STAT3011
Fall 2021
Exam 1 (A)
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 19 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. **Do not** use your books, notes, 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 I. (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) When a survey asked "On average, how many hours do you exercise?", of 200 responses, the mode was 1 (hour), the median was 1, the mean was 2.8. Standard deviation is 2.4. Based on these statistics, what would you conclude about the shape of the distribution?

   (A) This distribution is probably bell-shaped because the number of responses is large and the distributions of large data sets always have a bell shape.

   *** (B) **This distribution is probably skewed to the right because the mean is larger than the median, and the standard deviation is almost as large as the mean.**

   (C) This distribution is probably skewed to the left because the standard deviation is larger than mode.

   (D) The distribution is probably uniformly distributed because the mode and the median are similar.

2. (3 points) Suppose you collected data and constructed a histogram. It is approximately uniformly distributed. You deleted *one observation* from the sample. This observation's value is equal to the same mean. How will the sample mean and standard deviation change?

   (i) sample mean would increase.
   (ii) standard deviation would increase.
   (iii) sample mean would decrease.
   (iv) standard deviation would decrease.

   (A) (i) is correct.

   *** (B) **(ii) is correct.**

   (C) (iii) is correct.

   (D) (iv) is correct.

3. (3 points) Let A and B be events with $P(A) = 0.4$ and $P(B) = 0.7$. It is known that $P(A \cup B) = 0.82$. Which of the following statements about events A and B is correct?

   *** (A) **A and B are independent.**

   (B) A and B are disjoint.

   (C) A and B are both disjoint and independent.

   (D) A and B are neither disjoint nor independent.

4. (3 points) Consider rolling two fair dice (6-sided) at the same time. What is the conditional probability of getting doubles (outcomes of two dice are the same), *given that* the sum of two numbers is 4 or lower?

   (A) 1/6

   (B) 1/5

*** (C) **1/3**

   (D) 2/36

5. (3 points) In 4 independent and identical trials, the probability that event A happens is the same for each of all 4 trials. Suppose we know that the probability that event A happens at least one time in 4 trials is 65/81, then the probability that event A happens in a single trial is:

*** (A) **1/3**

   (B) 2/3

   (C) 1/2

   (D) 3/4

6. (3 points) Suppose X follows normal distribution $N(\mu = 0, \sigma = 2)$.
   If $P(X \geq 1) = p$, then $P(-1 < X < 0) =$:

   (A) 1-p

   (B) p

*** (C) **1/2-**p

   (D) 1/2+p

7. (3 points) Suppose students' test scores approximately follow normal distribution $N(\mu = 70, \sigma = 10)$. Then approximately _____ percentage of scores fall between 50 and 80. (Hint: the XX-XX-XX.X Rule)

   (A) 80.5

*** (B) **81.5**

   (C) 82.5

   (D) 83.5

8. (3 points) Select all **TRUE** statements about a normal distribution

   i The total area under a normal curve depends on its mean and standard deviation.

   ii Changing mean will result in a stretch in the normal curve.

   iii A normal distribution can be completely determined by its median and standard deviation.

   iv If $X \sim N(\mu, \sigma)$, we cannot calculate the probability $P(X = a)$ without knowing its mean and standard deviation of X.

