Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 026

Name: ANSWER KEY

This take-home exam is due **Monday, April 29** at the beginning of class.

You may use any notes, textbook, or online tools; however, you may not request help from any other human. If you believe a question is ambiguous, unanswerable, or erroneous, please let me know.

You will show your work on the pages with questions. When you are sure of your answers, you will put those answers in the boxes on the first few pages.

Unless you have an objection to doing so, please copy the honor-code text below and sign.

I understand that outside help is NOT allowed on this exam. On my honor, the work herein is my own.

Signature:

1. (a) **LB = 9.14**

   (b) **UB = 10.8**

2. (a) $H_0 : \mu_{\textbf{diff}} = 0$

   (b) $H_{\textbf{A}} : \mu_{\textbf{diff}} \neq 0$

   (c) $t^{\star} = 3.14$

   (d) $SE = 0.945$

   (e) $\mid t_{\textbf{obs}} \mid = 2.995$

   (f) $0.02 < p\textbf{-value} < 0.04$

   (g) **retain**

3. (a) $H_0 : \quad \mu_2 - \mu_1 = 0$

   (b) $H_0 : \quad \mu_2 - \mu_1 \neq 0$

   (c) $t^{\star} = 2.57$

   (d) $SE = 0.207$

   (e) $\mid t_{\textbf{obs}} \mid = 2.61$

   (f) $0.04 < p\textbf{-value} < 0.05$

   (g) **reject**

4. (a) **LB of p CI = 0.631 or** 63.1%

   (b) **UB of p CI = 0.641 or** 64.1%

5. $n \approx 660$

6. (a) $H_0 : \quad p_2 - p_1 = 0$

(b) $H_{\mathbf{A}}: \quad p_2 - p_1 \neq 0$

(c) $z^{\star} = 2.33$

(d) $SE = 0.049$

(e) $\mid z_{\mathbf{obs}} \mid = 2.47$

(f) p-**value** = 0.0136

(g) **reject**

1. **Problem**
   As an ornithologist, you wish to determine the average body mass of *Denrdoica magnolia*. You randomly capture 15 adults of *Denrdoica magnolia*, resulting in a sample mean of 9.96 grams and a sample standard deviation of 1.4 grams. You decide to report a 96% confidence interval.

   (a) Determine the lower bound of the confidence interval.
   (b) Determine the upper bound of the confidence interval.

   Solution
   We are given the sample size, sample mean, sample standard deviation, and confidence level.

$$\begin{aligned} n &= 15 \\ \bar{x} &= 9.96 \\ s &= 1.4 \\ CL &= 0.96 \end{aligned}$$

   Determine the degrees of freedom (because we don't know σ and we are doing inference so we need to use the t distribution).

$$df = n - 1 = 14$$

   Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.96$.

$$t^\star = 2.26$$

   Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.4}{\sqrt{15}} = 0.361$$

   We want to make an inference about the population mean.

$$\mu \;\approx\; \bar{x} \pm t^\star SE$$

   Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (9.96 - 2.26 \times 0.361,\ 9.96 + 2.26 \times 0.361) \\ &= (9.14,\ 10.8) \end{aligned}$$

   We are 96% confident that the population mean is between 9.14 and 10.8.

   (a) Lower bound = 9.14
   (b) Upper bound = 10.8

2. **Problem**
A teacher has 7 students who have each taken two quizzes. Perform a two-tail test with significance level 0.02 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6	student7
quiz 1:	73.5	87.9	54.6	82	81.5	85.5	58.7
quiz 2:	72.8	84.7	50.8	79.9	74.2	86	55.5

(a) State the null hypothesis.
(b) State the alternative hypothesis.
(c) Evaluate the critical value. (The critical value is either $z^{\star}$ or $t^{\star}$. Determine its value.)
(d) Determine the standard error of the relevant sampling distribution.
(e) Evaluate the absolute value of the test statistic. (The test statistic is either z_{obs} or t_{obs}. Determine its absolute value.)
(f) If possible, evaluate the p-value. Otherwise, describe an interval containing the p-value.
(g) Do we reject or retain the null?

