

STAT 350 Practice Exam 2

Name (Print): _____

Instructor (circle one): Ayu Findsen Sellke Troisi

Class Time: 9:30 AM 11:30 AM 12:30 PM
 1:30 PM 2:30 PM 3:30 PM ONLINE

Instructions:

1. Please write your name on every page.
2. You are expected to uphold the honor code of Purdue University. It is your responsibility to keep your work covered at all times. Anyone cheating on the exam will automatically fail the course, and will be reported to the Office of Dean of Students.
3. Please alert proctors if you observe any cheating during the exam. We highly appreciate it.
4. It is strictly prohibited to smuggle this exam outside. Your exam will be returned to you after it is graded.
5. You may have one double-sided 8.5 in x 11 in crib sheet to take this test. The crib sheet can be handwritten or typed.
6. The only materials that you are allowed are your calculator, writing utensils and erasers and your crib sheet. If you bring any other papers in to the exam, you will get a **zero** on the exam. We will provide scratch paper if you need more room.
7. Leave all your belongings except those permitted for the exam in the front of the room. We are not responsible for any loss.
8. If you share your calculator or use a cell phone, you will get a **zero** on the exam.
9. Breaks (including bathroom breaks) during the exam are **not** allowed. If you leave the exam room, you must turn in your exam and you will not be allowed to come back.
10. You must show **ALL** your work to obtain full credit. An answer without showing any work may result in **zero** credit.
11. All numeric answers should have **two decimal places** except answers from the z-table should have **four decimal places**.
12. If your work is not readable, it will be marked wrong.
13. After you complete the exam, please turn in your crib sheet, tables and any scrap paper along with the exam. Do not staple. Please show your Purdue picture ID.

Your exam is not valid without your signature below.

STUDENT: I attest here that I have followed the instructions above honestly while taking this test and that work submitted is my own, produced without assistance from books, other people, notes other than my own crib sheets, or other aids.

Signature of Student: _____

| | Points Earned |
|-----------------------|---------------|
| Problem 1 (30 points) | |
| Problem 2 (15 points) | |
| Problem 3 (20 points) | |
| Problem 4 (25 points) | |
| Problem 5 (10 points) | |
| Problem 6 (10 points) | |

TOTAL (105 / 100)

1. (30 points) Since 2007, the American Psychological Association has supported an annual nationwide survey to examine stress across the United States. A total of 340 Millennials (18- to 33-year-olds) were asked to indicate their average stress level (on a 10-point scale) during the past month. The mean score was 5.4. Assume that the population standard deviation is 2.3.

(a) (10 points) Find the 95% confidence interval using this sample. Be sure to interpret your result.

[solution] **Since the population standard deviation is known, use Z procedure.**

(3 pts) $z^* = 1.96$ for $C=95\%$

(3 pts) $5.4 \pm 1.96 \cdot 2.3 / \sqrt{340}$
(1 pt) $= 5.4 \pm 0.24$

(1 pt) (5.16, 5.64)

(2 pts) We are 95% confident that the mean stress level of all Millennials is between 5.16 and 5.64

(b) (5 points) What is the margin of error of your 95% confidence interval? If we want the confidence level to be 90%, will the margin of error increase or decrease?

(2 pts) The margin of the 95% CI is 0.24

(3 pts) The margin of error will decrease if we want the confidence level to be 90%

(b) (10 points) How many samples should we take if we want the margin of error to be 0.2 with 95% confidence?

(4 pts) $ME = 1.96 \cdot 2.3 / \sqrt{n} = 0.2$
 $\sqrt{n} = 1.96 \cdot 2.3 / 0.2 = 22.54$

(4 pts) $n = 508.05$

(2 pts) round up $\Rightarrow n = 509$

(c) (5 points) The data for this study are integer values between 1 and 10. Explain why the confidence interval based on the Normal distribution should be a good approximation.

2. (15 points) An environmental group conducted a study to determine whether crows in a certain region were ingesting food containing unhealthy levels of lead. The lead levels of 49 crows in the region were measured and recorded. The mean lead level of the 49 crows in the sample was 4.90 ppm, and the population standard deviation is 0.7 ppm. Carry out a two-sided test of significance that the true mean is 5.2 ppm, with a significance level 0.01. Make sure you follow the four-step procedure.

[Solution]

Since the population standard deviation is known (0.7 ppm), use Z procedure.