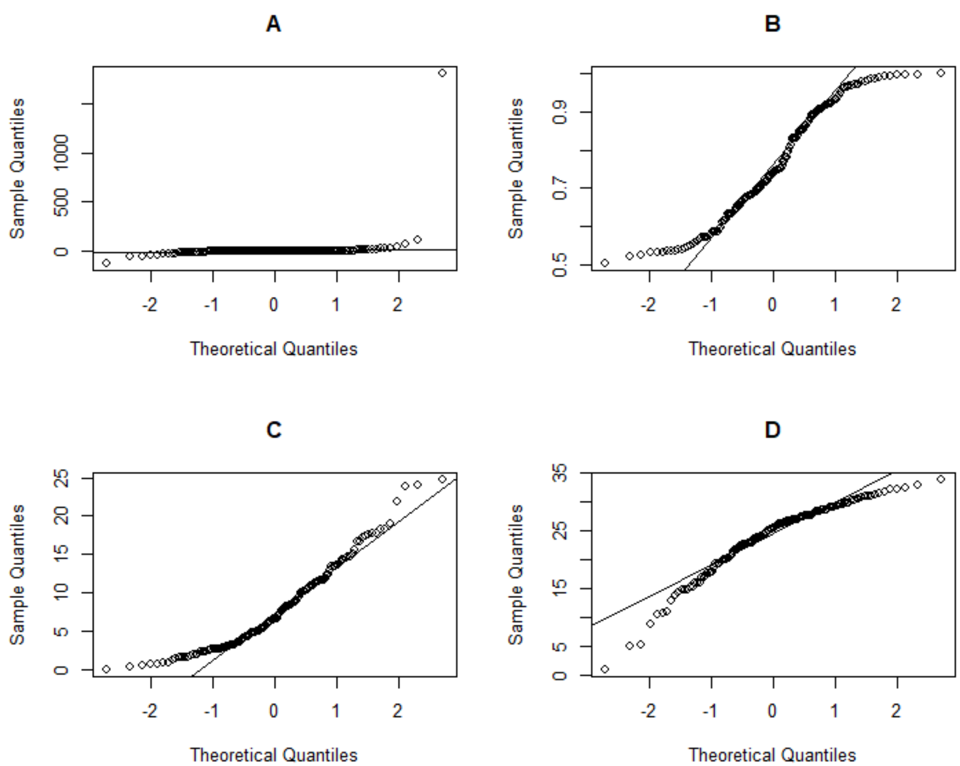
   (A) i and iv

*** (B) **iii only**

   (C) ii and iii

   (D) iii and iv

9. (3 points) Which one of the following figures is most likely to be a left skewed data?



   (A) A

   (B) B

   (C) C

*** (D) **D**

10. (3 points) Let X be a discrete random variable with the following probability distribution :

x	0	1	2	3
$P(X = x)$	a	0.34	b	0.22

If it is known that the expected value of X, $E(X) = 1.5$, find $P(X = 0)$.

(A) 0.25
*** (B) **0.19**
(C) 0.21
(D) 0.16

11. (3 points) In a large lecture, the professor has each student toss a fair coin 100 times and calculate the proportion of times the tosses come up heads. Roughly 95% of students should have proportions between which two numbers?

(A) 0.45, 0.55
(B) 0.49, 0.51
*** (C) **0.4, 0.6**
(D) 0.36, 0.64

For Questions 12 and 13: A random sample $X_1, X_2, \ldots X_{100}$ is drawn from a population that follows normal distribution with mean $\mu = 10$ and standard deviation $\sigma = 1$. Let $\bar{X}$ be the sample mean of $X_1, X_2, \ldots, X_n$.

12. (3 points) What are the mean and standard deviation of the sampling distribution of $\bar{x}$ with sample size $n = 100$?

*** (A) $\mu_{\bar{X}} = 10, \sigma_{\bar{X}} = 0.1$
(B) $\mu_{\bar{X}} = 1, \sigma_{\bar{X}} = 1$
(C) $\mu_{\bar{X}} = 1, \sigma_{\bar{X}} = 0.1$
(D) $\mu_{\bar{X}} = 10, \sigma_{\bar{X}} = 1$

13. (3 points) If the sample size changes from $n = 100$ to $n = 9$, then how does your answer change in the previous question?

(A) $\mu_{\bar{X}}$ increase, $\sigma_{\bar{X}}$ increase
(B) $\mu_{\bar{X}}$ increase, $\sigma_{\bar{X}}$ same
(C) $\mu_{\bar{X}}$ decrease, $\sigma_{\bar{X}}$ decrease
*** (D) $\mu_{\bar{X}}$ **same,** $\sigma_{\bar{X}}$ **increase**

14. (1 point) Did you circle your multiple choice answers on page 19?

*** (A) **No, but I will now.**
*** (B) **Yes.**
*** (C) **Yes.**
*** (D) **Yes.**

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

(Hypothetical study) A psychologist wanted to examine the effect of heartbeat sound on babies' development. He prepared three baby nurseries at a hospital; he arranged the first nursery to have the continuous sound of a heartbeat played, the second to have rhythmical sounds, the third to have no sounds. Each newborn of 66 was randomly assigned to one of the nurseries. At the end of 4 days, he measured how much weight the babies in the three nurseries had gained or lost in grams (g).

1. (4 points) a) What are the explanatory variable and response variable in this study?
   b) For each variable, identify the type of variables (categorical/quantitative. If quantitative, specify whether it is discrete or continuous).

   Explanatory variable : type of sounds baby heard, (or whether a baby heard a human heartbeat sounds, rhythmical sounds, or no sounds during 4 days at nursery.)
   Type of explanatory variable: Categorical
   Response variable : Babies weight gain/loss (in grams) after 4 days
   Type of response variable: Quantitative and Continuous

2. (2 points) Suppose that the average weight gain from the first nursery with heartbeat sound was significantly higher than the rest. Can we establish a cause-and-effect relationship? Explain why/why not.

