Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 001

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 = 9.48**

   (b) **UB = 10.3**

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

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

   (c) $t^\star = 2.61$

   (d) $SE = 1.12$

   (e) $\mid t_{\textbf{obs}} \mid = 2.448$

   (f) $0.04 < p$**-value** < 0.05

   (g) **retain**

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

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

   (c) $t^\star = 1.81$

   (d) $SE = 6.48$

   (e) $\mid t_{\textbf{obs}} \mid = 1.78$

   (f) $0.1 < p$**-value** < 0.2

   (g) **retain**

4. (a) **LB of p CI = 0.61 or** 61%

   (b) **UB of p CI = 0.698 or** 69.8%

5. $n \approx 170000$

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

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

(c) $z^{\star} = 2.05$

(d) $SE = 0.052$

(e) $\mid z_{\mathbf{obs}} \mid = 2.24$

(f) p-**value** = 0.025

(g) **reject**

1. **Problem**
   As an ornithologist, you wish to determine the average body mass of *Vireo griseus*. You randomly capture 34 adults of *Vireo griseus*, resulting in a sample mean of 9.9 grams and a sample standard deviation of 0.903 grams. You decide to report a 99% 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.

$$\begin{aligned} n &= 34 \\ \bar{x} &= 9.9 \\ s &= 0.903 \\ CL &= 0.99 \end{aligned}$$

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

$$df = n - 1 = 33$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.99$.

$$t^\star = 2.73$$

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{0.903}{\sqrt{34}} = 0.155$$

We want to make an inference about the population mean.

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

Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE, \ \bar{x} + t^\star SE) \\ &= (9.9 - 2.73 \times 0.155, \ 9.9 - 2.73 \times 0.155) \\ &= (9.48, \ 10.3) \end{aligned}$$

We are 99% confident that the population mean is between 9.48 and 10.3.

(a) Lower bound = 9.48

(b) Upper bound = 10.3

2. **Problem**
A teacher has 7 students who have each taken two quizzes. Perform a two-tail test with significance level 0.04 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6	student7
quiz 1:	50.4	78.9	88.6	59.2	87.7	74	62.8
quiz 2:	49.2	81	92.5	60.6	87.6	79.6	66.8

(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.

$$n = 7$$
$$\alpha = 0.04$$

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 $P(|T| > t^\star) = 0.04$.

$$t^\star = 2.61$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6	student7
quiz2-quiz1:	-1.2	2.1	3.9	1.4	-0.1	5.6	4

Find the sample mean.

$$\overline{x_{\text{diff}}} = 2.24$$

Find the sample standard deviation.

$$s_{\text{diff}} = 2.42$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 0.915$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{2.24 - 0}{0.915} = 2.448$$

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.

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

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) $t^{\star} = 2.61$

(d) $SE = 1.1234353$

(e) $|t_{\text{obs}}| = 2.448$

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

(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 7 cases and a control group with 7 cases. You decide to run a hypothesis test with a significance level of 0.1. Your data is below. Please use 10 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
30	42
54	33
46	24
57	36
61	46
57	65
63	42

(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.

$$n_1 = 7$$
$$n_2 = 7$$
$$\alpha = 0.1$$

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.

$$t^\star = 1.81$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 52.6$$
$$s_1 = 11.4$$
$$\bar{x}_2 = 41.1$$
$$s_2 = 12.8$$

Calculate the standard error.

$$SE = \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$$
$$= \sqrt{\frac{(11.4)^2}{7} + \frac{(12.8)^2}{7}}$$
$$= 6.48$$

Determine the test statistic.

$$t_{\text{obs}} = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE}$$
$$= \frac{(41.1 - 52.6) - (0)}{6.48}$$
$$= -1.78$$

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.

$$0.1 < p\text{-value} < 0.2$$

Compare p-value and α.

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

We conclude that we should retain the null hypothesis.

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

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

(c) $t^{\star} = 1.81$

(d) $SE = 6.48$

(e) $|t_{\text{obs}}| = 1.78$

(f) $0.1 < p\text{-value} < 0.2$

(g) retain the null

4. **Problem**
   From a very large population, a random sample of 940 individuals was taken. In that sample, 65.4% were special. Determine a 99.5% 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 $P(|Z| < z^\star) = 0.995$.

$$z^\star = 2.81$$

Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.654)(1-0.654)}{940}} = 0.0155$$

Calculate the margin of error.

$$ME = z^\star SE = (2.81)(0.0155) = 0.0436$$

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.

$$(0.61\ ,\ 0.698)$$

We are 99.5% confident that the true population proportion is between 61% and 69.8%.

(a) The lower bound = 0.61, which can also be expressed as 61%.
(b) The upper bound = 0.698, which can also be expressed as 69.8%.

5. **Problem**
   Your boss wants to know what proportion of a very large population is cold. She also wants to guarantee that the margin of error of a 90% confidence interval will be less than 0.002 (which is 0.2 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

   Solution
   Determine $z^\star$ such that $P(|Z| < z^\star) = 0.9$.

$$z^\star = 1.64$$

   Determine the maximal standard error.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.002}{1.64} = 0.00122$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.00122)^2} = 167965.600645$$

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

$$n = 167966$$

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

$$n = 170000$$

6. **Problem**
An experiment is run with a treatment group of size 143 and a control group of size 179. The results are summarized in the table below.

	treatment	control
happy	88	131
not happy	55	48

Using a significance level of 0.04, determine whether the treatment causes an effect on the proportion of cases that are happy.

(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 $P(|Z| > z^\star) = 0.04$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 2.05$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{88}{143} = 0.615$$

$$\hat{p}_2 = \frac{131}{179} = 0.732$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.732 - 0.615 = 0.117$$

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

$$\hat{p} = \frac{88 + 131}{143 + 179} = 0.68$$

Determine the standard error.

$$
\begin{aligned}
SE &= \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}} \\
&= \sqrt{\frac{(0.68)(0.32)}{143} + \frac{(0.68)(0.32)}{179}} \\
&= 0.0523
\end{aligned}
$$

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

$$H_0 : \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0523)$$

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.

$$
\begin{aligned}
z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\
&= \frac{(0.732 - 0.615) - 0}{0.0523} \\
&= 2.24
\end{aligned}
$$

Determine the *p*-value.

$$
\begin{aligned}
p\text{-value} &= 2 \cdot \Phi(-|z|) \\
&= 2 \cdot \Phi(-2.24) \\
&= 0.025
\end{aligned}
$$

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) $H_0 : \quad p_2 - p_1 = 0$

