STAT 3011
FALL 2022
Exam 1 (B)
Time Limit: 90 Minutes

Name (Print): ____________________________

Student ID: __________________________

Instructions:

- Do *not* begin or turn this page until you are instructed.
- Enter all requested information on the top and bottom of this page, and put your initials on the top of every page, in case the pages become separated.
- This exam contains 15 pages (including this cover page and the multiple choice answer sheet). Check to see if any pages are missing. There are 13 multiple-choice questions and 4 short answer problems.
- The exam is closed book. **Do not** use your books, or any wireless device on this exam.
- You may use a calculator and one sheet of paper (size A4 or 8.5" by 11") with formulas or other notes on both sides. **Do not** share calculators or notes!
- Show all your work on each problem for full credit except multiple choice problems. The following rules apply:
  - *Organize your work*, in a reasonably neat and coherent way, in the space provided. Work scattered all over the page without a clear ordering will receive very little credit.
  - *Mysterious or unsupported answers will not receive full credit* for short answer problems. A correct answer, unsupported by calculations, explanation, or algebraic work will not receive full credit; an incorrect answer supported by substantially correct calculations and explanations may still receive partial credit.
  - If you need more space, use the back of the pages; clearly indicate when you have done this.

Honesty Statement and Pledge:

I have not given or received any aid or assistance to or from any other student in this course during the exam period. Everything I have written on this exam represents my own work and knowledge. I sign this knowing that infringements on the University's Academic Honest policy may result in failure or expulsion.

Signed By: __________________________ **Date:** __________________________

Problem 1. (40 points) **Multiple Choice**

Choose ONLY ONE answer for each question. Circle your answers to all questions in the answer sheet provided on page 15. (NO explanation is needed).

1. (3 points) A random sample of patients who received a hip transplant operation at Stanford University Hospital in 2022 will be followed for 10 years after their operation to determine the success (or failure) of the transplant.
   Which of the following is correct?
   (A) This study is an observational study.
   (B) This study uses a volunteer sample.
   (C) Sample of this study is those who had a successful transplant operation in 2022.
   (D) None of the above are correct.

2. (3 points) Which of the following graphs is *not* useful for judging the shape of distribution for a quantitative variable?
   (A) Box plot
   (B) Bar graph
   (C) Histogram
   (D) Stem and leaf plot

3. (3 points) Which of the following is true?
   (A) Sample standard deviation can never be zero.
   (B) Mean and median can never be the same.
   (C) The mean is always one of the data points.
   (D) All of the above are false.

4. (3 points) Let A and B be independent events. It is known that $P(A) = 0.2$ and $P(A \cap B) = 0.16$. Find the probability $P(A \cup B)$.
   (A) 0.45
   (B) 0.80
   (C) 0.84
   (D) 1.00

5. (3 points) Let C and D be events such that $P(C) = 0.55, P(D) = 0.25$, and $P(C \mid D) = 0.20$. What is the probability that EXACTLY ONE of C or D occurs (but NOT BOTH)?
   (A) 0.40
   (B) 0.60
   (C) 0.75
   (D) 0.70

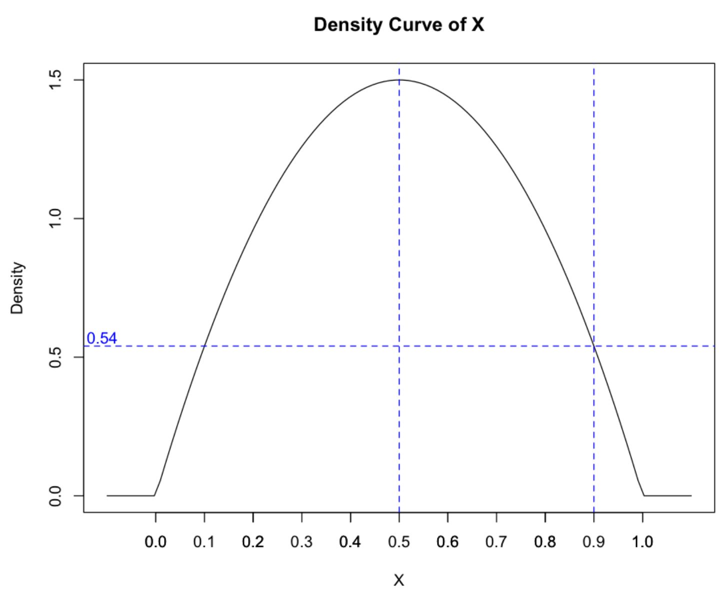
6. (3 points) The probability distribution of a random variable X is provided in the following table. Select the claim that is incorrect. (Hint: refer to the R command results.)