   Yes, we can conclude that listening to heartbeat sounds for 4 days in their lives caused newborns to gain more weight/lose less. We can establish a cause-and-effect relationship from an (randomized) experiment.

Do not use any information given above

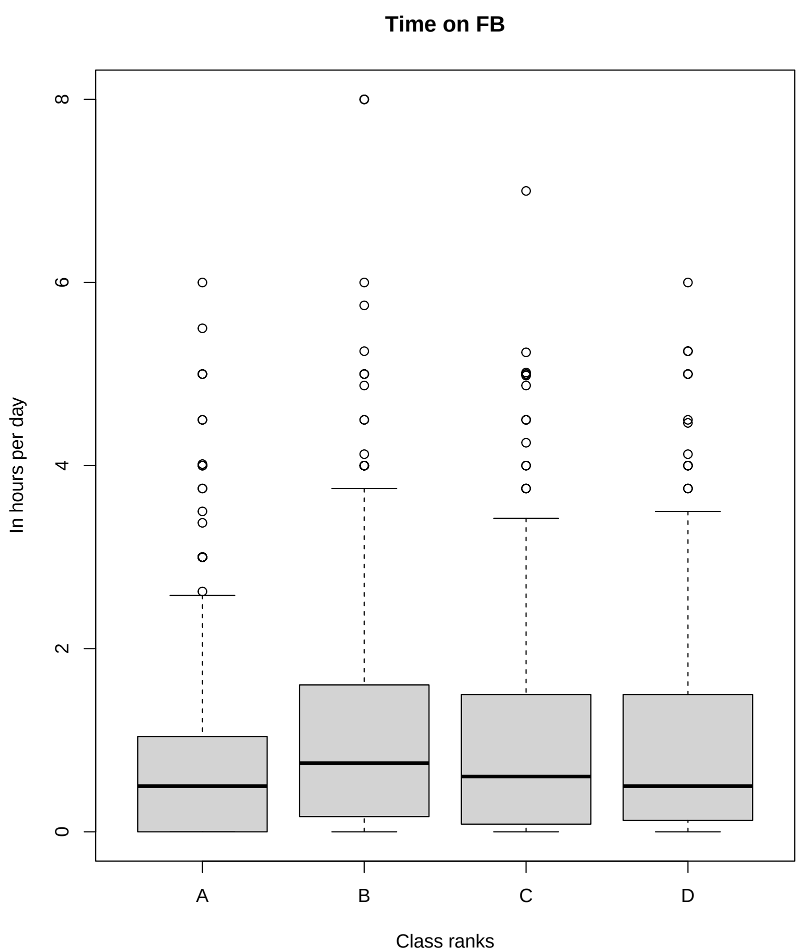
A study from 2015 investigate the amount of time (in hours) per day freshman, sophomores, juniors, and seniors spent on Facebook while doing school work. The students surveyed were U.S. residents from a 4-year, public, primarily residential institution in the

Northeastern United States.

Below is R commands and outputs. Answer the following questions.

```
> dat<-read.csv("timeFB.csv")
> attach(dat)
The following objects are masked _by_ .GlobalEnv:

    class.rank, time
```



3. (2 points) The following R command is used to make the plot above. Fill in ____(a)____

```
> boxplot(time ~ class.rank, main="____(a)____",
ylab="__________", xlab="_______")
```

Time on FB

4. (2 points) It is known that 5-number-summary of Senior's Facebook time use is:

$$0, \quad 0, \quad 0.5, \quad 1, \quad 6$$

Among the 4 box plots on page 7, which of A, B, C, D represents Senior class? Explain why briefly.

Class A represent senior. Middle 50% of the distribution (represented by the middle box) is between 0 to 1 hour. Both minimum and 1st quartile are 0 as we see the lower whisker (minimum) overlaps with the bottom middle box (Q1) Its maximum is 6 hours.