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

(c) $z^{\star} = 2.05$

(d) $SE = 0.0523$

(e) $|z_{\text{obs}}| = 2.24$

(f) p-value = 0.025

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 — Exam ID 002

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 = 20.9**

   (b) **UB = 23**

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

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

   (c) $t^{\star} = 1.94$

   (d) $SE = 1.19$

   (e) $\mid t_{\mathbf{obs}} \mid = 1.984$

   (f) $0.05 < p$**-value** < 0.1

   (g) **reject**

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

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

   (c) $t^{\star} = 2.45$

   (d) $SE = 33.2$

   (e) $\mid t_{\mathbf{obs}} \mid = 2.53$

   (f) $0.04 < p$**-value** < 0.05

   (g) **reject**

4. (a) **LB of p CI = 0.142 or** 14.2%

   (b) **UB of p CI = 0.162 or** 16.2%

5. $n \approx 41000$

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

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

(c) $z^{\star} = 1.28$

(d) $SE = 0.149$

(e) $\mid z_{\mathbf{obs}} \mid = 1.36$

(f) p-**value** = 0.1738

(g) **reject**

1. **Problem**
As an ornithologist, you wish to determine the average body mass of *Icterus spurius*. You randomly capture 32 adults of *Icterus spurius*, resulting in a sample mean of 21.99 grams and a sample standard deviation of 1.96 grams. You decide to report a 99.5% 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.

$$\begin{aligned} n &= 32 \\ \bar{x} &= 21.99 \\ s &= 1.96 \\ CL &= 0.995 \end{aligned}$$

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

$$df = n - 1 = 31$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.995$.

$$t^\star = 3.02$$

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.96}{\sqrt{32}} = 0.346$$

We want to make an inference about the population mean.

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

Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (21.99 - 3.02 \times 0.346,\ 21.99 - 3.02 \times 0.346) \\ &= (20.9,\ 23) \end{aligned}$$

We are 99.5% confident that the population mean is between 20.9 and 23.

   (a) Lower bound = 20.9

   (b) Upper bound = 23

2. **Problem**
A teacher has 7 students who have each taken two quizzes. Perform a two-tail test with significance level 0.1 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6	student7
quiz 1:	53	70.5	71.9	68.1	59.5	88.3	52.6
quiz 2:	47.9	66	69.4	64.9	46.7	89.3	54.4

(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.

$$n = 7$$
$$\alpha = 0.1$$

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 $P(|T| > t^\star) = 0.1$.

$$t^\star = 1.94$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6	student7
quiz2-quiz1:	-5.1	-4.5	-2.5	-3.2	-12.8	1	1.8

Find the sample mean.

$$\overline{x_{\text{diff}}} = -3.61$$

Find the sample standard deviation.

$$s_{\text{diff}} = 4.82$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 1.82$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{-3.61 - 0}{1.82} = -1.984$$

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.

$$0.05 < p\text{-value} < 0.1$$

We conclude that we should reject the null hypothesis.

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

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

(c) $t^{\star} = 1.94$

(d) $SE = 1.1928553$

(e) $|t_{\text{obs}}| = 1.984$

(f) $0.05 < p\text{-value} < 0.1$

(g) reject 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 7 cases and a control group with 7 cases. You decide to run a hypothesis test with a significance level of 0.05. Your data is below. Please use 6 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
530	720
510	620
530	550
480	490
510	590
520	700
530	530

(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.

$$n_1 = 7$$
$$n_2 = 7$$
$$\alpha = 0.05$$

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.

$$t^\star = 2.45$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 516$$
$$s_1 = 18.1$$
$$\bar{x}_2 = 600$$
$$s_2 = 86$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(18.1)^2}{7} + \frac{(86)^2}{7}} \\ &= 33.2 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(600 - 516) - (0)}{33.2} \\ &= 2.53 \end{aligned}$$

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) $t^{\star} = 2.45$

(d) $SE = 33.2$

(e) $|t_{\text{obs}}| = 2.53$

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

(g) reject the null

4. **Problem**

   From a very large population, a random sample of 4900 individuals was taken. In that sample, 15.2% were purple. 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.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.152)(1-0.152)}{4900}} = 0.00513$$

   Calculate the margin of error.

$$ME = z^{\star}SE = (1.96)(0.00513) = 0.0101$$

   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.

$$(0.142\,,\ 0.162)$$

   We are 95% confident that the true population proportion is between 14.2% and 16.2%.

   (a) The lower bound = 0.142, which can also be expressed as 14.2%.

   (b) The upper bound = 0.162, which can also be expressed as 16.2%.

5. **Problem**

   Your boss wants to know what proportion of a very large population is broken. She also wants to guarantee that the margin of error of a 99.5% confidence interval will be less than 0.007 (which is 0.7 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

   Solution

   Determine $z^\star$ such that $P(|Z| < z^\star) = 0.995$.

$$z^\star = 2.81$$

   Determine the maximal standard error.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.007}{2.81} = 0.00249$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.00249)^2} = 40321.9302914$$

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

$$n = 40322$$

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

$$n = 41000$$

6. **Problem**
An experiment is run with a treatment group of size 24 and a control group of size 21. The results are summarized in the table below.

	treatment	control
folksy	14	8
not folksy	10	13

Using a significance level of 0.2, determine whether the treatment causes an effect on the proportion of cases that are folksy.

(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 $P(|Z| > z^\star) = 0.2$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 1.28$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{14}{24} = 0.583$$

$$\hat{p}_2 = \frac{8}{21} = 0.381$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.381 - 0.583 = -0.202$$

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

$$\hat{p} = \frac{14 + 8}{24 + 21} = 0.489$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.489)(0.511)}{24} + \frac{(0.489)(0.511)}{21}}$$

$$= 0.149$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\, 0.149)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.381 - 0.583) - 0}{0.149} \\ &= -1.36 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-1.36) \\ &= 0.1738 \end{aligned}$$

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) $H_0: \quad p_2 - p_1 = 0$

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

(c) $z^\star = 1.28$

(d) $SE = 0.149$

(e) $|z_{\text{obs}}| = 1.36$

(f) p-value = 0.1738

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 003

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 = 11.3**

   (b) **UB = 12.3**

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

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

   (c) $t^{\star} = 3$

   (d) $SE = 2.01$

   (e) $\mid t_{\textbf{obs}} \mid = 3.092$

   (f) $0.02 < p$**-value** < 0.04

   (g) **reject**

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

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

   (c) $t^{\star} = 2.23$

   (d) $SE = 5.06$

   (e) $\mid t_{\textbf{obs}} \mid = 2.16$

   (f) $0.05 < p$**-value** < 0.1

   (g) **retain**

4. (a) **LB of p CI = 0.24 or** 24%

   (b) **UB of p CI = 0.258 or** 25.8%

5. $n \approx 20000$

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

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

(c) $z^{\star} = 2.05$

(d) $SE = 0.025$

(e) $\mid z_{\mathbf{obs}} \mid = 2.16$

(f) p-**value** = 0.0308

(g) **reject**

1. **Problem**
   As an ornithologist, you wish to determine the average body mass of *Dendroica coronata*. You randomly capture 13 adults of *Dendroica coronata*, resulting in a sample mean of 11.8 grams and a sample standard deviation of 0.962 grams. You decide to report a 90% 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.

$$\begin{aligned} n &= 13 \\ \bar{x} &= 11.8 \\ s &= 0.962 \\ CL &= 0.9 \end{aligned}$$

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

$$df = n - 1 = 12$$

   Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.9$.

$$t^\star = 1.78$$

   Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{0.962}{\sqrt{13}} = 0.267$$

   We want to make an inference about the population mean.

