STAT 3011 **Name (Print):** _____
Fall 2022
Exam 2 (A) **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 11 multiple-choice problems and 3 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 order 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 of the following is the correct $z_{\alpha/2}$ or $t_{\alpha/2}$ to construct an 80% confidence interval for μ with a random sample of size 12?
   Assume that the population distribution is normal and the population standard deviation is unknown.
   (A) `qnorm(0.8)`
   (B) `qt(0.9, df=11)`
   (C) `qt(0.8, df=12)`
   (D) `qt(0.8, df=11)`

2. (3 points) Which of the following is *not* a part of the formula for constructing a confidence interval for population proportion?
   (A) sample proportion
   (B) population proportion
   (C) Standard error of sample proportion
   (D) margin of error

3. (3 points) Which of the following is *not* true about the width of a confidence interval for μ?
   Holding everything else fixed:
   (A) increasing error probability α decreases the width of confidence interval.
   (B) increasing sample size decreases the width of the confidence interval.
   (C) increasing confidence level decreases the width of the confidence interval.
   (D) All of the above is true.

4. (3 points) To estimate the average earnings of students who work during summer break, the government surveyed 8000 students. A 95% confidence interval for μ is (\$5000, \$8000). Which of the following is the correct interpretation of confidence interval?
   (A) If the study were to be repeated many times and we construct a 95% confidence interval from each sample, 95% of those intervals will contain the population mean.
   (B) 95% of students who work during summer break earn between \$5000 and \$8000.
   (C) The probability sample mean falls between \$5000 and \$8000 is 95%.
   (D) All of the above is correct.

5. (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 0.1 hrs with 98% 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(0.98, df=14) = 2.26
qnorm(0.99) = 2.33
qnorm(0.98) = 2.05
```

(A) 97
(B) 421
(C) 511
(D) 543

6. (3 points) A one-sample t-test for testing $H_0 : \mu = -2.8$ vs. $H_a : \mu < -2.8$ resulted in a p-value of 0.96. Based on the same sample, would the 90% confidence interval and the 95% confidence interval for μ contain the value -2.8?

(A) Neither the 90% confidence interval nor the 95% confidence interval would contain -2.8
(B) Only the 95% confidence interval would contain -2.8
(C) Both the 90% confidence interval and the 95% confidence interval would contain -2.8
(D) We do not have enough information to tell because there is no equivalence between one-sided t-test and confidence interval.

7. (3 points) A study report estimates the population mean of the weights of eighteen-year-olds to be 127 pounds. Two students believe the report *underestimates* the mean weight. To confirm their conjecture, they randomly sampled 49 eighteen-year-olds and asked them about their weights. The mean weight of these 49 teenagers is 130 pounds and the sample standard deviation is 14 pounds. Which of the following claim is correct? You may find the following R command results helpful.

```
> pt(1.5, df = 48)
[1] 0.9299181
> pt(1.5, df = 49)
[1] 0.9299847
> pt(1.5, df = 50)
[1] 0.9300486
```

(A) The alternative hypothesis is $H_a : \mu < 127$
(B) At the significance level of 0.05, the potential type of error they could have made is Type I error.
(C) At the significance level of 0.07, the potential type of error they could have made is Type II error.
(D) None of the above

8. (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 $\hat{p} = 0.2$. Which of the following claim is correct?

   (A) The standard error used to compute the test statistic is $\sqrt{\frac{0.2(1-0.2)}{60}}$.

   (B) The p-value of the test is smaller than 0.5.

   (C) The assumptions are not satisfied, because the number of successes is $12 < 15$.

   (D) Under H_0, the test statistic follows a normal distribution $N\left(0.3, \sqrt{\frac{0.3(1-0.3)}{60}}\right)$.

9. (3 points) A student in STAT 3011 wants to compare the selling prices of textbooks at two internet bookstores (Amazon vs e-bay). She first took a random sample of 10 textbooks used that term in courses at the U, based on the list of texts compiled by the college bookstore. Then she collected the prices of those 10 textbooks at Amazon and e-bay. She wants to test whether e-bay prices are cheaper than Amazon prices.
   Which of the following is the correct alternative hypothesis?

   (A) $H_a : p \neq 0.5$

   (B) H_a: $\mu_{\text{Amazon}} > \mu_{\text{e-bay}}$

   (C) H_a: $\mu_D > 0$ where D=(Amazon price - e-bay price)

   (D) $H_a : \mu_{\text{Amazon}} \neq \mu_{\text{e-bay}}$

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 $\bar{x}_2 = 3.3$. The P-value for testing $H_a : \mu_1 \neq \mu_2$ is 0.07. The significance level $\alpha = 0.05$.
    Which of the following is true?

    (A) The probability of making Type I error is 0.07.

    (B) The probability that the null hypothesis is true is 0.07.

    (C) We fail to reject the null hypothesis at $\alpha = 0.05$.

    (D) The result is statistically significant at $\alpha = 0.05$.

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 1000 randomly chosen American women between the age of 45 and 64 asked them what medical condition they most feared. Of those sampled, 50% said breast cancer, 10% said heart disease, and the rest picked other conditions.
Let p represent the population proportion of women between the age 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(0.01) = -2.33
qnorm(0.025) = -1.96
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 a 98% 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 65 years and older, the 95% confidence interval is (0.512, 0.559).
   Based on this confidence interval, can we conclude that the majority (more than 50%) of women 65 years 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 a researcher wishes to be 95% confident that the sample proportion will be within 4% of the true population proportion? Suppose that he doesn't have any prior information to estimate p.

   You may find the following R commands useful.