Do not use information above

5. (2 points) For the following 3 values, calculate the sample variance s^2. Show your work.

$$3, 5, 7$$

$$\begin{aligned} s^2 &= \sum_x \frac{(x-\bar{x})^2}{n-1} \\ &= \frac{(3-5)^2+(5-5)^2+(7-5)^2}{2} \\ &= \frac{8}{2} \\ &= 4 \end{aligned}$$

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

A medical practice of 5 doctors tracks whether their patients get a flu shot so they can better understand what demographic group to target the flu shot message to. Within a given day, each of the 5 doctors sees 40 patients for a total of 200 patients. The office staff examined medical records of 200 patients on a given day and noted whether they received a flu shot (or said they received a flu shot) and their age. They reported the data below:

	Received a flu shot?		
Age	**Yes**	**No**	**Total**
6 months - 18 years	35	15	50
19 years - 39 years	15	5	20
40 years - 64 years	20	15	35
65 years and above	70	25	95
Total	140	60	200

1. (1 point) What is the probability that a randomly selected patient received a flu shot?

P(receive flu shot) = 140/200 = 0.70

2. (2 points) What is the probability that { a randomly selected patient is between 19 years and 39 years, or received a flu shot or both}?

P(19-39 years $\cup$ flu shot) = P(19 - 39 years) + P(flu shot) - P(19-39years$\cap$ flushot)
$=\frac{20+140-15}{200}=\frac{145}{200}=\mathbf{0.725}$

3. (4 points) Let event A represent a patient being 6 months - 18 years and event B denote receiving a flu shot. Determine whether A and B are independent. Statistical formula need to be used to support your answer for full credit.

	Received a flu shot?		
Age	**Yes**	**No**	**Total**
6 months - 18 years	35	15	50
19 years - 39 years	15	5	20
40 years - 64 years	20	15	35
65 years and above	70	25	95
Total	140	60	200

- **Check whether P($A \cap B$) = P(A)*P(B) P($A \cap B$) = 0.175 and P(A)*P(B) = 0.25*0.7=0.175. Yes, these are independent.**
- **Check whether $P(A|B) = P(A)$, $P(A|B)$ =35/140=0.25 and P(A) = 0.25. Yes, these are independent.**
- **Check whether $P(B|A) = P(B)$, $P(B|A) = 35/50 = 0.7$ and P(B) = 0.7. Yes, these are independent.**

4. (8 points) Which age group is least likely to receive a flu shot? Hint: your answer needs to incorporate conditional probability for full credit.

- **P(flu shot | 6 month - 18 years) = 35/50=0.700**
- **P(flu shot | 19 years - 39 years) = 15/20=0.750**
- **P(flu shot | 40 years - 64 years) = 20/35=0.571**
- **P(flu shot | 65 years and above) = 70/95=0.737**

40-64 years old are the least likely to receive a flu shot since the probability that they get a flu shot is 0.571, which is smaller than the remaining conditional probabilities.

(Students also can calculate the conditional probability of NOT flu shot)

Problem IV. (7 points) Be sure to show all work for full credit.

Suppose the scores of a GRE verbal test follow the normal distribution with median 151, but standard deviation σ is unknown. Tom scores 146 in the test, and he is told that his score marks the 30th-percentile.

Following R commands may/may not be useful.

```
> pnorm(0.3)
[1] 0.6179114
> qnorm(0.3)
[1] -0.524
```

1. (1 point) Interpret Tom's percentile.

30% of the scores are lower than 146.
or below his score, there are 30% of students' scores.
or above his score there are 70% of scores.

2. (2 points) Find the corresponding z-score of Tom's GRE verbal score and interpret it.

From R command `qnorm(0.3)=-0.524`
Z**-score of Thomas is -0.524, it means Tom's score (146) is -0.524 standard deviation below the mean**

3. (2 points) Find the variance of GER verbal score, σ^2.

Since $z = (x - \mu)/\sigma$ $-0.524 = (146 - 151)/\sigma$ **Hence** $\sigma = 9.54$
(or $x = \mu + z\sigma$ $\sigma^2 = 9.54^2 = 91.01$ **then solve for** σ**)**

4. (2 points) What is the interval that contains the middle 40% of the score distribution?

By symmetry, the middle 40% ranges from 146 **to** $151 + (151 - 146) = 156$**.**

Problem V. (13 points) In a statistics mid term exam, 60 points are from short answer questions, and 40 points are from multiple choice questions. Each multiple choice question carries equal points and has 4 possible answers, only one of which is correct. The correct answer of each question is equally likely to be any one of the 4 choices, and independent for each question.

1. (4 points) Suppose that there are 4 multiple choice questions in this exam. A student guesses the answer to each question at random. Let X denote the number of questions this student guesses correctly.
   a) State the name of the probability distribution of X and its parameter(s).
   b) Find the expected value and standard deviation of X.