(2 pts.) Step 0

μ = the population mean lead level of crows in the region

(2 pts.) Step 1:

$H_0: \mu = 5.2$

$H_a: \mu \neq 5.2$

(2 pts.) Step 2:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{4.9 - 5.2}{\frac{0.7}{\sqrt{49}}} = -3$$

(2 pts.) Step 3:

P-value = $2 * P(Z < -3) = 2(0.0013) = 0.0026$

(2 pts.) Step 4

Reject H_0 because $0.0026 \leq 0.01$

The data strongly supports ($P = 0.0026$) the claim that the population mean lead level of crows in the region is not 5.2.

Without Step 0:

(2 pts.) Step 1

(3 pts.) Step 2

(2 pts.) Step 3

(3 pts.) Step 4

3. (20 points) The College of Science at Purdue University is interested in determining how much time students in that college spend on homework each week. The table below shows the mean and the standard deviation of the amount of times spent each week (in minutes) for a random sample of 20 students in Computer Science and a separate random sample of 20 students in Statistics. The college would like to estimate the difference in mean times spent on homework for all CS and versus all STAT students.

| | Sample Size | Sample Mean | Sample Standard Deviation |
|------------------------|-------------|-------------|---------------------------|
| CS Students | 20 | 282 | 74 |
| Statistics Students | 20 | 164 | 65 |
| Difference (CS - STAT) | 20 | 118 | 75 |

a) (5 points) Should you use the two sample T or the matched pairs T procedure to analyze the data? Explain your answer.

(3 pts) We should use two-sample T procedure,

Explain: (1 pts) Each sample is drawn from a distinct population -- CS students and Stat students

(1 pts) The amount of time spent each week in for CS students in the sample are independent of those from Stat students. That is, there was no pairing in the study.
or no confounding variables mentioned

b) (3 points) Degree of Freedom:

If using the matched pairs T, DF = 19 (fill-in the blank)

If using two sample T procedure, DF = 37.38 from software approximation;

c) (12 points) Construct a 95% confidence interval. Be sure to interpret your interval.

Full credit if the procedure is consistent with (a).

(3 pts) A 95% confidence interval for $\mu_1 - \mu_2$ is given by:

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

(3 pts) The critical value for 95% confidence is $t^* = 2.042$ based on DF = 37

(DF = 30 in the T Table)

(3 pts) The 95% confidence interval for μ is therefore

$$118 \pm 2.042 \sqrt{\frac{74^2}{20} + \frac{65^2}{20}} = 118 \pm 2.042 * 22.024 = 118 \pm 44.97$$

(1 pt) which is the interval (73.03, 162.97) minutes.

(2 pts) We can be 95% **confident** that the difference of the **population** mean times spent on homework for all CS versus all STAT students in the College of Science at Purdue University is between 73 minutes and 163 minutes.

4. (25 points) The weight gain of women during pregnancy has an important effect on the birth weight of their children. If the weight gain is not adequate, the infant is more likely to be

small and will tend to be less healthy. In a study conducted in three countries, weight gains (in kilograms) of women during the third trimester of pregnancy were measured. The results are summarized in the following table:

| Country | N | Mean | StDev |
|---------|-----|------|-------|
| Egypt | 46 | 3.7 | 2.5 |
| Kenya | 111 | 3.1 | 1.8 |
| Mexico | 52 | 2.9 | 1.8 |

Analysis of Variance

| Source | SS | MS | F-Value | P-Value |
|-----------------|--------|------|---------|---------|
| Country (Model) | 17.22 | 8.61 | xxx | 0.1123 |
| Error | 802.89 | xxxx | | |

(a) (5 points) Is it reasonable to use the assumption of equal standard deviation when analyzing these data? Give a reason for your answer.

1pts Yes
4pts $2.5/1.8 < 2$

(b) (5 points) Find the pooled standard deviation. Be sure to state the value of MSE.

| | |
|------|---|
| 1pts | $DFE = 46 + 111 + 52 - 3 = 206$ |
| 2pts | $MSE = SSE / DEF = 802.89 / 206 = 3.90$ |
| 2pts | $s_p = \sqrt{3.90} = 1.97$ |

(c) (2 points) What are the numerator and denominator degrees of freedom for the F statistic?

| | |
|------|------------------------------|
| 2pts | DF (Denominator) = DFE = 206 |
| 3pts | DF (Numerator) = 3 - 1 = 2 |

(d) (3 points) Find the value of the F statistic.

| | |
|------|-------------------------|
| 2pts | $F = MSM / MSE$ |
| 3pts | $F = 8.61 / 3.9 = 2.21$ |

4. (cont'd)

(d) (10 points) Carry out a significance test to compare the mean birth weights for the three countries at a significance level 0.05. Please follow the four step procedure and make a conclusion in context.

