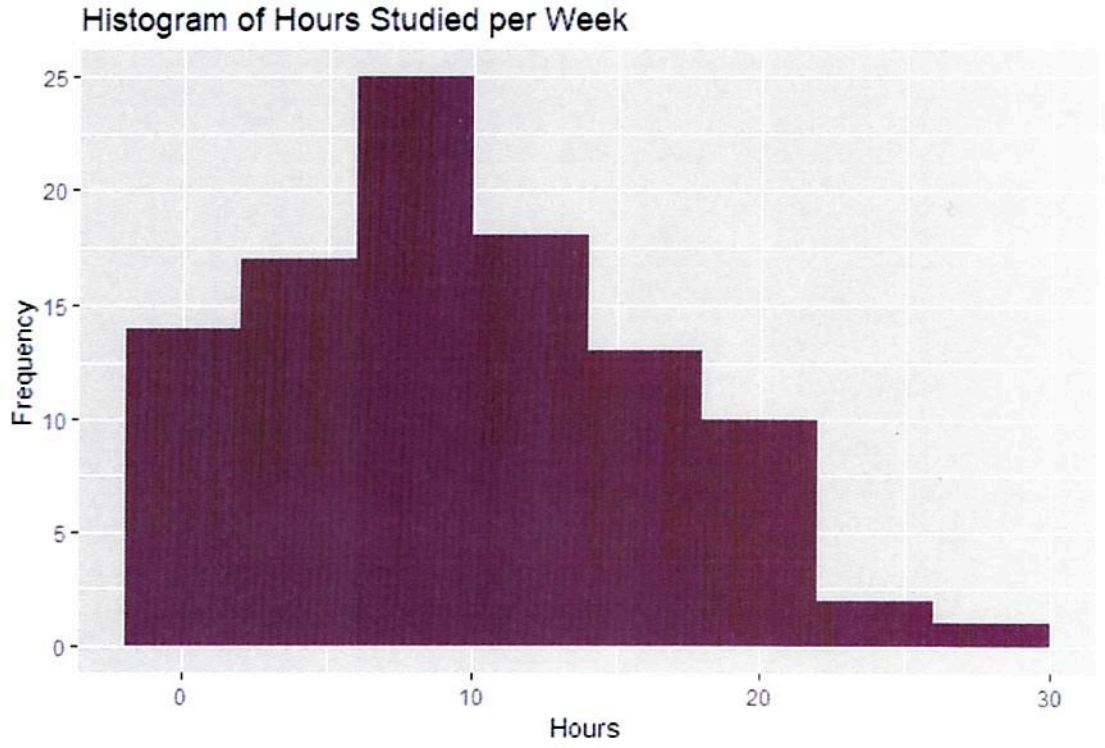
The data represents the number of hours studied for an exam by students in a lecture course. Create the following graphs:

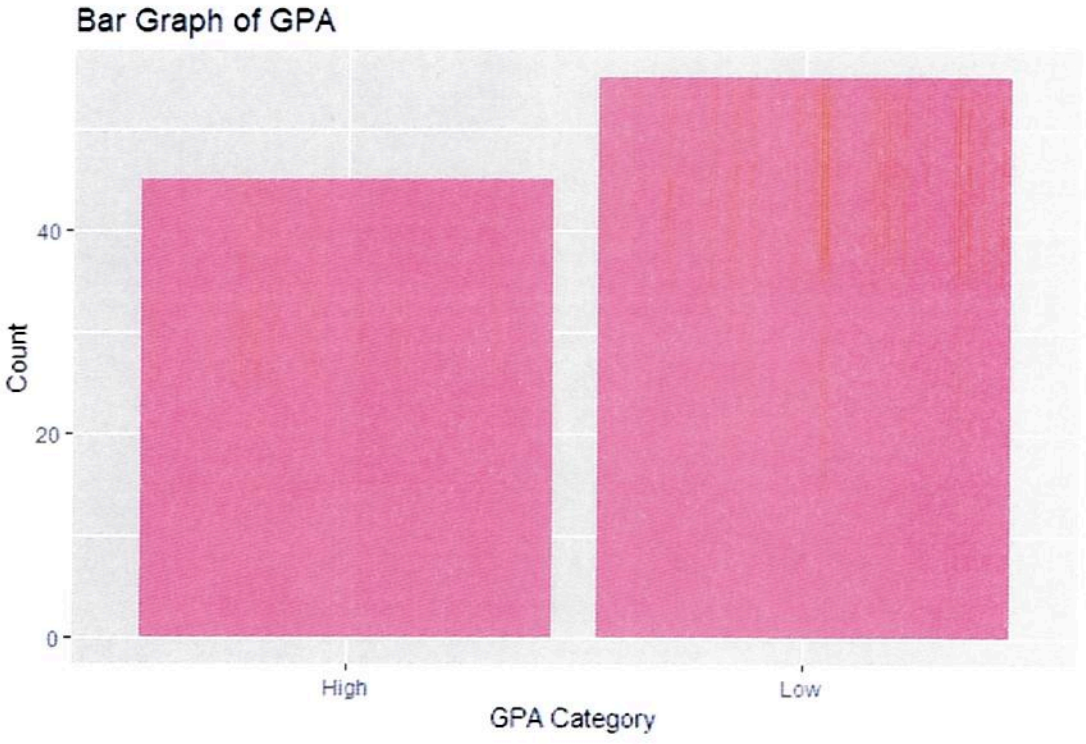
1. A histogram of "Hours Per Week". Your histogram should have a bin width of ~~1~~ 3-4. Label the graph appropriately.
2. Create a bar graph of GPA category (counts). Label the graph appropriately.
3. Create a comparative boxplot of Hours Per Week by GPA category. Label the graph appropriately.
4. Calculate a set of descriptive statistics for Hours Per Week. Be sure to have enough information to identify any extreme values.
5. Create a frequency table of GPA.
6. Find the indicated probabilities.
   a. Based on previous experience, it is found that 75% of people getting into a driving simulator put on their seatbelts. What is the probability that among the next 15 people that get into the simulator, more than 12 of them will put on their seatbelts?