   $X \sim$ **Bin**$(4, 1/4)$**, so** $E(X) = 4 \times 1/4 = 1$**, and** $\sigma = \sqrt{4 \times 1/4 \times 3/4} = \sqrt{3/4} \approx 0.866$**.**

2. (6 points) Let Y be the score he gets from 4 multiple choice questions.
   a) Construct the probability distribution table of Y.
   b) Set up the equation to find $E(Y)$.
   You do not need to simplify your answers for a) and b). For instance, expressions such as $\binom{10}{2}$ or ${}_{10}C_2 \frac{(0.5)^{10}}{2^2} + 3 \times \frac{5^2}{0.4^2}$ are acceptable.

   Each question carries 10 points.
   Y **has 5 possible outcomes:**
   $P(Y = 0) = P(X = 0) = (3/4)^4$**,**
   $P(Y = 10) = P(X = 1) = 4(1/4)(3/4)^3$
   $P(Y = 20) = P(X = 2) = 6(1/4)^2(3/4)^2$**,**
   $P(Y = 30) = P(X = 3) = 4(1/4)^3(3/4)$**,**
   $P(Y = 40) = P(X = 4) = (1/4)^4$**.**
   $E(Y) = 10 \times 4(1/4)(3/4)^3 + 20 \times 6(1/4)^2(3/4)^2 + 30 \times 4(1/4)^3(3/4) + 40 \times (1/4)^4 = 10 \times E(X) = 10$**.**

3. (3 points) Suppose this student is sure that he has got 35 points from short answer questions. What is the probability that this student's total exam score (points earned from short answer questions + points earned from multiple choice questions) is greater than equal to 60 points? Show your work.

$Y + 35 \geq 60$, **so** $Y \geq 25$ **or must answer at least three questions correctly.**
$P(Y \geq 25) = P(Y = 30) + P(Y = 40) = 4(1/4)^3(3/4) + (1/4)^4 \approx 0.05$**.**

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

Suppose that 64% of University of Minnesota students were fully vaccinated with COVID-19 Vaccine. Let p = the population proportion of University of Minnesota students that were fully vaccinated and let $\hat{p}$ = sample proportion of fully vaccinated from a random sample of 100 University of Minnesota students.

1. (2 points) Is $\hat{p}$ approximately normally distributed? Explain why.

   Yes. By the central limit theorem, if the sample size is large enough to have the expected number of successes and failures are at least 15 ($np \geq 15$ and $n(1-p) \geq 15$), $\hat{p}$ will approximately follow a normal distribution.

2. (3 points) Find the mean and standard deviation of the sampling distribution of $\hat{p}$.

   $$\mu_{\hat{p}} = p = 0.64$$
   $$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.64(1-.64)}{100}} = .048$$

3. (4 points) Regardless of your answer from 1, suppose $\hat{p}$ is approximately normally distributed. What is the probability that $\hat{p}$ will be between 0.61 and 0.67? Please keep three decimal places.

   Use the following R command if needed.

```
> pnorm(-0.625)
[1] 0.2659855
```

$$
\begin{aligned}
\Pr(0.61 < \hat{p} < 0.67) &= \Pr\left(\frac{0.61 - 0.64}{0.048} < z < \frac{0.67 - 0.64}{0.048}\right) \\
&= \Pr(-0.625 < \hat{p} < 0.625) \\
&= \texttt{pnorm(0.625)-pnorm(-0.625)} \\
&= 1 - 2 * \texttt{pnorm(-0.625)} \\
&= 0.468
\end{aligned}
$$

4. (3 points) Suppose, this time, we took a random sample with 1000 students. Let $\hat{p}_2$ = the sample proportion of those fully vaccinated from a random sample of 1000 University of Minnesota students. Consider the probability that $\hat{p}_2$ will be between 0.61 and 0.67. Will this probability be higher than, lower than or the same as your answer to Question 3? Explain your answer. (You don't need to calculate this new probability.)

The probability that $\hat{p}_2$ is between 0.61 and 0.67 will be higher than the answer to Question 3.

Proportions from larger samples have lower variability and are more likely to be close to center $- p$. Therefore, probability that the sample average will fall in a fixed interval that contains center $- p$ will be higher when the sample size is 1000 than when the sample size is 100.

Name: ___________________________

Lecture Section:	001	005	009	013	017
Lecture time:	9:05 am	8:00 am	12:20 pm	10:10 am	11:15 am
(**Circle One**)	Zhao	Shen	Xu	Park	Song

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	I	II	III	IV	V	VI	100
Score							
Total	40	12	15	7	13	12	100