Solution
We are given paired data. We are considering a mean of differences. Label the given information.

$$n = 7$$
$$\alpha = 0.02$$

State the hypotheses.

$$H_0 : \quad \mu_{\text{diff}} = 0$$
$$H_A : \quad \mu_{\text{diff}} \neq 0$$

Determine the degrees of freedom.

$$df = n - 1 = 6$$

We determine $t^{\star}$ such that $P(|T| > t^{\star}) = 0.02$.

$$t^{\star} = 3.14$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6	student7
quiz2-quiz1:	-0.7	-3.2	-3.8	-2.1	-7.3	0.5	-3.2

Find the sample mean.

$$\overline{x_{\text{diff}}} = -2.83$$

Find the sample standard deviation.

$$s_{\text{diff}} = 2.5$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 0.945$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{-2.83 - 0}{0.945} = -2.995$$

Compare $|t_{\text{obs}}|$ and $t^{\star}$.

$$|t_{\text{obs}}| < t^{\star}$$

We can determine an interval for the p-value using the t table.

$$0.02 < p\text{-value} < 0.04$$

We conclude that we should retain the null hypothesis.

(a) $H_0: \quad \mu_{\text{diff}} = 0$

(b) $H_A: \quad \mu_{\text{diff}} \neq 0$

(c) $t^{\star} = 3.14$

(d) $SE = 0.945$

(e) $|t_{\text{obs}}| = 2.995$

(f) $0.02 < p\text{-value} < 0.04$

(g) retain the null

3. **Problem**
   You are interested in whether a treatment causes an effect on a continuously measurable attribute. You use a treatment group with 5 cases and a control group with 5 cases. You decide to run a hypothesis test with a significance level of 0.05. Your data is below. Please use 5 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
5.3	4.9
5.4	4.8
6.1	4.6
5.3	5.4
5.5	5.2

   (a) State the null hypothesis.
   (b) State the alternative hypothesis.
   (c) Evaluate the critical value. (The critical value is either $z^\star$ or $t^\star$. Determine its value.)
   (d) Determine the standard error of the relevant sampling distribution.
   (e) Evaluate the absolute value of the test statistic. (The test statistic is either z_{obs} or t_{obs}. Determine its absolute value.)
   (f) If possible, evaluate the p-value. Otherwise, describe an interval containing the p-value.
   (g) Do we reject or retain the null?

   Solution
   We are given unpaired data. We are considering a difference of means. Label the given information.

$$n_1 = 5$$
$$n_2 = 5$$
$$\alpha = 0.05$$

   State the hypotheses.

$$H_0 : \quad \mu_2 - \mu_1 = 0$$
$$H_A : \quad \mu_2 - \mu_1 \neq 0$$

   We are using a two-tail test. Find $t^\star$ such that $P(|T| > t^\star)$ by using a t table.

$$t^\star = 2.57$$

   Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 5.52$$
$$s_1 = 0.335$$
$$\bar{x}_2 = 4.98$$
$$s_2 = 0.319$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(0.335)^2}{5} + \frac{(0.319)^2}{5}} \\ &= 0.207 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(4.98 - 5.52) - (0)}{0.207} \\ &= -2.61 \end{aligned}$$

Compare $|t_{\text{obs}}|$ and $t^\star$.

$$|t_{\text{obs}}| > t^\star$$

We can determine an interval for the p-value using the t table.

$$0.04 < p\text{-value} < 0.05$$

Compare p-value and α.

$$p\text{-value} < \alpha$$

We conclude that we should reject the null hypothesis.

(a) $H_0 : \ \mu_2 - \mu_1 = 0$

(b) $H_A : \ \mu_2 - \mu_1 \neq 0$

(c) $t^\star = 2.57$

(d) $SE = 0.207$

(e) $|t_{\text{obs}}| = 2.61$

(f) $0.04 < p\text{-value} < 0.05$

(g) reject the null

4. **Problem**
   From a very large population, a random sample of 31000 individuals was taken. In that sample, 63.6% were asleep. Determine a 95% confidence interval of the population proportion.

