STAT 3011 **Name (Print):** _____
Fall 2022
Exam 2 (C) **Student ID:** _____
Time Limit: 90 Minutes

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 13 pages (including this cover page and the multiple choice answer sheet). Check to see if any pages are missing. There are 13 multiple-choice problems and 4 short-answer problems.
- The exam is closed book. **Do not** use your books, or any electronic devices 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. (33 points) **Multiple Choice**

Choose ONLY ONE answer for each question. Circle your answers to all questions in the answer sheet provided on page 13. (NO explanation is needed).

1. (3 points) Which is true?
   (A) P-value is the probability that the test statistic takes a value more extreme than the one observed when H_a is true.
   (B) If a test statistic is positive and $H_a : p < p_0$, then we will reject the null hypothesis at $\alpha = 0.05$.
   (C) Both (A) and (B) are true.
   (D) Neither (A) nor (B) are true.

2. (3 points) According to the 2015 American Community Survey, the 90% confidence interval of the mean travel time to work for US workers who drive alone is $(26.3, 26.5)$ minutes. Which of the following statements is TRUE?
   (A) 90% of all US workers who drive alone spend between 26.3 and 26.5 minutes to travel to work.
   (B) Among all possible samples of US workers who drive alone, 90% of which have sample mean travel times to work between 26.3 and 26.5 minutes.
   (C) If we draw 100 random samples of US workers who drive alone, and for each sample we calculate a 90% confidence interval for the population mean travel time to work, then about 10 of these 100 intervals will miss out the true value of population mean.
   (D) All of the statements above are true.

3. (3 points) Suppose you want to estimate, on average, how much time college students spent on social media applications in a typical day. You wish your estimate to be within 0.2 hrs with 95% confidence. How large should your sample be?
   Use sample standard deviation 1 (hr) as an educated guess for standard deviation.
   You may find the following R output helpful.

```
qt(0.975, df=14) = 2.145
qnorm(0.975) = 1.96
qnorm(0.95) = 1.645
```

   (A) 49
   (B) 68
   (C) 97
   (D) 116

4. (3 points) Let Z, T_1, and T_5 represent a random variable that follows standard normal distribution (N(0, 1)), T-distribution with degrees of freedom 1, and degrees of freedom 5, respectively.
   If k is a positive constant number and $P(-k < T_1 < k) = p$, then
   (A) $P(-k < Z < k) > p$
   (B) $P(-k < T_5 < k) > p$
   (C) Both (A) and (B) are correct.
   (D) Both (A) and (B) are wrong.

5. (3 points) A one-sample t-test for testing $H_0 : \mu = 3.2$ vs. $H_a : \mu > 3.2$ resulted in a p-value of 0.98. Based on the same sample, would the 90% confidence interval and the 95% confidence interval for μ contain the value 3.2?
   (A) Both the 90% confidence interval and the 95% confidence interval would contain 3.2
   (B) Neither the 90% confidence interval nor the 95% confidence interval would contain 3.2
   (C) Only the 95% confidence interval would contain 3.2
   (D) We do not have enough information to tell, because there is no equivalence between one-sided t-test and confidence interval.

6. (3 points) A study report estimates the population mean of the weights of eighteen-year-old to be 127 pounds. Two students believe the report *overestimates* the mean weight. To confirm their conjecture, they randomly sampled 64 eighteen-year-old and asked them about their weights. The mean weight of these 64 teenagers is 124 pounds and the sample standard deviation is 12 pounds. Which of the following claim is correct? You may find the following R command results helpful.

```
> pt(2, df = 63)
[1] 0.9750921
```

   (A) At the significance level of 0.02, we will reject the null hypothesis.
   (B) At the significance level of 0.05, the potential type of error we could have made is Type I error.
   (C) At the significance level of 0.1, the potential type of error we could have made is Type II error.
   (D) The alternative hypothesis is $H_a : \mu > 127$

7. (3 points) Suppose we want to conduct a hypothesis test with $H_0 : p = 0.3$ vs. $H_a : p < 0.3$ and we collect a random sample of 60 observations. The sample proportion is $\hat{p} = 0.4$. Which of the following claim is correct?
   (A) The standard error used to compute the test statistic is $\sqrt{\frac{0.4(1-0.4)}{60}}$.
   (B) The p-value of the test is smaller than 0.5.
   (C) The assumptions are satisfied, because the expected number of successes is $24 > 15$.
   (D) Under H_0, $\hat{p}$ approximately follows a normal distribution $N\left(0.3, \sqrt{\frac{0.3(1-0.3)}{60}}\right)$.

