STAT 350 Exam 2 SPRING 2024

V1

Name: ___________________________________

PUID ____________________________

Instructor (circle one): Anand Dixit Timothy Reese Halin Shin Heekyung Ahn

Class Start Time: ○ **9:30 AM** ○ **11:30 PM** ○ **1:30 PM** ○ **2:30 PM** ○ **3:30 PM** ○ **Online**

As a boilermaker pursuing academic excellence, I pledge to be honest and true in all that I do. Accountable together - we are Purdue.

Instructions:

1. **IMPORTANT** Please write your **name** and **PUID clearly** on every **odd page.**
2. **Write your work in the box. Do not run over into the next question space.**
3. You are expected to uphold the honor code of Purdue University. It is your responsibility to keep your work covered at all times. Anyone caught cheating on the exam will automatically fail the course and will be reported to the Office of the Dean of Students.
4. It is strictly prohibited to smuggle this exam outside. Your exam will be returned to you on Gradescope after it is graded.
5. The only materials that you are allowed during the exam are your **scientific calculator**, **writing utensils, erasers, your crib sheet**, and **your picture ID**. If you bring any other papers into the exam, you will get a **zero** on the exam. Colored scratch paper will be provided if you need more room for your answers. Please write your name at the top of that paper also.
6. The crib sheet can be a handwritten or type double-sided 8.5in x 11in sheet.
7. Keep your bag closed and cellphone stored away securely at all times during the exam.
8. If you share your calculator or have a cell phone at your desk, you will get a **zero** on the exam.
9. The exam is only 60 minutes long so there will be no breaks (including bathroom breaks) during the exam. If you leave the exam room, you must turn in your exam, and you will not be allowed to come back.
10. **For free response questions you must show ALL your work to obtain full credit.** An answer without showing any work may result in **zero** credit. If your work is not readable, it will be marked wrong. Remember that work has to be shown for all numbers that are not provided in the problem or no credit will be given for them. All explanations must be in complete English sentences to receive full credit.
11. All numeric answers should have **four decimal places** unless stated otherwise.
12. After you complete the exam, please turn in your exam as well as your table and any scrap paper that you used. Please be prepared to **show your Purdue picture ID**. You will need to **sign a sheet** indicating that you have turned in your exam.

Your exam is not valid without your signature below. This means that it won't be graded.

I attest here that I have read and followed the instructions above honestly while taking this exam and that the work submitted is my own, produced without assistance from books, other people (including other students in this class), notes other than my own crib sheet(s), or other aids. In addition, I agree that if I tell any other student in this class anything about the exam BEFORE they take it, I (and the student that I communicate the information to) will fail the course and be reported to the Office of the Dean of Students for Academic Dishonesty.

Signature of Student: ____

You may use this page as scratch paper.
The following is for your benefit only.

Question Number	Total Possible	Your points
Problem 1 (True/False) (2 points each)	12	
Problem 2 (Multiple Choice) (3 points each)	9	
Problem 3	32	
Problem 4	20	
Problem 5	32	
Total	105	

1. **(12 points, 2 points each) True/False Questions.** Indicate the correct answer by completely filling in the appropriate circle. If you indicate your answer by any other way, you may be marked incorrect.

   1.1. If a simple random sample is taken from a normally distributed population,

   Ⓣ or Ⓕ then the distribution of the sample means follows a normal distribution, regardless of the sample size.

   1.2. If a simple random sample of size 2 or greater is taken from a normally distributed population,

   Ⓣ or Ⓕ then the variance of the sample mean is always greater than the population variance.

   1.3. In a simulation run where differences arise from a normal distribution in a paired sample procedure, and 92% confidence intervals are constructed for the mean difference across 1000 independent sets of paired samples,

   Ⓣ or Ⓕ exactly 80 of these intervals will not contain the true mean difference.

   1.4. When the significance level (α) of a statistical test is reduced while holding all other factors constant,

   Ⓣ or Ⓕ the power of the test increases.