```
qnorm(0.01) = -2.33
qnorm(0.025) = -1.96
qnorm(0.98) = 2.05
```

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

A study group supposes more than 80% of UMN students exercise at least one 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(4.324)
[1] 0.9999923
> pt(4.324, df = 390)
[1] 0.9999903
```

1. (12 points) Write down the five steps to perform the hypothesis test at the 0.01 significance level. Round your test statistic to three decimal places and p-value to six decimal places.

Copy of Problem 3 description from the previous page

A study group supposes more than 80% of UMN students exercise at least one 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 Problem 3 Part 1 on page 7? 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 one hour per day is significantly *different* from 80%, what is the p-value?

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

To examine the cognitive benefits of interacting with nature, researchers at the University of Michigan designed a randomized experiment. 144 volunteer students were randomly assigned to take a 50-minute walk in either (i) a nature trail in the Ann Arbor Arboretum or (ii) heavy traffic downtown Ann Arbor.
After the walk, participants performed a short-term memory test. Each subject received a score on a scale of 1 to 10 based on their performance. Below are summary statistics from the two groups.

	sample mean	sample standard deviation	sample size
Group 1 (Nature trail group)	9.4	3	72
Group 2 (Downtown group)	8.4	2	72

Let μ_1 be the population mean cognitive performance after walking in a nature trail and μ_2 be the population mean cognitive performance after walking in traffic-heavy street.

Answer the following questions.

1. (4 points) What is the point estimate of $\mu_1 - \mu_2$? Use correct statistical notation and specify its value.

Copy of Problem 4 description from the previous page

To examine the cognitive benefits of interacting with nature, researchers at the University of Michigan designed a randomized experiment. 144 volunteer students were randomly assigned to take a 50-minute walk in either (i) a nature trail in the Ann Arbor Arboretum or (ii) heavy traffic downtown Ann Arbor. After the walk, participants performed a short-term memory test. Each subject received a score on a scale of 1 to 10 based on their performance. Below is summary statistics from the two groups.

	sample mean	sample standard deviation	sample size
Group 1 (Nature trail group)	9.4	3	72
Group 2 (Downtown group)	8.4	2	72

Let μ_1 be the population mean cognitive performance after walking in a nature trail and μ_2 be the population mean cognitive performance after walking in traffic-heavy street.

For Part 2 - 6: Researchers want to test if the mean cognitive score is higher after interacting with nature (walking in a nature trail) than otherwise (walking in traffic-heavy street).

2. (4 points) State the null and alternative hypotheses. Use statistical notations.

   - H_0 :
   - H_a :

3. (4 points) Calculate the value of test statistic.

4. (4 points) P-value is 0.011. Use $\alpha = 0.05$ and draw a conclusion and interpret in context.

5. (4 points) Do you think the result is statistically significant at $\alpha = 0.05$? Do you think the result is practically significant? Use one or two sentences to support your answer.

6. (4 points) State the assumptions of the hypothesis test. Determine whether each assumption is satisfied. Explain.

This page is left blank on purpose.

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