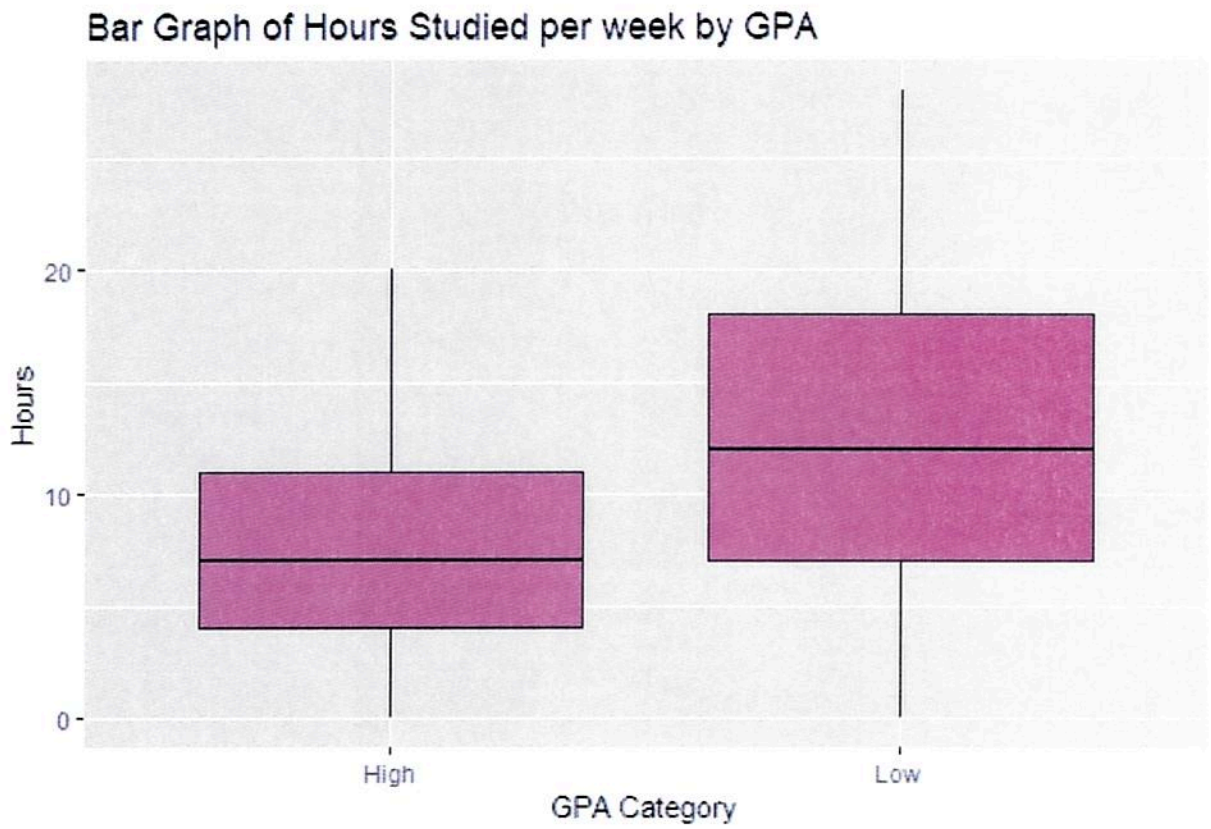
   b. A drive thru at a particular bank sees 15 customers pass through during a particular hour of the day. Determine the probability that the drive thru will see 5 or more passengers in the next 20 minutes?

   c. The height women has a mean of 64.3" and a standard deviation of 3.1". What is the probability that a randomly selected woman will stand more than 71" tall?

MTH 324, Fall 2023, Exam #1 At-home Analysis








```
data1$`Hours per week`
       n  missing distinct     Info     Mean      Gmd      .05      .10      .25
     100        0       23    0.996    10.12    7.551      0.0      1.9      5.0
     .50      .75      .90      .95
     9.0     15.0     20.0     22.0

lowest :  0  1  2  4  5, highest: 19 20 22 23 28
```

IQR=10

```
GPA
High  Low
  45   55
```

6a. 0.2360878
6b. 0.5595067
6c. 0.01533646

MTH 324, Exam #1, Fall 2023 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 study hours and GPA in our dataset.