   1.5. In a two-sample independent t-test using the Welch procedure to account for unequal variances between groups,

   Ⓣ or Ⓕ the test statistic is assumed to adhere to an exact t-distribution, provided the assumption of normality holds.

   1.6. In a completely randomized experimental design,

   Ⓣ or Ⓕ random assignment of experimental units to treatments helps to minimize potential biases by helping to distribute extraneous variables more evenly across treatment groups.

2. **(9 points, 3 points each) Multiple Choice Questions.** Indicate the correct answer by completely filling in the appropriate circle. If you indicate your answer by any other way, you may be marked incorrect. **For each question, there is only one correct option letter choice.**

2.1. In a randomized block design, when we block experimental units based on a specific characteristic, the primary objective is to:

Ⓐ Increase the variability arising from extraneous variables by grouping similar experimental units into blocks, thereby enhancing the detection of treatment effects.

Ⓑ Decrease the variability arising from extraneous variables by grouping similar experimental units into blocks, thereby enhancing the detection of treatment effects.

Ⓒ To allocate treatments to experimental units across blocks in a manner that conceals the treatment identities from both the participants and researchers.

Ⓓ Equalize the allocation of treatments to experimental units within each block to facilitate the administrative convenience of the experiment.

Ⓔ Balance the number of experimental units across blocks to primarily focus on the uniformity of treatment application without direct concern for extraneous or confounding variables.

2.2. Suppose a simple random sample of size 400 is taken from a skewed population with a known population mean of 200 units and a population standard deviation of 50 units. Which of the following statements is TRUE regarding the standard deviation of the sample mean?

Ⓐ The standard deviation of the sample mean is equal to the population standard deviation, which is 50 units.

Ⓑ The standard deviation of the sample mean cannot be accurately determined from the given information due to the population's skewed distribution.

Ⓒ The standard deviation of the sample mean, indicative of the sampling distribution's variability, amounts to 0.125 units.

Ⓓ The standard deviation of the sample mean is as large as 2500 units, indicative of the sampling distribution's total variability.

Ⓔ The standard deviation of the sample mean, indicative of the sampling distribution's variability, amounts to 2.5 units.

2.3. When estimating the difference between two population means using confidence inte rvals, if a researcher incorrectly uses a pooled variance estimator under the false ass umption of equal variances, despite the populations having unequal variances, how d oes this affect the margin of error for the confidence interval?

Ⓐ The margin of error is unaffected, as the pooled estimator adjusts for variance differences.

Ⓑ The margin of error decreases, reflecting an underestimated standard error due to the assumption violation.

Ⓒ The margin of error increases, reflecting an overestimated standard error due to the assumption violation.

Ⓓ The margin of error may inaccurately reflect the true variability, underestimating or overestimating it based on the sample sizes and actual variances.

Ⓔ The margin of error becomes zero, indicating a failure of the pooled estimator to account for variance differences.

Free Response Questions 3-5. Show all work, clearly label your answers, and use **four decimal places**.

3. **(32 points)** According to the famous Stanford marshmallow study, children's ability to wait longer for rewards is positively correlated with their educational achievements. However, subsequent research has identified several factors, such as family background, home environment, and cultural differences, that might influence both a child's waiting time for rewards and their educational achievements.

   a) **(3 points)** Which of the following statements is **false**?

   Ⓐ The positive correlation ensures causation because the marshmallow experiment is co nducted before observing their educational achievements.

   Ⓑ The positive correlation might be insignificant after considering lurking variables.

   Ⓒ The underlying factors are lurking variables in the famous marshmallow study.

   Ⓓ None of the above

b) **(3 points)** A researcher is investigating the connection between the birth order of children and their ability to wait longer for rewards. The study will involve **15 randomly chosen households** within the Greater Lafayette area, each with at least two children. For each household, the first and second child will be included in the experiment.