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

   Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (11.8 - 1.78 \times 0.267,\ 11.8 - 1.78 \times 0.267) \\ &= (11.3,\ 12.3) \end{aligned}$$

   We are 90% confident that the population mean is between 11.3 and 12.3.

   (a) Lower bound = 11.3

   (b) Upper bound = 12.3

2. **Problem**
A teacher has 5 students who have each taken two quizzes. Perform a two-tail test with significance level 0.04 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5
quiz 1:	76.8	78.8	62	55.1	58
quiz 2:	72.8	69.2	56.4	55.3	48.4

(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.

$$n = 5$$
$$\alpha = 0.04$$

State the hypotheses.

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

Determine the degrees of freedom.

$$df = n - 1 = 4$$

We determine $t^{\star}$ such that $P(|T| > t^{\star}) = 0.04$.

$$t^{\star} = 3$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5
quiz2-quiz1:	-4	-9.6	-5.6	0.2	-9.6

Find the sample mean.

$$\overline{x_{\text{diff}}} = -5.72$$

Find the sample standard deviation.

$$s_{\text{diff}} = 4.13$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 1.85$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{-5.72 - 0}{1.85} = -3.092$$

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 reject the null hypothesis.

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

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

(c) $t^\star = 3$

(d) $SE = 2.0133485$

(e) $|t_{\text{obs}}| = 3.092$

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

(g) reject 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 7 cases and a control group with 7 cases. You decide to run a hypothesis test with a significance level of 0.05. Your data is below. Please use 10 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
107	78
96	84
78	83
81	82
113	83
96	87
86	84

(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.

$$n_1 = 7$$
$$n_2 = 7$$
$$\alpha = 0.05$$

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.

$$t^\star = 2.23$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 93.9$$
$$s_1 = 13.1$$
$$\bar{x}_2 = 83$$
$$s_2 = 2.71$$

Calculate the standard error.

$$SE = \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$$
$$= \sqrt{\frac{(13.1)^2}{7} + \frac{(2.71)^2}{7}}$$
$$= 5.06$$

Determine the test statistic.

$$t_{\text{obs}} = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE}$$
$$= \frac{(83 - 93.9) - (0)}{5.06}$$
$$= -2.16$$

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.

$$0.05 < p\text{-value} < 0.1$$

Compare p-value and α.

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

We conclude that we should retain the null hypothesis.

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

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

(c) $t^{\star} = 2.23$

(d) $SE = 5.06$

(e) $|t_{\text{obs}}| = 2.16$

(f) $0.05 < p\text{-value} < 0.1$

(g) retain the null

4. **Problem**
   From a very large population, a random sample of 12000 individuals was taken. In that sample, 24.9% were bitter. Determine a 98% 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 $P(|Z| < z^\star) = 0.98$.

$$z^\star = 2.33$$

   Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.249)(1-0.249)}{12000}} = 0.00395$$

   Calculate the margin of error.

$$ME = z^\star SE = (2.33)(0.00395) = 0.0092$$

   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.

$$(0.24\,,\ 0.258)$$

   We are 98% confident that the true population proportion is between 24% and 25.8%.

   (a) The lower bound = 0.24, which can also be expressed as 24%.
   (b) The upper bound = 0.258, which can also be expressed as 25.8%.

5. **Problem**
   Your boss wants to know what proportion of a very large population is glowing. She also wants to guarantee that the margin of error of a 95% confidence interval will be less than 0.007 (which is 0.7 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

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

$$z^\star = 1.96$$

   Determine the maximal standard error.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.007}{1.96} = 0.00357$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.00357)^2} = 19615.689413$$

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

$$n = 19616$$

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

$$n = 20000$$

6. **Problem**
An experiment is run with a treatment group of size 251 and a control group of size 287. The results are summarized in the table below.

	treatment	control
sorry	30	19
not sorry	221	268

Using a significance level of 0.04, determine whether the treatment causes an effect on the proportion of cases that are sorry.

(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 $P(|Z| > z^\star) = 0.04$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 2.05$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{30}{251} = 0.12$$

$$\hat{p}_2 = \frac{19}{287} = 0.0662$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.0662 - 0.12 = -0.0538$$

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

$$\hat{p} = \frac{30 + 19}{251 + 287} = 0.0911$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.0911)(0.9089)}{251} + \frac{(0.0911)(0.9089)}{287}}$$

$$= 0.0249$$

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

$$H_0 : \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0249)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.0662 - 0.12) - 0}{0.0249} \\ &= -2.16 \end{aligned}$$

Determine the *p*-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.16) \\ &= 0.0308 \end{aligned}$$

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) $H_0 : \quad p_2 - p_1 = 0$

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

(c) $z^\star = 2.05$

(d) $SE = 0.0249$

(e) $|z_{\text{obs}}| = 2.16$

(f) p-value = 0.0308

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 004

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 = 12.4**

   (b) **UB = 13.2**

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

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

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

   (d) $SE = 1.6$

   (e) $\mid t_{\textbf{obs}} \mid = 2.406$

   (f) $0.05 < p$**-value** < 0.1

   (g) **retain**

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

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

   (c) $t^{\star} = 3.25$

   (d) $SE = 0.164$

   (e) $\mid t_{\textbf{obs}} \mid = 3.34$

   (f) $0.005 < p$**-value** < 0.01

   (g) **reject**

4. (a) **LB of p CI = 0.665 or** 66.5%

   (b) **UB of p CI = 0.679 or** 67.9%

5. $n \approx 1700$

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

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

(c) $z^{\star} = 2.58$

(d) $SE = 0.096$

(e) $|\ z_{\mathbf{obs}}\ | = 2.86$

(f) p-**value** = 0.0042

(g) **reject**

1. **Problem**

As an ornithologist, you wish to determine the average body mass of *Vermivora peregrina*. You randomly capture 28 adults of *Vermivora peregrina*, resulting in a sample mean of 12.78 grams and a sample standard deviation of 1.01 grams. You decide to report a 95% 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.

$$n = 28$$
$$\bar{x} = 12.78$$
$$s = 1.01$$
$$CL = 0.95$$

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

$$df = n - 1 = 27$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.95$.

$$t^\star = 2.05$$

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.01}{\sqrt{28}} = 0.191$$

We want to make an inference about the population mean.

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

Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (12.78 - 2.05 \times 0.191,\ 12.78 - 2.05 \times 0.191) \\ &= (12.4,\ 13.2) \end{aligned}$$

We are 95% confident that the population mean is between 12.4 and 13.2.

(a) Lower bound = 12.4

(b) Upper bound = 13.2

2. **Problem**
A teacher has 6 students who have each taken two quizzes. Perform a two-tail test with significance level 0.05 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6
quiz 1:	50.4	83.1	71.6	75.1	63.4	80.9
quiz 2:	46.1	84.4	68.6	64.9	62	75.4

(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.

$$n = 6$$
$$\alpha = 0.05$$

State the hypotheses.