   (a) Find the lower bound of the confidence interval.
   (b) Find the upper bound of the condifence interval.

   Solution
   Determine $z^{\star}$ such that $P(|Z| < z^{\star}) = 0.95$.

$$z^{\star} = 1.96$$

   Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.636)(1-0.636)}{31000}} = 0.00273$$

   Calculate the margin of error.

$$ME = z^{\star}SE = (1.96)(0.00273) = 0.00535$$

   To find the confidence interval's bounds, find the sample proportion plus or minus the margin of error.

$$p \approx \hat{p} \pm ME$$

   Determine the interval.

$$(0.631\ ,\ 0.641)$$

   We are 95% confident that the true population proportion is between 63.1% and 64.1%.

   (a) The lower bound = 0.631, which can also be expressed as 63.1%.
   (b) The upper bound = 0.641, which can also be expressed as 64.1%.

5. **Problem**
   Your boss wants to know what proportion of a very large population is asleep. She also wants to guarantee that the margin of error of a 96% confidence interval will be less than 0.04 (which is 4 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

   Solution
   Determine $z^\star$ such that $P(|Z| < z^\star) = 0.96$.

$$z^\star = 2.05$$

   Determine the maximal standard error.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.04}{2.05} = 0.0195$$

   Calculate n. Because we have no idea what p is, we will use a conservative approach and use $p = 0.5$.

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.0195)^2} = 657.4621959$$

   When determining a necessary sample size, always round up (ceiling).

$$n = 658$$

   Also, for simplicity, you should probably only use about 2 significant digits, while still rounding up.

$$n = 660$$

6. **Problem**
An experiment is run with a treatment group of size 198 and a control group of size 169. The results are summarized in the table below.

	treatment	control
sick	73	42
not sick	125	127

Using a significance level of 0.02, determine whether the treatment causes an effect on the proportion of cases that are sick.

(a) State the null hypothesis.
(b) State the alternative hypothesis.
(c) Evaluate the critical value. (The critical value is either $z^\star$ or $t^\star$. Determine its value.)
(d) Determine the standard error of the relevant sampling distribution.
(e) Evaluate the absolute value of the test statistic. (The test statistic is either z_{obs} or t_{obs}. Determine its absolute value.)
(f) If possible, evaluate the p-value. Otherwise, describe an interval containing the p-value.
(g) Do we reject or retain the null?

Solution
State the hypotheses.

$$H_0: \quad p_2 - p_1 = 0$$
$$H_A: \quad p_2 - p_1 \neq 0$$

Find $z^\star$ such that $P(|Z| > z^\star) = 0.02$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 2.33$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{73}{198} = 0.369$$

$$\hat{p}_2 = \frac{42}{169} = 0.249$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.249 - 0.369 = -0.12$$

Determine the pooled proportion (because the null assumes the population proportions are equal).

$$\hat{p} = \frac{73 + 42}{198 + 169} = 0.313$$

Determine the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}$$

$$= \sqrt{\frac{(0.313)(0.687)}{198} + \frac{(0.313)(0.687)}{169}}$$

$$= 0.0486$$

We can be more specific about what the null hypothesis claims.

$$H_0 : \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0486)$$

We want to describe how unusual our observation is under the null by finding the p-value. To do so, first find the z score.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.249 - 0.369) - 0}{0.0486} \\ &= -2.47 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.47) \\ &= 0.0136 \end{aligned}$$

Compare the p-value to the signficance level.

$$p\text{-value} < \alpha$$

So, we reject the null hypothesis. Thus the difference in proportions is significant.

(a) $H_0 : \ p_2 - p_1 = 0$

(b) $H_A : \ p_2 - p_1 \neq 0$

(c) $z^\star = 2.33$

(d) $SE = 0.0486$

(e) $|z_{obs}| = 2.47$

(f) p-value = 0.0136

(g) reject the null