In the study, which of the following variables is **NOT indirectly** controlled by the paired design?

Ⓐ Parent's occupation

Ⓑ Child's birth order

Ⓒ Cultural background

Ⓓ Home environment

c) **(3 points)** The researcher believes that the first child will wait longer for the rewards than the second child. Drawing on previous research, the **standard deviation** of the waiting time difference between the first and second children is established at **37 seconds**, and this research consistently indicates that these differences follow a normal distribution. The researchers set a significance level of $\boldsymbol{\alpha = 0.05}$ and asserts that a meaningful difference between the first and second child waiting times would need to exhibit a **true difference of 15 seconds**. This threshold is determined with the understanding that such a difference would be substantively meaningful.

Select the researcher's hypotheses. Define $\boldsymbol{D}$ = **First Child Wait Time – Second Child Wait Time**

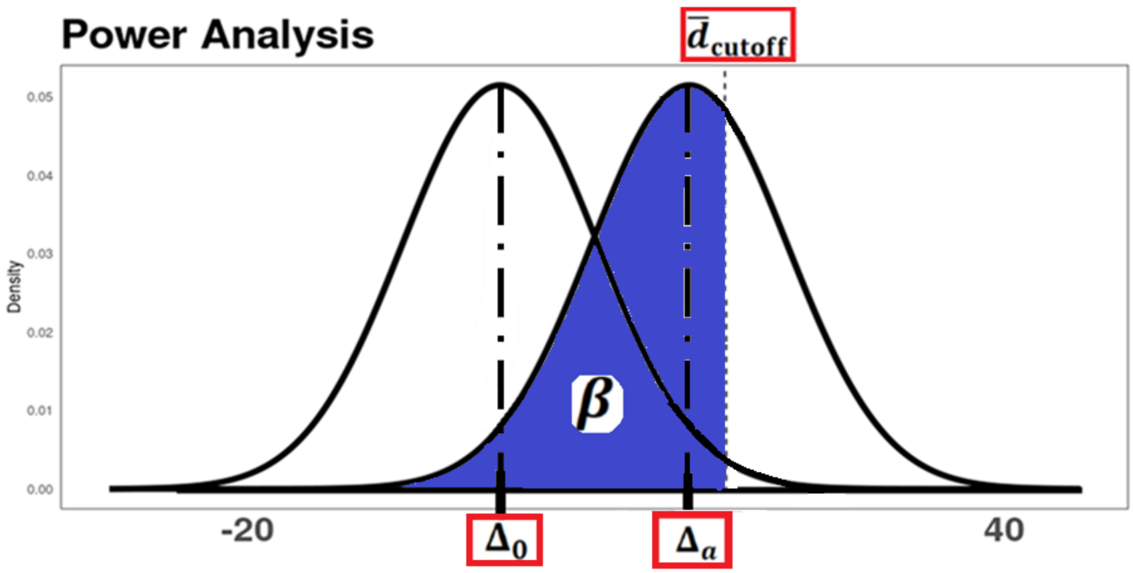
Ⓐ $\boldsymbol{H_0: \mu_D \leq 0;} \quad \boldsymbol{H_a: \mu_D > 0}$

Ⓑ $\boldsymbol{H_0: \mu_D \geq 0;} \quad \boldsymbol{H_a: \mu_D < 0}$

Ⓒ $\boldsymbol{H_0: \mu_D = 0;} \quad \boldsymbol{H_a: \mu_D \neq 0}$

Ⓓ None of the above

d) (8 points) In the power graph below **clearly label** and **shade** in the region on the graph that signifies the **Type II error** $\boldsymbol{\beta}$. Additionally, provide the values of $\boldsymbol{\Delta_0}$ and $\boldsymbol{\Delta_a}$, representing the mean difference under the null hypothesis and the meaningful difference for the alternative hypothesis, respectively.