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

Determine the degrees of freedom.

$$df = n - 1 = 5$$

We determine $t^\star$ such that $P(|T| > t^\star) = 0.05$.

$$t^\star = 2.57$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6
quiz2-quiz1:	-4.3	1.3	-3	-10.2	-1.4	-5.5

Find the sample mean.

$$\overline{x_{\text{diff}}} = -3.85$$

Find the sample standard deviation.

$$s_{\text{diff}} = 3.91$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 1.6$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{-3.85 - 0}{1.6} = -2.406$$

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.

$$0.05 < p\text{-value} < 0.1$$

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) $t^\star = 2.57$

(d) $SE = 1.5978527$

(e) $|t_{\text{obs}}| = 2.406$

(f) $0.05 < p\text{-value} < 0.1$

(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 7 cases and a control group with 7 cases. You decide to run a hypothesis test with a significance level of 0.01. Your data is below. Please use 9 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
2.88	2.01
3.04	3.15
2.89	2.19
2.78	2.12
2.7	2.57
2.73	2.04
3.06	2.14

(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.

$$n_1 = 7$$
$$n_2 = 7$$
$$\alpha = 0.01$$

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.

$$t^\star = 3.25$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 2.87$$
$$s_1 = 0.143$$
$$\bar{x}_2 = 2.32$$
$$s_2 = 0.411$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(0.143)^2}{7} + \frac{(0.411)^2}{7}} \\ &= 0.164 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(2.32 - 2.87) - (0)}{0.164} \\ &= -3.34 \end{aligned}$$

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.

$$0.005 < p\text{-value} < 0.01$$

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) $t^\star = 3.25$

(d) $SE = 0.164$

(e) $|t_{\text{obs}}| = 3.34$

(f) $0.005 < p\text{-value} < 0.01$

(g) reject the null

4. **Problem**
   From a very large population, a random sample of 41000 individuals was taken. In that sample, 67.2% were floating. Determine a 99.5% 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 $P(|Z| < z^\star) = 0.995$.

$$z^\star = 2.81$$

Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.672)(1-0.672)}{41000}} = 0.00232$$

Calculate the margin of error.

$$ME = z^\star SE = (2.81)(0.00232) = 0.00652$$

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.

$$(0.665\,,\ 0.679)$$

We are 99.5% confident that the true population proportion is between 66.5% and 67.9%.

(a) The lower bound = 0.665, which can also be expressed as 66.5%.

(b) The upper bound = 0.679, which can also be expressed as 67.9%.

5. **Problem**

   Your boss wants to know what proportion of a very large population is angry. She also wants to guarantee that the margin of error of a 90% confidence interval will be less than 0.02 (which is 2 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

   Solution

   Determine $z^\star$ such that $P(|Z| < z^\star) = 0.9$.

$$z^\star = 1.64$$

   Determine the maximal standard error.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.02}{1.64} = 0.0122$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.0122)^2} = 1679.6560064$$

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

$$n = 1680$$

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

$$n = 1700$$

6. **Problem**
An experiment is run with a treatment group of size 71 and a control group of size 44. The results are summarized in the table below.

	treatment	control
reclusive	29	30
not reclusive	42	14

Using a significance level of 0.01, determine whether the treatment causes an effect on the proportion of cases that are reclusive.

(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.

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

Find $z^\star$ such that $P(|Z| > z^\star) = 0.01$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 2.58$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{29}{71} = 0.408$$

$$\hat{p}_2 = \frac{30}{44} = 0.682$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.682 - 0.408 = 0.274$$

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

$$\hat{p} = \frac{29 + 30}{71 + 44} = 0.513$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.513)(0.487)}{71} + \frac{(0.513)(0.487)}{44}}$$

$$= 0.0959$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0959)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.682 - 0.408) - 0}{0.0959} \\ &= 2.86 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.86) \\ &= 0.0042 \end{aligned}$$

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) $H_0: \quad p_2 - p_1 = 0$

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

(c) $z^\star = 2.58$

(d) $SE = 0.0959$

(e) $|z_{\text{obs}}| = 2.86$

(f) p-value $= 0.0042$

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 — Exam ID 005

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 = 14.4**

   (b) **UB = 16.1**

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

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

   (c) $t^{\star} = 2.02$

   (d) $SE = 1.36$

   (e) $\mid t_{\textbf{obs}} \mid = 1.872$

   (f) $0.1 <$ p**-value** < 1

   (g) **retain**

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

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

   (c) $t^{\star} = 3.5$

   (d) $SE = 4.06$

   (e) $\mid t_{\textbf{obs}} \mid = 3.45$

   (f) $0.01 <$ p**-value** < 0.02

   (g) **retain**

4. (a) **LB of p CI = 0.609 or** 60.9%

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

5. $n \approx 4200$

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

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

(c) $z^\star = 2.81$

(d) $SE = 0.046$

(e) $\mid z_{\mathbf{obs}} \mid = 3.03$

(f) p**-value** = 0.0024

(g) **reject**

1. **Problem**
   As an ornithologist, you wish to determine the average body mass of *Passerina cyanea*. You randomly capture 35 adults of *Passerina cyanea*, resulting in a sample mean of 15.25 grams and a sample standard deviation of 1.93 grams. You decide to report a 99% 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.

$$\begin{aligned} n &= 35 \\ \bar{x} &= 15.25 \\ s &= 1.93 \\ CL &= 0.99 \end{aligned}$$

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

$$df = n - 1 = 34$$

   Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.99$.

$$t^\star = 2.73$$

   Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.93}{\sqrt{35}} = 0.326$$

   We want to make an inference about the population mean.

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

   Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (15.25 - 2.73 \times 0.326,\ 15.25 - 2.73 \times 0.326) \\ &= (14.4,\ 16.1) \end{aligned}$$

   We are 99% confident that the population mean is between 14.4 and 16.1.

   (a) Lower bound = 14.4

   (b) Upper bound = 16.1

2. **Problem**
A teacher has 6 students who have each taken two quizzes. Perform a two-tail test with significance level 0.1 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6
quiz 1:	63.5	74.3	62.7	56	70.4	85.9
quiz 2:	63	77.6	65.3	56.1	70.5	87.3

(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.

$$n = 6$$
$$\alpha = 0.1$$

State the hypotheses.

