

STAT3011
Spring 2022
Exam 1 (B)
Time Limit: 120 Minutes

Name (Print): SOLUTION

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 17 pages (including this cover page and the multiple choice answer sheet). Check to see if any pages are missing. There are 14 multiple choice questions and 5 short answer problems.
- The exam is closed book. You may *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. You may *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 explanation 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 the **ONLY ONE** correct answer for each question. Circle your answers to all questions in the answer sheet provided. (NO explanation is needed).

1. (3 points) Which statement about sample standard deviation s is *false*?
*** (A) **s can never be zero.**
(B) s can never be negative.
(C) s is not resistant to outliers.
(D) For any normal distributions, about 95% of the data fall within 2 standard deviations from the mean.

2. (3 points) If a distribution is right skewed, which of the following is **true**?
(A) Median is larger than the mean.
*** (B) **The distribution has longer tail on the right than on the left.**
(C) The empirical rule (68-95-99.7 rule) applies.
(D) The distribution is bimodal.

3. (3 points) Which of the following is a continuous variable?
(A) the score of a 20 true/false quiz where each question is worth 1 point.
(B) the number of grains of sand in a cubic mile.
(C) the proportion of males in a statistic class with 100 students.
*** (D) **None of the above are continuous variable**

4. (3 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 9.5 million voters in California voted in 2010, a TV exit poll used surveyed 3889 voters and found 53.1% of them voted for Jerry Brown (Democratic candidate in 2010 California gubernatorial race). After all the votes were counted, Jerry Brown won 53.8% of votes.
Based on the information above, which of the following definitions best describes the number 53.8%
*** (A) **Parameter**
(B) Statistic
(C) Sample
(D) Population

53.8% is the true population proportion of voters who voted for Jerry Brown. Hence parameter.

5. (3 points) In probability theory, events which can never occur together are best described as

- (A) independent events
- (B) dependent events
- *** (C) **disjoint events**
- (D) None of the above

6. (3 points) A previous study show that about 25% of all deaths are due to cancer. Of the deaths that are due to cancer, 30% are attributed to tobacco, 40% to diet, and 30% to other causes. What is the probability that a death is due to cancer and tobacco?

*** (A) **0.075**
(B) 0.75
(C) 0.30
(D) 0.03

$$P(C \text{ and } T) = P(C \cap T) = P(C)P(T|C) = (0.25)(0.3) = 0.075$$

7. (3 points) In a best out of three series played between teams A and B, the team that gets two wins first wins the entire series. Assume that the probability that the series end after two games is 25%. Which below is the mean of the probability distribution of X , where X is number of games played in the series?

(A) 2.95
*** (B) **2.75**
(C) 2.65
(D) 2.4

8. (3 points) Suppose X follows a binomial distribution $\text{Bin}(n, p)$, and we know that the expected value $E(X) = 4$ and the variance $\sigma^2 = 2.4$. What is the value of the parameter n ?

*** (A) **10**
(B) 9
(C) 8
(D) 7

9. (3 points) Find the value of $a > 0$ such that $P(-a < Z < a) = 0.6$

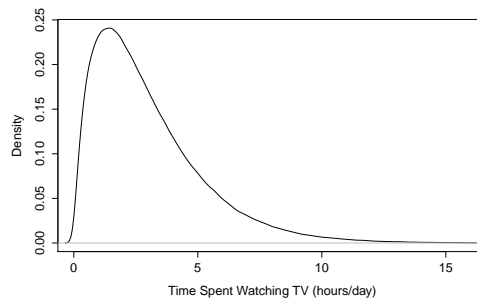
*** (A) **qnorm(0.8)**
(B) qnorm(0.7)
(C) qnorm(0.6)
(D) qnorm(0.4)

10. (3 points) The incomes of trainees at a local mill are normally distributed with a mean of \$1100 and a standard deviation of \$150. We want to find the percentage of trainees earning less than \$900 a month. Which of the following is **CLOSEST** to the true percentage? (Hint: the XX-XX-XX.X Rule)

(A) 2%

- *** (B) **10%**
(C) 35%
(D) 90%

For Questions 11 and 12: Let X = number of hours an American watches TV on an average day. The following is a density curve for random variable X :