i. $\Delta_0 = 0$

ii. $\Delta_a = 15$

e) (8 points) Select the appropriate critical value for determining the cutoff value and calculate the cutoff value $\bar{d}_{\text{cutoff}}$. Your answer must indicate both the critical value and the cutoff value.

qnorm(0.01, lower = FALSE) [1] 2.326348	qnorm(0.025, lower = FALSE) [1] 1.959964
qnorm(0.05, lower = FALSE) [1] 1.644854	qnorm(0.95, lower = FALSE) [1] -1.644854

Critical Value: qnorm(0.05, lower = FALSE)

$z = 1.644854$

$$\bar{d}_{\text{cutoff}} = \Delta_0 + z\frac{\sigma}{\sqrt{n}} = 1.644854 \times \frac{37}{\sqrt{15}} = 15.71388$$

f) **(7 points)** Utilize the cutoff value calculated in part **e)** to calculate the power of the test. **Clearly set up the probability to be calculated** and show the mathematical steps required to obtain the power of the test and select the correct code and output for computing the power of this test from the table below.

pnorm(1.239448, lower.tail = FALSE) [1] 0.1075898	pnorm(1.239448, lower.tail = TRUE) [1] 0.8924102
pnorm(0.07472524, lower.tail = FALSE) [1] 0.4702167	pnorm(0.07472524, lower.tail = TRUE) [1] 0.5297833
pnorm(0.3898356, lower.tail = FALSE) [1] 0.3483291	pnorm(0.3898356, lower.tail = TRUE) [1] 0.6516709

Power:

$$P(\bar{D} > \bar{d}_{\text{cutoff}} | \Delta = \Delta_a) = P\left(Z > \frac{\bar{d}_{\text{cutoff}} - \Delta_a}{\frac{\sigma}{\sqrt{n}}}\right) =$$

$$P\left(Z > \frac{15.71388 - 15}{37/\sqrt{15}}\right) = P(Z > 0.07472524) = 0.4702167$$

4. **(20 points)** Suppose weights of undergraduate students from Lumia University come from a minor positively skewed population with an **average weight** of **180 lb** and with a **standard deviation** of **20 lb**. A researcher randomly selects a sample of 40 undergraduate students from this population.

a) **(4 points)** What is the distribution of the mean weights of the 40 undergraduate students? Clearly specify the name of the distribution and its parameters.

The distribution of the mean weights of a random sample of 40 undergraduate students would follow a **normal distribution** with **mean** $\mu_{\bar{X}} = 180$ and **standard deviation of** $\sigma_{\bar{X}} = \frac{20}{\sqrt{40}} = \sqrt{10} = 3.162278$ or $\sigma^2_{\bar{X}} = 10.$

b) **(8 points)** What is the probability of the mean weight of the 40 undergraduate students being **less than 175 lb**? **Clearly set up the probability to be calculated** and show the mathematical steps required to obtain the probability. You may use the following R output in your calculations.

pnorm(-0.25, lower.tail = TRUE) 0.4012937	pnorm(-0.25, lower.tail = FALSE) 0.5987063
pnorm(-1.58, lower.tail = TRUE) 0.05705343	pnorm(-1.58, lower.tail = FALSE) 0.9429466
pnorm(-2.53, lower.tail = TRUE) 0.005703126	pnorm(-2.53, lower.tail = FALSE) 0.9942969

$$P(\bar{X} < 175) = P\left(Z < \frac{175 - 180}{\frac{20}{\sqrt{40}}}\right)$$

$$= P(Z < -1.58) = \mathbf{0.05705343}$$

c) **(8 points)** What is the **99th percentile** of the mean weight of the 40 undergraduate students? You may use the following R output in your calculations. **Show your work.**

Qnorm(0.01/2, lower.tail = FALSE) 2.575829	qnorm(0.01, lower.tail = FALSE) 2.326348
qnorm(0.04/2, lower.tail = FALSE) 2.053749	qnorm(0.04, lower.tail = FALSE) 1.750686