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

Determine the degrees of freedom.

$$df = n - 1 = 5$$

We determine $t^\star$ such that $P(|T| > t^\star) = 0.1$.

$$t^\star = 2.02$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6
quiz2-quiz1:	-0.5	3.3	2.6	0.1	0.1	1.4

Find the sample mean.

$$\overline{x_{\text{diff}}} = 1.17$$

Find the sample standard deviation.

$$s_{\text{diff}} = 1.53$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 0.625$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{1.17 - 0}{0.625} = 1.872$$

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.

$$0.1 < p\text{-value} < 1$$

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) $t^{\star} = 2.02$

(d) $SE = 1.3614515$

(e) $|t_{\text{obs}}| = 1.872$

(f) $0.1 < p\text{-value} < 1$

(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 7 cases and a control group with 7 cases. You decide to run a hypothesis test with a significance level of 0.01. Your data is below. Please use 7 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
89	92
86	104
83	106
87	87
83	118
84	95
86	94

(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.

$$n_1 = 7$$
$$n_2 = 7$$
$$\alpha = 0.01$$

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.

$$t^\star = 3.5$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 85.4$$
$$s_1 = 2.23$$
$$\bar{x}_2 = 99.4$$
$$s_2 = 10.5$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(2.23)^2}{7} + \frac{(10.5)^2}{7}} \\ &= 4.06 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(99.4 - 85.4) - (0)}{4.06} \\ &= 3.45 \end{aligned}$$

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.

$$0.01 < p\text{-value} < 0.02$$

Compare p-value and α.

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

We conclude that we should retain the null hypothesis.

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

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

(c) $t^{\star} = 3.5$

(d) $SE = 4.06$

(e) $|t_{\text{obs}}| = 3.45$

(f) $0.01 < p\text{-value} < 0.02$

(g) retain the null

4. **Problem**
   From a very large population, a random sample of 5200 individuals was taken. In that sample, 62.5% were blue. Determine a 98% 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 $P(|Z| < z^\star) = 0.98$.

$$z^\star = 2.33$$

   Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.625)(1-0.625)}{5200}} = 0.00671$$

   Calculate the margin of error.

$$ME = z^\star SE = (2.33)(0.00671) = 0.0156$$

   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.

$$(0.609\,,\ 0.641)$$

   We are 98% confident that the true population proportion is between 60.9% and 64.1%.

   (a) The lower bound = 0.609, which can also be expressed as 60.9%.
   (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 angry. She also wants to guarantee that the margin of error of a 99% confidence interval will be less than 0.02 (which is 2 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

   Solution

   Determine $z^\star$ such that $P(|Z| < z^\star) = 0.99$.

$$z^\star = 2.58$$

   Determine the maximal standard error.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.02}{2.58} = 0.00775$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.00775)^2} = 4162.3309053$$

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

$$n = 4163$$

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

$$n = 4200$$

6. **Problem**
An experiment is run with a treatment group of size 247 and a control group of size 225. The results are summarized in the table below.

	treatment	control
reclusive	122	80
not reclusive	125	145

Using a significance level of 0.005, determine whether the treatment causes an effect on the proportion of cases that are reclusive.

(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 $P(|Z| > z^\star) = 0.005$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 2.81$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{122}{247} = 0.494$$

$$\hat{p}_2 = \frac{80}{225} = 0.356$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.356 - 0.494 = -0.138$$

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

$$\hat{p} = \frac{122 + 80}{247 + 225} = 0.428$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.428)(0.572)}{247} + \frac{(0.428)(0.572)}{225}}$$

$$= 0.0456$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0456)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.356 - 0.494) - 0}{0.0456} \\ &= -3.03 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-3.03) \\ &= 0.0024 \end{aligned}$$

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) $H_0: \quad p_2 - p_1 = 0$

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

(c) $z^\star = 2.81$

(d) $SE = 0.0456$

(e) $|z_{\text{obs}}| = 3.03$

(f) p-value = 0.0024

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 — Exam ID 006

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 = 42.4**

   (b) **UB = 50.1**

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

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

   (c) $t^\star = 2.76$

   (d) $SE = 1.5$

   (e) $\mid t_{\textbf{obs}} \mid = 2.706$

   (f) $0.04 < p$**-value** < 0.05

   (g) **retain**

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

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

   (c) $t^\star = 3.11$

   (d) $SE = 9.96$

   (e) $\mid t_{\textbf{obs}} \mid = 3.08$

   (f) $0.01 < p$**-value** < 0.02

   (g) **retain**

4. (a) **LB of p CI = 0.937 or** 93.7%

   (b) **UB of p CI = 0.949 or** 94.9%

5. $n \approx 39000$

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) $SE = 0.041$

(e) $\mid z_{\mathbf{obs}} \mid = 2.52$

(f) p-**value** = 0.0118

(g) **reject**

1. **Problem**
As an ornithologist, you wish to determine the average body mass of *Agelaius Phoeniceus*. You randomly capture 24 adults of *Agelaius Phoeniceus*, resulting in a sample mean of 46.27 grams and a sample standard deviation of 6.72 grams. You decide to report a 99% 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.

$$\begin{aligned} n &= 24 \\ \bar{x} &= 46.27 \\ s &= 6.72 \\ CL &= 0.99 \end{aligned}$$

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

$$df = n - 1 = 23$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.99$.

$$t^\star = 2.81$$

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{6.72}{\sqrt{24}} = 1.37$$

We want to make an inference about the population mean.

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

Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (46.27 - 2.81 \times 1.37,\ 46.27 - 2.81 \times 1.37) \\ &= (42.4,\ 50.1) \end{aligned}$$

We are 99% confident that the population mean is between 42.4 and 50.1.

   (a) Lower bound = 42.4

   (b) Upper bound = 50.1

2. **Problem**
A teacher has 6 students who have each taken two quizzes. Perform a two-tail test with significance level 0.04 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6
quiz 1:	86.1	54.2	60.5	59.8	57	69.8
quiz 2:	87.9	50.4	54.6	52.7	55.6	63

(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.

$$n = 6$$
$$\alpha = 0.04$$

State the hypotheses.

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

Determine the degrees of freedom.

$$df = n - 1 = 5$$

We determine $t^\star$ such that $P(|T| > t^\star) = 0.04$.

$$t^\star = 2.76$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6
quiz2-quiz1:	1.8	-3.8	-5.9	-7.1	-1.4	-6.8

Find the sample mean.

$$\overline{x_{\text{diff}}} = -3.87$$

Find the sample standard deviation.

$$s_{\text{diff}} = 3.5$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 1.43$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{-3.87 - 0}{1.43} = -2.706$$

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.

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

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) $t^{\star} = 2.76$

(d) $SE = 1.496348$

(e) $|t_{\text{obs}}| = 2.706$

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

(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 7 cases and a control group with 7 cases. You decide to run a hypothesis test with a significance level of 0.01. Your data is below. Please use 11 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
108	124
60	130
92	125
101	103
90	76
82	134
78	132

(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.

$$n_1 = 7$$
$$n_2 = 7$$
$$\alpha = 0.01$$

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.

$$t^{\star} = 3.11$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 87.3$$
$$s_1 = 15.8$$
$$\bar{x}_2 = 118$$
$$s_2 = 21.1$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(15.8)^2}{7} + \frac{(21.1)^2}{7}} \\ &= 9.96 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(118 - 87.3) - (0)}{9.96} \\ &= 3.08 \end{aligned}$$

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.

$$0.01 < p\text{-value} < 0.02$$

Compare p-value and α.