11. (3 points) What is the shape of the sampling distribution of the sample mean \bar{X} after drawing 40,000 samples with $n = 2$?
- *** (A) **Unimodal, right-skewed**
(B) Unimodal, left-skewed
(C) Symmetric, approximately bell-shaped
(D) Uniform distribution
12. (3 points) What is the shape of the sampling distribution of the sample mean \bar{X} after drawing 40,000 samples with $n = 100$?
- (A) Unimodal, right-skewed
(B) Unimodal, left-skewed
*** (C) **Symmetric, approximately bell-shaped**
(D) Uniform distribution
13. (3 points) Which of the following properties are TRUE regarding the sampling distribution of sample mean?
- (A) By the central limit theorem, the distribution of \bar{X} is normal for any sample size n .
*** (B) **When the population being sampled follows a normal distribution, the distribution of \bar{X} is normal for any sample size n .**
(C) The mean of all possible sample means are smaller than the population mean.
(D) Sampling distributions can only be constructed for the sample mean and sample standard deviation
14. (1 point) Did you circle your multiple choice answers on page 17 ?
- *** (A) **No, but I will now.**

*** (B) **Yes.**

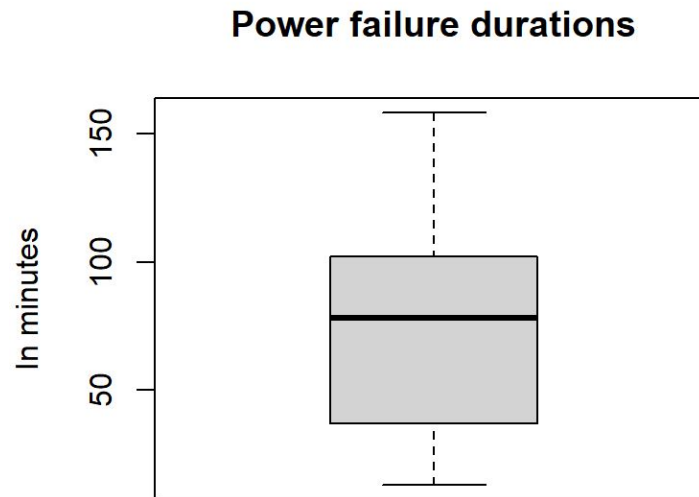
*** (C) **Yes.**

*** (D) **Yes.**

Problem 2. (15 points) Be sure to show all work for full credit.

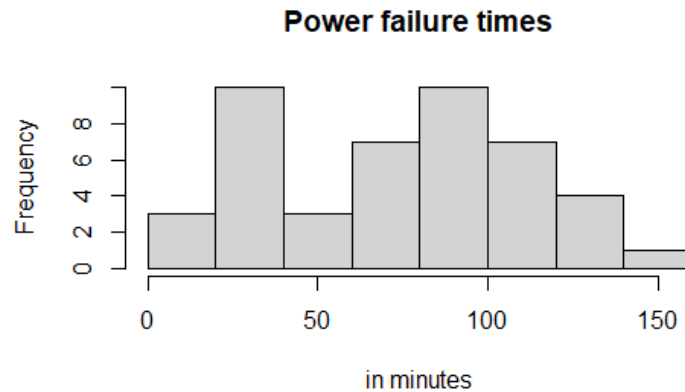
A power company wants to investigate power failures in the last 20 years. The length (duration) of the most recent 45 power failures, in minutes, are recorded.

1. (4 points) Use the boxplot below and estimate median and IQR.



Median is around 80 (minutes) and IQR is about $100 - 40 = 60$ minutes.

2. (3 points) Below is a histogram of the 45 power failure. Describe the distribution. Is there any information/characteristic that you couldn't see from the boxplot?



Bi-modal, not symmetric, ((slightly skewed to the right)), no outliers. We didn't see the bimodal feature in boxplot.

3. (3 points) The sample mean of power failure times is 74 minutes. Based on two plots from the previous page, which of the following is the most reasonable value to be the sample standard deviation of power failure? Explain why the other options are eliminated.
- A) -5
 - B) 0
 - C) 40
 - D) 100

C) 40.
A) standard deviation can't be negative
B) too small considering variability in the data set
3 standard deviation rule doesn't apply with $s=100$ D) too large

Do not use any information above

4. (5 points) The following 3 numbers represent the number of pairs of shoes owned by randomly selected students.