1. What is the modal class on your histogram?

   7-10

2. What is the shape of the Hours Per Week histogram? Roughly symmetric, right-skewed, left-skewed?

   right-skewed

3. Which GPA level has the fewest members in the sample? How many students is that?

   high, 45

4. Based on your comparative box plot, which GPA group appears to study the least? Does this make sense?

   high

   it does seem somewhat counter intuitive

5. Calculate the upper and lower fences for the Hours Per Week data, and the extreme upper and lower fences. Are there any outliers in the data? Are there any extreme outliers in the data?

   IQR = 10 $\quad$ 1.5×10 = 15 $\quad$ 3×10 = 30

   Q1 = 5 $\quad$ 5 − 15 = −10 lower fence

   Q3 = 15 $\quad$ 15 + 15 = 30 upper fence

   Q1 & EF $\quad$ 5 − 30 = −25 $\quad$ 15 + 30 = 45 extreme upper/lower fences

   there are no outliers in the data

6. Find the indicated probabilities.
   a. Based on previous experience, it is found that 75% of people getting into a driving simulator put on their seatbelts. What is the probability that among the next 15 people that get into the simulator, more than 12 of them will put on their seatbelts?

0.2360878

   b. A drive thru at a particular bank sees 15 customers pass through during a particular hour of the day. Determine the probability that the drive thru will see 5 or more passengers in the next 20 minutes?

0.5595067

   c. The height women has a mean of 64.3" and a standard deviation of 3.1". What is the probability that a randomly selected woman will stand more than 71" tall?

0.01533646

Part II:

7. Describe the procedure for calculating a simple random sample, and a systematic sample. Highlight how they differ from each other. Describe a situation in which each method is used.

Phone survey

in a simple random sample, number the population (sampling frame) draw uniformly distributed random #s from list.

in a line

a systematic sample take pop size N and divides by sample size n to get step size in sampling frame $k = N/n$. Then choose a starting value between 1 and k, and then each sample after is prev # plus k.

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

to see how subsets of the population may behave differently

9. Why are IRBs (Institutional Review Boards) involved in research on human subjects? What is their purpose?

they are there to ensure that studies on people are conducted ethically

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. Hair color nominal

Categorical, NA

b. Richter scale

numerical, interval, continuous

c. PIN number

Categorical, nominal, NA

d. Number of students in a class

numerical, ratio, discrete

e. Weight in grams

numerical, ratio, continuous

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 cycling?

$110/200 = 11/20$

b. What is the probability that someone selected randomly from this sample is male?

$100/200 = 1/2$

c. What is the probability that someone selected randomly from this sample is a male and prefers cycling?

$60/200 = \frac{3}{10}$

d. What is the probability that someone selected randomly from this sample is a male or prefers cycling?

$$\frac{100+110-60}{200} = \frac{150}{200} = \frac{3}{4}$$

e. What is the probability that someone selected randomly from this sample is a male given that they prefer cycling?

$60/110 = \frac{6}{11}$

f. What is the probability that someone selected randomly from this sample does not prefer cycling?

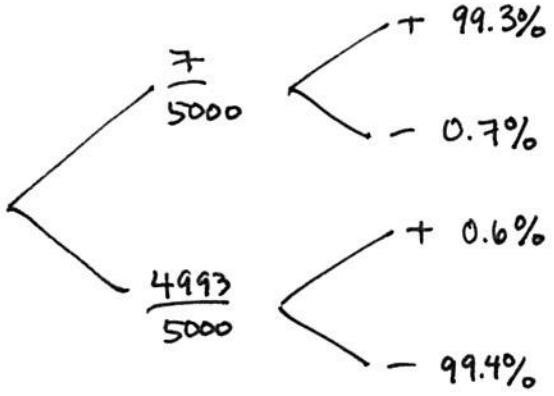
$90/200 = \frac{9}{20}$

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

$P(\text{male}) = 1/2$ these are not equal

$P(\text{male} \mid \text{cycling}) = 6/11$ so dependent

12. A particular rare disease occurs in just 7 out of 5000 people in population. A test for that disease correctly identifies those with the disease 99.3% of the time. For people without the disease, the test correctly identifies that they do not have the disease 99.4% of the time. If a patient tests positive for the disease, what is the probability that they actually have the disease?



about 19%

$$\frac{\frac{7}{5000} \times 0.993}{\frac{7}{5000} \times 0.993 + \frac{4993}{5000} \times 0.006} = 0.188328...$$

b. What is $P(X_1 \leq X_2)$

0.74

c. Find the marginal probability distributions of X_1 and X_2.

X_1	0	1	2	3	4
$P(x)$	0.19	0.3	0.25	0.14	0.12

X_2	0	1	2	3
$P(x)$	0.19	0.30	0.28	0.23

d. Find the mean of X_1.

1.7

e. Find $P(X_1 = 4|X_2 = 2)$

$$\frac{.05}{0.12} = 0.41\overline{6} = 5/12$$