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

We conclude that we should retain the null hypothesis.

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

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

(c) $t^\star = 3.11$

(d) $SE = 9.96$

(e) $|t_{\text{obs}}| = 3.08$

(f) $0.01 < p\text{-value} < 0.02$

(g) retain the null

4. **Problem**
   From a very large population, a random sample of 6500 individuals was taken. In that sample, 94.3% were angry. Determine a 96% 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 $P(|Z| < z^\star) = 0.96$.

   $$z^\star = 2.05$$

   Calculate the standard error.

   $$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.943)(1-0.943)}{6500}} = 0.00288$$

   Calculate the margin of error.

   $$ME = z^\star SE = (2.05)(0.00288) = 0.0059$$

   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.

   $$(0.937\,,\ 0.949)$$

   We are 96% confident that the true population proportion is between 93.7% and 94.9%.

   (a) The lower bound = 0.937, which can also be expressed as 93.7%.
   (b) The upper bound = 0.949, which can also be expressed as 94.9%.

5. **Problem**
   Your boss wants to know what proportion of a very large population is shiny. She also wants to guarantee that the margin of error of a 95% confidence interval will be less than 0.005 (which is 0.5 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

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

$$z^\star = 1.96$$

   Determine the maximal standard error.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.005}{1.96} = 0.00255$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.00255)^2} = 38446.7512495$$

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

$$n = 38447$$

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

$$n = 39000$$

6. **Problem**
An experiment is run with a treatment group of size 181 and a control group of size 205. The results are summarized in the table below.

	treatment	control
special	134	173
not special	47	32

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

(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.

$$\hat{p}_1 = \frac{134}{181} = 0.74$$

$$\hat{p}_2 = \frac{173}{205} = 0.844$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.844 - 0.74 = 0.104$$

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

$$\hat{p} = \frac{134 + 173}{181 + 205} = 0.795$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.795)(0.205)}{181} + \frac{(0.795)(0.205)}{205}}$$

$$= 0.0412$$

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

$$H_0 : \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0412)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.844 - 0.74) - 0}{0.0412} \\ &= 2.52 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.52) \\ &= 0.0118 \end{aligned}$$

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) $H_0 : \quad p_2 - p_1 = 0$

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

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

(d) $SE = 0.0412$

(e) $|z_{\text{obs}}| = 2.52$

(f) p-value = 0.0118

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 007

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 = 9.3**

   (b) **UB = 10.7**

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

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

   (c) $t^{\star} = 2.13$

   (d) $SE = 1.68$

   (e) $\mid t_{\textbf{obs}} \mid = 1.957$

   (f) $0.1 <$ ***p*-value** < 1

   (g) **retain**

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

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

   (c) $t^{\star} = 1.94$

   (d) $SE = 112$

   (e) $\mid t_{\textbf{obs}} \mid = 1.68$

   (f) $0.1 <$ ***p*-value** < 0.2

   (g) **retain**

4. (a) **LB of p CI = 0.587 or** 58.7%

   (b) **UB of p CI = 0.609 or** 60.9%

5. $n \approx 4200$

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

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

(c) $z^{\star} = 1.28$

(d) $SE = 0.136$

(e) $\mid z_{\mathbf{obs}} \mid = 1.47$

(f) p**-value** = 0.1416

(g) **reject**

1. **Problem**
   As an ornithologist, you wish to determine the average body mass of *Wilsonia citrina*. You randomly capture 31 adults of *Wilsonia citrina*, resulting in a sample mean of 9.99 grams and a sample standard deviation of 1.88 grams. You decide to report a 95% 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.

$$\begin{aligned} n &= 31 \\ \bar{x} &= 9.99 \\ s &= 1.88 \\ CL &= 0.95 \end{aligned}$$

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

$$df = n - 1 = 30$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.95$.

$$t^\star = 2.04$$

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.88}{\sqrt{31}} = 0.338$$

We want to make an inference about the population mean.

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

Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (9.99 - 2.04 \times 0.338,\ 9.99 - 2.04 \times 0.338) \\ &= (9.3,\ 10.7) \end{aligned}$$

We are 95% confident that the population mean is between 9.3 and 10.7.

   (a) Lower bound = 9.3

   (b) Upper bound = 10.7

2. **Problem**
A teacher has 5 students who have each taken two quizzes. Perform a two-tail test with significance level 0.1 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5
quiz 1:	82.6	51	68.8	62.7	63.4
quiz 2:	88.1	62.4	73	59.4	68.5

(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.

$$n = 5$$
$$\alpha = 0.1$$

State the hypotheses.

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

Determine the degrees of freedom.

$$df = n - 1 = 4$$

We determine $t^\star$ such that $P(|T| > t^\star) = 0.1$.

$$t^\star = 2.13$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5
quiz2-quiz1:	5.5	11.4	4.2	-3.3	5.1

Find the sample mean.

$$\overline{x_{\text{diff}}} = 4.58$$

Find the sample standard deviation.

$$s_{\text{diff}} = 5.24$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 2.34$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{4.58 - 0}{2.34} = 1.957$$

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.

$$0.1 < p\text{-value} < 1$$

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) $t^\star = 2.13$

(d) $SE = 1.6812805$

(e) $|t_{\text{obs}}| = 1.957$

(f) $0.1 < p\text{-value} < 1$

(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 0.1. Your data is below. Please use 6 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
740	610
1030	990
1280	700
950	940
1010	820

(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.

$$n_1 = 5$$
$$n_2 = 5$$
$$\alpha = 0.1$$

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.

$$t^\star = 1.94$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 1000$$
$$s_1 = 193$$
$$\bar{x}_2 = 812$$
$$s_2 = 159$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(193)^2}{5} + \frac{(159)^2}{5}} \\ &= 112 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(812 - 1000) - (0)}{112} \\ &= -1.68 \end{aligned}$$

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.

$$0.1 < p\text{-value} < 0.2$$

Compare p-value and α.

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

We conclude that we should retain the null hypothesis.

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

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

(c) $t^{\star} = 1.94$

(d) $SE = 112$

(e) $|t_{\text{obs}}| = 1.68$

(f) $0.1 < p\text{-value} < 0.2$

(g) retain the null

4. **Problem**
From a very large population, a random sample of 5400 individuals was taken. In that sample, 59.8% were asleep. Determine a 90% 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 $P(|Z| < z^{\star}) = 0.9$.

$$z^{\star} = 1.64$$

Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.598)(1-0.598)}{5400}} = 0.00667$$

Calculate the margin of error.

$$ME = z^{\star}SE = (1.64)(0.00667) = 0.0109$$

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.

$$(0.587\,,\ 0.609)$$

We are 90% confident that the true population proportion is between 58.7% and 60.9%.

 (a) The lower bound = 0.587, which can also be expressed as 58.7%.
 (b) The upper bound = 0.609, which can also be expressed as 60.9%.