2, 4, 6

Calculate the sample mean and sample standard deviation and interpret the results.

$$\text{mean } \bar{x} = (2 + 4 + 6)/3 = 4$$

standard deviation:

$$s = \sqrt{\sum (x - \bar{x})^2 / (n - 1)}$$

$$= \sqrt{\{(2 - 4)^2 + (4 - 4)^2 + (6 - 4)^2\} / 2}$$

$$= \sqrt{4} = 2 \text{ Average number of shoes own by students is 4 pairs.}$$

Typical students have 2 pairs less or more than the mean.

Problem 3. (12 points) Be sure to show all work for full credit.

A doctor decides to analyze his patients' COVID vaccination records during the year of 2021. He examined medical records of randomly selected 100 patients and noted whether they are vaccinated or not. The data is categorized into two groups: those under age of 40, and those age 40 or older.

Results showed that among those under 40 years old, 20 out of the 30 patients are vaccinated. Among those 40 years or older, 60 out of the 70 patients are vaccinated.

1. (1 point) What is the probability that a randomly selected patient is not vaccinated?

Among the 100 patients, 80 got vaccinated, so $P(\text{no vaccination}) = (100 - 80)/100 = 0.2$

2. (1 point) Calculate the probability of a randomly selected patient is vaccinated in each age category. Which age group of patients is more likely to get the vaccine?

The probability for getting the vaccination below age of 40 is $20/30=0.667 < 60/70=0.857$, which is the probability for getting the vaccination above age of 40. So age 40 and above is more likely to get the vaccine.

Do Not Use any information above.

(Hypothetical study) The doctor wants to investigate whether there were severe side effects after the first shot or after the second shot. Let A be the event of severe side effects after the first shot, and B be the event of severe side effects after the second shot. He estimates $P(A) = 0.2$ and $P(B) = 0.25$.

Address Question 3 and 4 based on Situation 1 and Question 5 and 6 based on Situation 2.

Situation 1: The occurrences of severe side effects are independent.

3. (2 points) Find the probability of severe side effects either after the first shot or after the second shot (or both). Write the formula and calculate the final answer.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.2 + 0.25 - 0.2 * 0.25 = 0.4$$

4. (3 points) *Given that* there are no severe side effects after the first shot, find the probability of severe side effects after the second shot. Write the formula and calculate the final answer.

$$P(B|A^c) = P(B \cap A^c) / P(A^c) = P(B)P(A^c) / P(A^c) = 0.25$$

Situation 2: The occurrences of severe side effects are disjoint.

5. (2 points) Find the probability of severe side effects after the first shot or after the second shot (or both). Write the formula and calculate the final answer.

$$P(A \cup B) = P(A) + P(B) = 0.2 + 0.25 = 0.45$$

6. (3 points) *Given that* there are no severe side effects after the first shot, find the probability of any severe side effect after the second shot. Write the formula and calculate the final answer.

$$P(B|A^c) = P(B \cap A^c)/P(A^c) = P(B)/(1 - P(A)) = 0.25/0.8 = 0.3125$$

Problem 4. (8 points) Be sure to show all work for full credit.

Suppose the height of female follows normal distribution $N(65, 3.5)$ and the height of male follows normal distribution $N(70, 4.0)$.

1. (2 points) Suppose Emma's (female) height is 70, find the corresponding z-score and interpret it.

$$1) \text{ z-score} = \frac{x - \mu}{\sigma} = \frac{70 - 65}{3.5} = 1.43. \text{ Her height is 1.43 standard deviation above the mean 65.}$$

2. (2 points) Suppose Tom's (male) height is 75, compare with Emma, who is relatively taller? Show your work/explain.

$$2) \text{ z-score for Tom: } (75-70)/4 = 1.25 < \text{z-score of Emma, so Emma is relatively taller.}$$

3. (2 points) Find the 40th percentile of female's height. Recall: the distribution of female's height follows a normal distribution $N(65, 3.5)$.

You may / may not find following R commands useful.

```
pnorm(0.4) = 0.655  
qnorm(0.4) = -0.253  
pnorm(3) = 0.999
```

$$\textbf{3) } \mu + z_{40\%}\sigma = 65 - 0.25 \times 3.5 = 64.125$$

4. (2 points) Find the probability that a randomly selected female is exactly 72 inches tall. Show your work or explain.

4) Since it is continuous variable, $P(X = 72) = 0$.