8. (3 points) Which of the following examples involves paired data?
   (A) A study comparing the average satisfaction rating taken by a random sample of 100 customers before employee training with the average satisfaction rating taken by a separate random sample of 100 customers after employee training at a large call center.
   (B) A group of 50 students had their blood pressured measured before and after watching a movie containing violence. The difference between blood pressure before and after the movie was calculated.
   (C) Both (A) and (B)
   (D) Neither (A) nor (B).

9. (3 points) Two types of flares are tested for their burning times (in minutes). Assuming that a hypothesis test of the claim that Brand X has a longer burning time than Brand Y has been conducted and that the conclusion is to fail reject the null hypothesis, state the conclusion in non-technical terms.
   (A) There is sufficient evidence to support the claim that Brand X and Brand Y has the same burning time.
   (B) There is sufficient evidence to support the claim that the mean burning time of Brand X is longer than Brand Y's.
   (C) There is not sufficient evidence to support the claim that the mean burning time of Brand X is longer than Brand Y's.
   (D) There is sufficient evidence to support the claim that the difference between two burning time is positive.

10. (3 points) A random sample of freshmen has a sample mean GPA $\bar{x}_1 = 3.7$ and a random sample of seniors has a sample mean GPA $\bar{x}_2 = 3.3$. You want to test if there is a significant difference between freshman's average GPA and seniors' average GPA. Test statistic is 1.85 and the p-value is 0.07. Which of the following is correct?
    (A) The probability we make Type 1 error is 0.07.
    (B) The probability that two population mean CPA's are equal is 0.07.
    (C) The probability that Freshman's mean GPA and senior's mean GPA are different is 0.07.
    (D) If there is no difference between two groups' population mean GPAs, then the probability of observing test statistics of 1.85 or more extreme is 0.07.

11. (3 points) Did you circle multiple choice answers on page 13?
    (A) Yes, I did.
    (B) I will now.
    (C) I will now.
    (D) I will now.

Problem 2. (20 points) Be sure to show all work for full credit.

For Problem 2 Part 1 and 2, the population of interest is women between the age of 45 and 64.
A recent survey of randomly chosen 1000 American women between the age of 45 and 64 asked them what medical condition they most feared. Of those samples, 60% said breast cancer, 10% said hear disease, and the rest picked other conditions. Let p represent the population proportion of women between the ages of 45 and 64 who most fear breast cancer. Answer each of the following questions below. You may find the following R commands useful.

```
qnorm(0.03) = -1.88
qnorm(0.94) = 1.55
qnorm(0.98) = 2.05
```

1. (2 points) Calculate the point estimate for p. Use correct statistical notation and specify its value.

2. (9 points) Construct the 94% confidence interval to estimate p. Remember to check the assumptions and interpret the result. Round your answer to four decimal places.

In Problem 2 Part 3: We have a different population of interest from Part 1 & 2; Women 65 years or older. So don't use any information from the previous page.

3. (4 points) Suppose another researcher found that for women with the age of 65 and older, the 95% confidence interval is (0.489, 0.539). Based on this confidence interval, can we conclude that the majority (more than 50%) of women with the ages of 65 and older most feared breast cancer? Explain why or why not.

In Problem 2 Part 4: We have a different population of interest: Women between the ages of 25 and 44. So don't use any information above.

4. (5 points) Let p be the proportion of women between the ages of 25 and 44 who most fear breast cancer.
   How large a sample is needed if he wishes to be 96% confident that the sample proportion will be within 3% of the true population proportion? Suppose that he doesn't have any prior information to estimate p.
   You may find the following R command(s) useful.

```
qnorm(0.03) = -1.88
qnorm(0.94) = 1.55
qnorm(0.98) = 2.05
```

Problem 3. (23 points) Be sure to show all work for full credit.