5. **Problem**

   Your boss wants to know what proportion of a very large population is special. She also wants to guarantee that the margin of error of a 99% confidence interval will be less than 0.02 (which is 2 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

   Solution

   Determine $z^{\star}$ such that $P(|Z| < z^{\star}) = 0.99$.

   $$z^{\star} = 2.58$$

   Determine the maximal standard error.

   $$ME = z^{\star} SE$$

   $$SE = \frac{ME}{z^{\star}} = \frac{0.02}{2.58} = 0.00775$$

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

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

   $$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.00775)^2} = 4162.3309053$$

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

   $$n = 4163$$

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

   $$n = 4200$$

6. **Problem**

An experiment is run with a treatment group of size 19 and a control group of size 41. The results are summarized in the table below.

	treatment	control
happy	5	19
not happy	14	22

Using a significance level of 0.2, determine whether the treatment causes an effect on the proportion of cases that are happy.

(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 $P(|Z| > z^\star) = 0.2$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 1.28$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{5}{19} = 0.263$$

$$\hat{p}_2 = \frac{19}{41} = 0.463$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.463 - 0.263 = 0.2$$

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

$$\hat{p} = \frac{5 + 19}{19 + 41} = 0.4$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.4)(0.6)}{19} + \frac{(0.4)(0.6)}{41}}$$

$$= 0.136$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0, 0.136)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.463 - 0.263) - 0}{0.136} \\ &= 1.47 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-1.47) \\ &= 0.1416 \end{aligned}$$

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) $H_0: \quad p_2 - p_1 = 0$

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

(c) $z^\star = 1.28$

(d) $SE = 0.136$

(e) $|z_{\text{obs}}| = 1.47$

(f) p-value = 0.1416

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 008

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 = 22.3**

   (b) **UB = 24.1**

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

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

   (c) $t^{\star} = 3.36$

   (d) $SE = 0.866$

   (e) $\mid t_{\textbf{obs}} \mid = 3.238$

   (f) $0.02 <$ p**-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) $t^{\star} = 2.36$

   (d) $SE = 50.8$

   (e) $\mid t_{\textbf{obs}} \mid = 2.15$

   (f) $0.05 <$ p**-value** < 0.1

   (g) **retain**

4. (a) **LB of p CI = 0.435 or** 43.5%

   (b) **UB of p CI = 0.491 or** 49.1%

5. $n \approx 21000$

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

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

(c) $z^{\star} = 1.64$

(d) $SE = 0.055$

(e) $\mid z_{\mathbf{obs}} \mid = 1.9$

(f) p-**value** = 0.0574

(g) **reject**

1. **Problem**
   As an ornithologist, you wish to determine the average body mass of *Seiurus aurocapillus*. You randomly capture 23 adults of *Seiurus aurocapillus*, resulting in a sample mean of 23.17 grams and a sample standard deviation of 1.4 grams. You decide to report a 99.5% 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.

$$\begin{aligned} n &= 23 \\ \bar{x} &= 23.17 \\ s &= 1.4 \\ CL &= 0.995 \end{aligned}$$

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

$$df = n - 1 = 22$$

   Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.995$.

$$t^\star = 3.12$$

   Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.4}{\sqrt{23}} = 0.292$$

   We want to make an inference about the population mean.

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

   Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (23.17 - 3.12 \times 0.292,\ 23.17 - 3.12 \times 0.292) \\ &= (22.3,\ 24.1) \end{aligned}$$

   We are 99.5% confident that the population mean is between 22.3 and 24.1.

   (a) Lower bound = 22.3

   (b) Upper bound = 24.1

2. **Problem**
A teacher has 6 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
quiz 1:	66.5	64	74.6	81.8	69.9	69.5
quiz 2:	65.6	68.5	77.3	84.2	74.6	75.7

(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.

$$n = 6$$
$$\alpha = 0.02$$

State the hypotheses.

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

Determine the degrees of freedom.

$$df = n - 1 = 5$$

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

$$t^\star = 3.36$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6
quiz2-quiz1:	-0.9	4.5	2.7	2.4	4.7	6.2

Find the sample mean.

$$\overline{x_{\text{diff}}} = 3.27$$

Find the sample standard deviation.

$$s_{\text{diff}} = 2.47$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 1.01$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{3.27 - 0}{1.01} = 3.238$$

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) $t^\star = 3.36$

(d) $SE = 0.8656677$

(e) $|t_{\text{obs}}| = 3.238$

(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 6 cases and a control group with 6 cases. You decide to run a hypothesis test with a significance level of 0.05. Your data is below. Please use 7 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
580	400
490	510
570	500
550	530
490	330
480	240

(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.

$$n_1 = 6$$
$$n_2 = 6$$
$$\alpha = 0.05$$

State the hypotheses.

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

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

$$t^\star = 2.36$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 527$$
$$s_1 = 45$$
$$\bar{x}_2 = 418$$
$$s_2 = 116$$

Calculate the standard error.

$$SE = \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$$
$$= \sqrt{\frac{(45)^2}{6} + \frac{(116)^2}{6}}$$
$$= 50.8$$

Determine the test statistic.

$$t_{\text{obs}} = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE}$$
$$= \frac{(418 - 527) - (0)}{50.8}$$
$$= -2.15$$

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.

$$0.05 < p\text{-value} < 0.1$$

Compare p-value and α.

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

We conclude that we should retain the null hypothesis.

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

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

(c) $t^\star = 2.36$

(d) $SE = 50.8$

(e) $|t_{\text{obs}}| = 2.15$

(f) $0.05 < p\text{-value} < 0.1$

(g) retain the null

4. **Problem**
   From a very large population, a random sample of 1200 individuals was taken. In that sample, 46.3% were glowing. 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.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.463)(1-0.463)}{1200}} = 0.0144$$

   Calculate the margin of error.

$$ME = z^\star SE = (1.96)(0.0144) = 0.0282$$

   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.

$$(0.435\,,\ 0.491)$$

   We are 95% confident that the true population proportion is between 43.5% and 49.1%.

   (a) The lower bound = 0.435, which can also be expressed as 43.5%.
   (b) The upper bound = 0.491, which can also be expressed as 49.1%.