Let μ_1 , μ_2 , μ_3 be the population weight gains (in kilograms) of women during the third trimester of pregnancy in Egypt, Kenya, and Mexico respectively.

| | |
|------|---|
| 1pt | $H_0: \mu_1 = \mu_2 = \mu_3$ |
| 2pts | H_1 : not all of the μ_i are equal |
| 1pt | $F(2, 206) = 2.21$ |
| 2pts | P-Value = 0.1123 > 0.05 |
| 2pts | Fail to reject H_0 |
| 2pts | The population weight gains of women during the third trimester of pregnancy in the three countries are the same. |

If using a 5 step process (Findsen):

| | |
|-------|--|
| 2 pt. | Let μ_1 , μ_2 , μ_3 be the population weight gains (in kilograms) of women during the third trimester of pregnancy in Egypt, Kenya, and Mexico respectively. |
| 1pt | $H_0: \mu_1 = \mu_2 = \mu_3$ |
| 1pts | H_1 : not all of the μ_i are equal |
| 1pt | $F(2, 206) = 2.21$ |
| 2pts | P-Value = 0.1123 > 0.05 |
| 1pts | Fail to reject H_0 |
| 2pts | The population weight gains of women during the third trimester of pregnancy in the three countries are the same. |

5. (10 points) A researcher wants to conduct a study to test whether listening to soothing music for 20 minutes helps to reduce diastolic blood pressure in patients with high blood pressure, compared to simply sitting quietly in a noise-free environment for 20 minutes. One hundred patients with high blood pressure at a large medical clinic are available to participate in this study. The null hypothesis for this study is that there is no difference in the mean reduction of diastolic blood pressure for the two treatments and the alternative hypothesis is that the mean reduction in diastolic blood pressure is greater for the music treatment. If the null hypothesis is rejected, the clinic will offer this music therapy as a free service to their patients with high blood pressure. Describe Type I and Type II errors and the consequences of each in the context of this study, and discuss which one you think is more serious.

a) (2 pts) Type I Error Description :

Concluding that soothing music does reduce mean diastolic blood pressure compared to sitting quietly, when in fact it does not.

b) (2 pts) Consequence of Type I Error in the context of this study:

The consequence of this type of error is that the clinic will offer music therapy when it is not effective.

c) (2 pts) Type II Error Description:

Soothing music does reduce diastolic blood pressure compared to sitting quietly, but we fail to detect this and conclude that it does not.

d) (2 pts) Consequence of Type II Error in the context of this study:

The consequence of this type of error is that the clinic will choose not to offer music therapy when it would have been effective.

e) (2 pts) Discussion on which type of error is more serious.

Which type of error is more serious? A case can be made for either type of error, and the student can take either side as long as a reasonable justification is given.

Type I error is more serious because it will cost the clinic money with no benefit

Type II error is more serious because the clinic will miss an opportunity to improve the health and well-being of its patients or the patients will miss out to getting better.

6. (5 points) You must decide which of two discrete distributions a random variable X has. We will call the distributions p_0 and p_1 . Here are the probabilities they assign to the values x of X :

| x | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|-------|-----|-----|-----|-----|-----|-----|-----|
| p_0 | 0.1 | 0.1 | 0.2 | 0.1 | 0.1 | 0.1 | 0.3 |
| p_1 | 0.2 | 0.2 | 0.2 | 0.1 | 0.1 | 0.1 | 0.1 |

You have a single observation on X and wish to test

$$H_0: p_0 \text{ is correct}$$

$$H_a: p_1 \text{ is correct}$$

One possible decision procedure is to reject H_0 only if $X \leq 2$, and accept H_0 if $X \geq 3$.

(a) (2.5 points) Find the probability of a Type I error.

(0.5 pt)

Type I error = reject H_0 when H_0 is actually true

(1 pt)

$P(\text{Type I error}) = P(X \leq 2 \text{ when the true distribution is } p_0)$

(1 pt)

$= P(0) + P(1) + P(2) \text{ [use } p_0]$

$= 0.1 + 0.1 + 0.2 = 0.4$

(b) (2.5 points) Find the probability of a Type II error.

(0.5 pt)

Type II error = fail to reject H_0 when H_0 is actually false

(1.5 pt)

$P(\text{Type II error}) = P(X \leq 2 \text{ when the true distribution is } p_1)$

$= P(0) + P(1) + P(2) \text{ [use } p_1]$

$= 0.2 + 0.2 + 0.2 = 0.6$