A study group supposes fewer than 40% of UMN students exercise at least two hour per day. To confirm their guess, they collect a random sample of 391 UMN students and get their reported exercise hours per day as in the table below.

Exercise Hours	< 0.5	0.5	0.6	1	1.5	2	2.5	3	3.5	≥ 4
Number of Students	25	18	1	179	20	117	2	16	1	12

Table 1: Frequency Table of Students' Exercise Hours

You may find the following R commands helpful.

```
> pnorm(0.867)
[1] 0.807029
> pt(0.867, df = 390)
[1] 0.8067624
```

1. (12 points) Write down the five steps to perform the hypothesis test at the 0.05 significance level. Round your test statistic to three decimal places and p-value to four decimal places.

Copy of Problem 3 description from the previous page
A study group supposes fewer than 40% of UMN students exercise at least two hour per day. To confirm their guess, they collect a random sample of 391 UMN students and get their reported exercise hours per day as in Table 1 on page 7.

2. (3 points) State the Type I error in the context of this problem.

3. (3 points) State the Type II error in the context of this problem.

4. (3 points) Which type of error (Type I or Type II) could you have made in Question 1? Explain why.

5. (2 points) If the study group uses the same sample to test whether the proportion of UMN students who exercise at least two hours per day is significantly different from 40%. What is the p-value?

Problem 4. (24 points) Be sure to show all work for full credit.

To find out whether a new serum will arrest leukemia, 9 mice, all with an advanced stage of the disease, are selected. Five mice receive the treatment and 4 do not. Survival times, in years, from the time the experiment commenced are measured. Assume that population distributions of survival time are approximately normally distributed. Below is data from the study.

	Survival time	Sample mean	Sample standard deviation
Treatment	2.1, 5.3, 1.4, 4.6, 0.9	2.86	1.97
No treatment	1.9, 0.5, 2.8, 3.1	2.075	1.17

For Problem 4 Part 1 - 4: At the 0.05 level of significance, can the serum be said to have a positive effect on survival time? (effective in increasing surviving time)

1. (4 points) State the necessary assumptions of this test. Determine whether the samples given in the problem description meet each assumption.

Copy of Problem 4 description from the previous page
To find out whether a new serum will arrest leukemia, 9 mice, all with an advanced stage of the disease, are selected. Five mice receive the treatment and 4 do not. Survival times, in years, from the time the experiment commenced are measured. Assume that population distributions of survival time are approximately normally distributed. Below is data from the study.

	Survival time	Sample mean	Sample standard deviation
Treatment	2.1, 5.3, 1.4, 4.6, 0.9	2.86	1.97
No treatment	1.9, 0.5, 2.8, 3.1	2.075	1.17

For Problem 4 Part 1 - 4 At the 0.05 level of significance, can the serum be said to have a positive effect on survival time? (effective in increasing surviving time)

2. (4 points) State the null and alternative hypotheses. Let μ_1 be population mean surviving time of treatment group and μ_2 population mean surviving time of no treatment group.

3. (4 points) State the distribution of test statistic under the null hypothesis.
Calculate the value of test statistic using the sample information provided. Round your answer to the nearest one decimal place.

4. (4 points) Calculate the p-value and draw the conclusion. Use $\alpha = 0.05$.
   Following R commands may or may not be useful.

```
pt(0.7, df=3) = 0.733
pt(0.7, df=7) = 0.747
pnorm(0.7) = 0.758
```

5. (4 points) (Multiple choice) Based on your conclusion from the previous questions, do you think a 95% confidence interval for the difference between two population means : (select one)
   Explain why

   (A) contains only negative numbers.
   (B) contains only positive numbers.
   (C) contains 0.
   (D) Not enough information provided.

6. (4 points) Do you think the result from Part 4 above is statistically significant at $\alpha = 0.01$? Do you think the result is practically significant? Use one or two sentences to support your answer.

This page is left blank intentionally.

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

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

Question	I	II	III	IV	100
Score					
Total	33	20	23	24	100