x	0	1	2	3
$P(X = x)$	0.343	0.441	0.189	0.027

```
> dbinom(x = 0, size = 3, prob = 0.3)
[1] 0.343
> dbinom(x = 1, size = 3, prob = 0.3)
[1] 0.441
> dbinom(x = 2, size = 3, prob = 0.3)
[1] 0.189
> dbinom(x = 3, size = 3, prob = 0.3)
[1] 0.027
```

(A) $X \sim Bin(3, 0.3)$
(B) This is a valid probability distribution.
(C) The standard deviation of X equals to $\sqrt{0.07}$.
(D) The mean of X equals to 0.9.

7. (3 points) We have a random variable X with mean 0.5 and standard deviation 0.224. Its density curve is as follows. We know that the density at 0.9 is 0.54 and the probability that X is smaller than 0.1 equals to 0.028. What is the probability that X is smaller than or equal to 0.9?



(A) 0.486
(B) 0.54
(C) 0.972
(D) `pnorm(0.9, mean = 0.5, sd = 0.224)`

8. (3 points) Suppose we have a random variable $X \sim N(2, 3)$, based on the 68-95-99.7 rule, what is the probability that X is greater than or equal to -1 and smaller than 11?
   (A) 0.9735
   (B) 0.975
   (C) 0.8385
   (D) 0.815

9. (3 points) A college senior scored 95 on the psychology exam and scored 85 on the history exam. The mean score for the psychology exam is 89 with a standard deviation 3. The mean score for the history exam is 80 with a standard deviation 2. Suppose both distributions are normal. Relative to others, which exam did this student do better on?
   (A) Psychology
   (B) History
   (C) Equally well
   (D) Cannot decide based on the current information

10. (3 points) Suppose the height of female undergraduate students follows a normal distribution $N(68, 2)$. What is the correct R command that gives us the probability that a randomly selected female undergraduate student is greater than or equal to 69.5 inches?
   (A) `pnorm(0.75, mean = 68, sd = 2)`
   (B) `pnorm(69.5, mean = 68, sd = 2) + dnorm(69.5, mean = 68, sd = 2)`
   (C) `1 - pnorm(0.75)`
   (D) `pnorm(0.75)`

11. (3 points) Suppose the probability is $p = 0.2$ that a person who purchases an instant lottery ticket wins money, and this probability holds for every ticket purchased. Consider different random samples of $n = 100$ purchased tickets. Let $\hat{p}$ be sample proportion of winning tickets in a sample of 100 tickets.
    Which of the following is true?
   (A) The sampling distribution of sample proportion is approximately normal.
   (B) Standard deviation of sampling distribution of sample proportion is $\sqrt{(0.2)(0.8)/100} = 0.04$.
   (C) According to the 68-95-99.7 rule, approximately 95% of sample proportion of tickets that will be money winners will be between 12% and 28%.
   (D) All of the above are true.

12. (3 points) The Central Limit Theorem states :
    (A) For a large n, the shape of the sampling distribution of sample mean is approximately normal.
    (B) the shape of the sampling distribution of sample mean is always normal.
    (C) the mean of sampling distribution of sample mean is equal to the population mean.
    (D) the standard deviation of sampling distribution of sample mean is always less than the standard deviation of the population for $n > 1$.

13. (4 points) Did you circle multiple choice answers on page 15?
    (A) Yes, I did.
    (B) I will now.
    (C) I will now.
    (D) I will now.

Problem 2. (13 points) Be sure to show all work for full credit.

1. (4 points) In elections, television networks often declare the winner well before all the votes have been counted. They do this using exit polling, surveying voters after they leave the voting booth. To predict how (i) 9.5 million voters in California voted in 2010, a TV exit poll used surveyed (ii) 3889 voters and found (iii) 53.1% of them voted for Jerry Brown (Democratic candidate in 2010 California gubernatorial race). After all the votes were counted, Jerry Brown won (iv)53.8% of votes.