Problem 5. (10 points) In one instant lottery, the winning rate for each lottery ticket is 0.2. A person buys 5 tickets. Let X be the number of winning tickets among the 5 tickets that are purchased. Assume the tickets are independent with each other.

1. (4 points) What is the distribution of X ? Write down the name of the distribution of X and corresponding parameters.

$$X \sim \text{Bin}(5, 0.2).$$

2. (2 points) What is the probability that there are at least 4 winning tickets among the 5?

$$P(X \geq 4) = P(X = 4) + P(X = 5) = 5(0.2)^4(0.8) + 1(0.2)^5 = 0.00672$$

3. (4 points) Find and interpret the mean and standard deviation of X . If you don't know how to find the mean and standard deviation of X , make up numbers and interpret.

$$\begin{aligned}\mu &= E(X) = np = 5(0.2) = 1, \\ \sigma &= \sqrt{np(1-p)} = \sqrt{5(0.2)(0.8)} = 0.89.\end{aligned}$$

If one buys 5 lottery tickets, the long-run average number of winning ticket is 1. or If we pick many samples with 5 tickets, on average we observe 1 winning ticket.

If we pick many samples with 5 tickets, on average we observe the number of winning ticket is $\sigma = 0.89$ away from the mean $\mu = 1$.

Problem 6. (15 points) Be sure to show all work for full credit.

Assuming the population proportion of **females** among all adult Facebook users in 2011 is 0.56.

1. (3 points) Let a random variable \mathbf{X} be defined as:
 $\mathbf{X} = 1$ when a random selected user is a female and $\mathbf{X} = 0$ when the user is not a female.
Find the mean and standard deviation of \mathbf{X} .

$$\mu_X = p = 0.56, \sigma_X = \sqrt{(1-p)^2 p + p^2(1-p)} = \sqrt{p(1-p)} = \sqrt{0.44 * 0.56} = 0.4964$$

2. (3 points) Suppose that we randomly select $n = 1000$ Facebook users. Find the mean and standard deviation of the **sampling distribution** for the sample proportion of female adult Facebook owners at 2011.

$$\text{mean } \mu_{\hat{p}} = p = 0.56, \text{ standard deviation } \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.56*0.44}{1000}} = 0.0157$$

3. (4 points) If we **decrease** the sample size, how will each of the following measurements change? Circle one of (Increase, Decrease or No change). No Explanation Required.

- Mean of X : increase, decrease, no change
- Standard deviation of X : increase, decrease, no change
- Mean of \hat{p} : increase, decrease, no change
- Standard deviation of \hat{p} : increase, decrease, no change

no change, no change, no change, increase
Population mean, population standard deviation and sample mean no change.
standard deviation of sample proportion increase as n increases.

4. (2 points) Is it reasonable to assume a normal shape for the sampling distribution of the sample proportion of female adult Facebook users in 2011? Explain briefly.

Yes. Both np and $n(1-p)$ are larger than 15. According to CLT, the sampling distribution of the sample proportion is approximately normal.

Do not use any information given above

A 2011 survey asked randomly sampled adult Facebook users about their Facebook privacy settings. One of the questions on the survey was, “Do you know how to adjust your Facebook privacy settings to control what people can and cannot see?” The responses are cross-tabulated based on sex.

		Sex		Total
		Male	Female	
Response	Yes	288	378	666
	No	61	62	123
	Not sure	10	7	17
	Total	359	447	806

5. (3 points) i) What is the **sample size** in this survey?
- ii) Give a **point estimate** \hat{p}_1 for the true proportion of adult Facebook users {who are **male** and answered **No** to “Do you know how to adjust your Facebook privacy settings to control what people can and cannot see?”} and
- iii) give a **point estimate** \hat{p}_2 for the true proportion of adult Facebook users {who answered **No** to “Do you know how to adjust your Facebook privacy settings to control what people can and cannot see?”}.

Sample size is 806,
point estimate \hat{p}_1 is $\frac{61}{806} = 0.0757$ and
point estimate \hat{p}_2 is $\frac{123}{806} = 0.1526$

Name: _____

Lecture Section: 001 006 011 016
Lecture time: 11:15 am 8:00 am 12:20 pm 10:10 am
(Circle One) Park Yang Xu Shen

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
14	A	B	C	D

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

Question	1	2	3	4	5	6	100
Score							
Total	40	15	12	8	10	15	100