The **99th percentile satisfies:**

$$P(Z < 2.326348) = 0.01$$

$$P(\bar{X} < \bar{x}_{0.99}) = 0.01$$

Where,

$$\bar{x}_{0.99} = \mu_{\bar{X}} + z\sigma_{\bar{X}} = 180 + 2.326348\frac{20}{\sqrt{40}} = 187.3566$$

5. **(32 points)** For all pharmaceutical companies producing pills of Medicine A, federal regulations mandate that the ratio of the average weight to a fixed field standard value must **not exceed 1.2**. For medicine A, this **fixed field standard is 210 mg**.

   a) **(4 points)** Formulate the government regulation requirement as an inequality using the average weight (μ) of the pills. The inequality should be structured to place the average weight of the pills on one side by itself and a numerical value on the other side. How does this value relate to the hypothesis we formulate regarding the average weight of the pills?

$$\frac{\mu}{210} \leq 1.2$$

$$\mu \leq 252$$

The value 252 is the null value for the hypothesis test.

   b) **(4 points)** Placebo-Potion Pharmaceuticals was selected during the initial screening to conduct a formal hypothesis test, with $n = 150$, $\alpha = 0.01$, on whether the average pill weight exceeds the regulatory limit, indicating failure to meet the regulation. State the appropriate **null** and **alternative hypotheses**.

$H_0: \mu \leq 252$
$H_a: \mu > 252$

   c) **(16 points)** Placebo-Potion Pharmaceuticals collected an SRS from their production lines accordingly, and found that the **sample mean** was **256.5 mg** and the **sample standard deviation** was **25 mg**.

      i. **(6 points)** Compute the appropriate **test statistic**. Show all work.

$$t_{TS} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{256.5 - 252}{25/\sqrt{150}} = 2.2045$$

      ii. **(3 points)** Select the appropriate code to compute the ***p*-value** from the table below.

Ⓐ pnorm(test_statistic)	Ⓑ pt(test_statistic, df = 150)	Ⓒ pt(test_statistic, df = 149)
Ⓓ pnorm(test_statistic,lower.tail = FALSE)	Ⓔ pt(test_statistic, df = 150,lower.tail = FALSE)	Ⓕ pt(test_statistic, df = 149,lower.tail = FALSE)
Ⓖ pnorm(**abs**(test_statistic),lower.tail = FALSE)	Ⓗ pt(**abs**(test_statistic), df = 150,lower.tail = FALSE)	Ⓘ pt(**abs**(test_statistic), df = 149,lower.tail = FALSE)

iii. **(9 points)** The $\boldsymbol{p}$**-value** was found to be 0.0145 using the appropriate test statistic and code from above. State the decision and provide a formal conclusion about Placebo-Potion Pharmaceuticals' compliance with the regulation based on this hypothesis test.

The $\boldsymbol{p}$**-value** $= \mathbf{0.0145} > \mathbf{0.01}$ therefore we do not have evidence to reject the null hypothesis.

The data **does not** give support ($\boldsymbol{p}$**-value** = 0.0145) to the claim that the average weight of medicine A at Placebo-Potion Pharmaceuticals exceeds the regulation requirements. In other words, it appears they meet the regulation requirements.

d) **(6 points)** In addition to the hypothesis test, the company was instructed to establish a confidence region for the true mean weight of the pills. Based on the results of the hypothesis test, which confidence region selection aligns most closely with those findings?

i. **(3 points) Type**

Ⓐ Confidence interval

Ⓑ Lower confidence bound

Ⓒ Upper confidence bound

ii. **(3 points) Interval/Bound Values:**

Ⓐ **(251.17, 261.83)**

Ⓑ **(253.17, 259.83)**

Ⓒ **(251.6997, ∞)**

Ⓓ **(253.6997, ∞)**

Ⓔ **(-∞, 261.3003)**

This page is intentionally blank. You may use this as scrap paper.