   Based on the information above, match each of the definitions a) - d) to one of (i)-(iv). If it is unknown, state "unknown".

   a) Population of interest

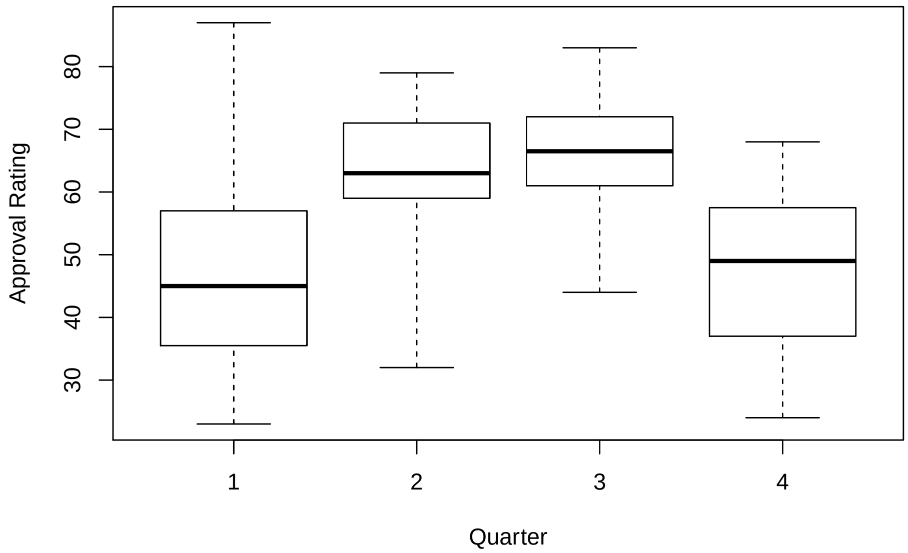
   b) Parameter

   c) Sample

   d) Statistic

Be sure to show all work for full credit.

Political scientists have noticed that presidential approval ratings tend to fluctuate with the seasons. To examine this claim, we use data on presidential quarterly approval ratings from 1945 to 1976. The data are displayed in the boxplots below.



Answer the following questions regarding the presidential approval rating data.

2. (3 points) During which quarter does the presidential approval rating seem to be lowest? Clearly state your reasoning; state explicitly the statistic you are basing your answer on.

3. (3 points) Give an estimate for the IQR for the first quarter approval ratings.

4. (3 points) Below is R commands used to construct the side-by-side box-plot.

```
> dat <- read.csv(file.choose())
> names(dat)
[1] "ratings" "quarters"
> boxplot(___________(a)____________, xlab="Quarter", ylab="Approval Rating")
```

Fill in the blank ____(a)____ that makes a side-by-side boxplot.

Problem 3. (15 points) Be sure to show all work for full credit.

Below is a table that summarizes 302 students' college affiliations (College of Biology Science (CBS) or College of Liberal Arts (CLA)) and their favorite seasons.

	Favorite season				Total
	Spring	Summer	Fall	Winter	
CBS	27	35	52	19	133
CLA	6	59	87	17	169
Total	33	94	139	36	302

Let :

- A be the event that a randomly selected student is in CBS
- B be the event a randomly selected student's favorite season is summer
- C be the event a randomly selected student's favorite season is fall.

1. (4 points) Find the probability that a randomly selected student either {is in CBS OR likes summer the most}.
   Use A, B, and/or C to define this event and calculate the probability. Round your answer to the nearest 3 decimal places.

2. (4 points) *Given that* a randomly selected student's favorite season is summer, what is the probability this student is in CBS?
   Use A, B, and/or C to define this event and calculate the probability. Round your answer to the nearest 3 decimal places.

Copy of the table from the previous page

	Favorite season				Total
	Spring	Summer	Fall	Winter	
CBS	27	35	52	19	133
CLA	6	59	87	17	169
Total	33	94	139	36	302

Let :

- A be the event that a randomly selected student is in CBS
- B be the event a randomly selected student's favorite season is summer
- C be the event a randomly selected student's favorite season is fall.

3. (3 points) Which pairs of events (A, B), (A, C), or (B, C), if any, are disjoint?

4. (4 points) Are A and B independent? Show your work to support your answer.

Problem 4. (20 points) Be sure to show all work for full credit.

Suppose the weights of eighteen-year-old follow a normal distribution with mean 127 and standard deviation 12 (in pound). We randomly sample 4 eighteen-year-olds and ask if their weight is larger than 127 pounds.

You may find the following R command results helpful.

```
pnorm(0.5) = 0.6915;   qnorm(0.242) = -0.7;   qnorm(1 - 0.242) = 0.7
```

1. (2 points) Let W represent the weight of a randomly selected eighteen-year-old. What is the distribution of W? Use statistical notation.

2. (4 points) Tom's weight is 133 pounds. What is his z-score? Interpret this z-score.

3. (4 points) What is the probability that a randomly selected teen is heavier than Tom? Round your answer to three decimal places.

Copy of Problem 4 description from the previous page

Suppose the weights of eighteen-year-old follow a normal distribution with mean 127 and standard deviation 12 (in pound). We randomly sample 4 eighteen-year-olds and ask if their weight is larger than 127 pounds.

You may find the following R command results helpful.

`pnorm(0.5) = 0.6915;   qnorm(0.242) = -0.7;   qnorm(1 - 0.242) = 0.7`

4. (4 points) In order to be lighter than at least 24.2% of eighteen-year-olds, what is the maximum weight that Tom should have? Does he need to control his weight? Round your answer to one decimal place.

5. (2 points) We randomly sample 4 eighteen-year-olds and ask if their weight is larger than 127 pounds. Let V represent the number of teens who say they are heavier than 127 pounds. What is the distribution of V? Write down the statistical notation.

6. (4 points) What is the probability that more than 2 teens in our sample say they are heavier than 127 pounds? Show your work including formulas with numbers plugged in.

Problem 5. (12 points) Be sure to show all work for full credit.

Sheila's doctor is concerned that she may suffer from gestational diabetes (high blood glucose levels during pregnancy). There is variation both in the actual glucose level and in the blood test that measures the level. A patient is classified as having gestational diabetes if the glucose level is above 170 milligrams per deciliter (mg/dl) one hour after a sugary drink is digested. Sheila's measured glucose level one hour after ingesting the sugary drink varies according to the normal distribution with $\mu = 150$mg/dl and $\sigma = 20$mg/dl.

You may find the following R code helpful.

```
pnorm(1.0) = 0.8413, pnorm(1.5) = 0.9332, pnorm(2.0) = 0.97725,
qnorm(0.01) = -2.33, qnorm(0.025) = -1.96, qnorm(0.05) = -1.64.
```

1. (4 points) If measurements are made on randomly selected four separate days and sample mean is computed, describe the shape, center, and spread of the sampling distribution of sample mean. Write down the statistical notation for this sampling distribution.

2. (4 points) If the sample mean of Sheila's four readings is compared with the criterion 170mg/dl, what is the probability that Sheila is diagnosed as having gestational diabetes?

Copy of Problem 5 description from the previous page.

Sheila's doctor is concerned that she may suffer from gestational diabetes (high blood glucose levels during pregnancy). There is variation both in the actual glucose level and in the blood test that measures the level. A patient is classified as having gestational diabetes if the glucose level is above 170 milligrams per deciliter (mg/dl) one hour after a sugary drink is digested. Sheila's measured glucose level one hour after ingesting the sugary drink varies according to the normal distribution with $\mu = 150$mg/dl and $\sigma = 20$mg/dl.

You may find the following R code helpful.

`pnorm(1.0) = 0.8413, pnorm(1.5) = 0.9332, pnorm(2.0) = 0.97725,`
`qnorm(0.01) = -2.33, qnorm(0.025) = -1.96, qnorm(0.05) = -1.64.`

3. (4 points) What is the level L such that there is a probability of 0.05 that sample mean glucose level of four test results falls above L for Sheila's glucose level distribution?

Name: ___________________________

Lecture Section:	001	006	0011	016	021
Lecture time:	9:05 am	8:00 am	10:10 am	11:15 am	12:20 pm
(**Circle One**)	Zhang	Yang	Park	Park	Park

Question	Answer			
1	A	B	C	D
2	A	B	C	D
3	A	B	C	D
4	A	B	C	D
5	A	B	C	D
6	A	B	C	D
7	A	B	C	D
8	A	B	C	D
9	A	B	C	D
10	A	B	C	D
11	A	B	C	D
12	A	B	C	D
13	A	B	C	D

Please do NOT write in the following table. This is for grading purpose only!

Question	I	II	III	IV	V	
Score						
Total	40	13	15	20	12	100