5. **Problem**
   Your boss wants to know what proportion of a very large population is messy. She also wants to guarantee that the margin of error of a 99% confidence interval will be less than 0.009 (which is 0.9 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

   Solution
   Determine $z^\star$ such that $P(|Z| < z^\star) = 0.99$.

$$z^\star = 2.58$$

   Determine the maximal standard error.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.009}{2.58} = 0.00349$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.00349)^2} = 20525.2830437$$

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

$$n = 20526$$

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

$$n = 21000$$

6. **Problem**
An experiment is run with a treatment group of size 132 and a control group of size 176. The results are summarized in the table below.

	treatment	control
sick	55	55
not sick	77	121

Using a significance level of 0.1, 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 $P(|Z| > z^\star) = 0.1$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 1.64$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{55}{132} = 0.417$$

$$\hat{p}_2 = \frac{55}{176} = 0.312$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.312 - 0.417 = -0.105$$

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

$$\hat{p} = \frac{55 + 55}{132 + 176} = 0.357$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.357)(0.643)}{132} + \frac{(0.357)(0.643)}{176}}$$

$$= 0.0552$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0552)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.312 - 0.417) - 0}{0.0552} \\ &= -1.9 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-1.9) \\ &= 0.0574 \end{aligned}$$

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) $H_0: \quad p_2 - p_1 = 0$

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

(c) $z^\star = 1.64$

(d) $SE = 0.0552$

(e) $|z_{\text{obs}}| = 1.9$

(f) p-value $= 0.0574$

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 009

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 = 20.9**

   (b) **UB = 22.9**

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

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

   (c) $t^{\star} = 2.13$

   (d) $SE = 1.11$

   (e) $\mid t_{\textbf{obs}} \mid = 2.042$

   (f) $0.1 <$ p**-value** < 1

   (g) **retain**

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

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

   (c) $t^{\star} = 3.17$

   (d) $SE = 31.2$

   (e) $\mid t_{\textbf{obs}} \mid = 3.46$

   (f) $0.005 <$ p**-value** < 0.01

   (g) **reject**

4. (a) **LB of p CI = 0.241 or** 24.1%

   (b) **UB of p CI = 0.249 or** 24.9%

5. $n \approx 11000$

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

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

(c) $z^{\star} = 1.28$

(d) $SE = 0.033$

(e) $\mid z_{\mathbf{obs}} \mid = 1.32$

(f) p-**value** $= 0.1868$

(g) **reject**

1. **Problem**
   As an ornithologist, you wish to determine the average body mass of *Icterus spurius*. You randomly capture 26 adults of *Icterus spurius*, resulting in a sample mean of 21.92 grams and a sample standard deviation of 2.05 grams. You decide to report a 98% 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.

$$\begin{aligned} n &= 26 \\ \bar{x} &= 21.92 \\ s &= 2.05 \\ CL &= 0.98 \end{aligned}$$

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

$$df = n - 1 = 25$$

   Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.98$.

$$t^\star = 2.49$$

   Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{2.05}{\sqrt{26}} = 0.402$$

   We want to make an inference about the population mean.

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

   Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (21.92 - 2.49 \times 0.402,\ 21.92 - 2.49 \times 0.402) \\ &= (20.9,\ 22.9) \end{aligned}$$

   We are 98% confident that the population mean is between 20.9 and 22.9.

   (a) Lower bound = 20.9

   (b) Upper bound = 22.9

2. **Problem**
A teacher has 5 students who have each taken two quizzes. Perform a two-tail test with significance level 0.1 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5
quiz 1:	83.2	85.9	57	61.5	56.3
quiz 2:	80.5	90.1	61.9	64.9	61.1

(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.

$$n = 5$$
$$\alpha = 0.1$$

State the hypotheses.

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

Determine the degrees of freedom.

$$df = n - 1 = 4$$

We determine $t^{\star}$ such that $P(|T| > t^{\star}) = 0.1$.

$$t^{\star} = 2.13$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5
quiz2-quiz1:	-2.7	4.2	4.9	3.4	4.8

Find the sample mean.

$$\overline{x_{\text{diff}}} = 2.92$$

Find the sample standard deviation.

$$s_{\text{diff}} = 3.2$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 1.43$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{2.92 - 0}{1.43} = 2.042$$

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.

$$0.1 < p\text{-value} < 1$$

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) $t^{\star} = 2.13$

(d) $SE = 1.109171$

(e) $|t_{\text{obs}}| = 2.042$

(f) $0.1 < p\text{-value} < 1$

(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 7 cases and a control group with 7 cases. You decide to run a hypothesis test with a significance level of 0.01. Your data is below. Please use 10 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
700	490
720	600
720	660
650	500
700	690
720	630
670	550

(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.

$$n_1 = 7$$
$$n_2 = 7$$
$$\alpha = 0.01$$

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.

$$t^\star = 3.17$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 697$$
$$s_1 = 27.5$$
$$\bar{x}_2 = 589$$
$$s_2 = 77.8$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(27.5)^2}{7} + \frac{(77.8)^2}{7}} \\ &= 31.2 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(589 - 697) - (0)}{31.2} \\ &= -3.46 \end{aligned}$$

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.

$$0.005 < p\text{-value} < 0.01$$

Compare p-value and α.

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

We conclude that we should reject the null hypothesis.

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

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

(c) $t^\star = 3.17$

(d) $SE = 31.2$

(e) $|t_{\text{obs}}| = 3.46$

(f) $0.005 < p\text{-value} < 0.01$

(g) reject the null

4. **Problem**
From a very large population, a random sample of 46000 individuals was taken. In that sample, 24.5% were blue. 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.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.245)(1-0.245)}{46000}} = 0.00201$$

Calculate the margin of error.

$$ME = z^\star SE = (1.96)(0.00201) = 0.00394$$

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.

$$(0.241\ ,\ 0.249)$$

We are 95% confident that the true population proportion is between 24.1% and 24.9%.

   (a) The lower bound = 0.241, which can also be expressed as 24.1%.
   (b) The upper bound = 0.249, which can also be expressed as 24.9%.

5. **Problem**
   Your boss wants to know what proportion of a very large population is shiny. She also wants to guarantee that the margin of error of a 96% confidence interval will be less than 0.01 (which is 1 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.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.01}{2.05} = 0.00488$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.00488)^2} = 10497.8500403$$

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

$$n = 10498$$

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

$$n = 11000$$

6. **Problem**
An experiment is run with a treatment group of size 258 and a control group of size 216. The results are summarized in the table below.

	treatment	control
green	45	28
not green	213	188

Using a significance level of 0.2, determine whether the treatment causes an effect on the proportion of cases that are green.

(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 $P(|Z| > z^\star) = 0.2$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 1.28$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{45}{258} = 0.174$$

$$\hat{p}_2 = \frac{28}{216} = 0.13$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.13 - 0.174 = -0.044$$

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

$$\hat{p} = \frac{45 + 28}{258 + 216} = 0.154$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.154)(0.846)}{258} + \frac{(0.154)(0.846)}{216}}$$

$$= 0.0333$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0333)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.13 - 0.174) - 0}{0.0333} \\ &= -1.32 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-1.32) \\ &= 0.1868 \end{aligned}$$

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) $H_0: \quad p_2 - p_1 = 0$

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

(c) $z^\star = 1.28$

(d) $SE = 0.0333$

(e) $|z_{\text{obs}}| = 1.32$

(f) p-value = 0.1868

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 010

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 = 12.5**

   (b) **UB = 13.1**

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

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

   (c) $t^\star = 3.36$

   (d) $SE = 0.997$

   (e) $\mid t_{\textbf{obs}} \mid = 3.336$

   (f) $0.02 < p$**-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) $t^\star = 2.02$

   (d) $SE = 8.22$

   (e) $\mid t_{\textbf{obs}} \mid = 1.85$

   (f) $0.1 < p$**-value** < 0.2

   (g) **retain**

4. (a) **LB of p CI = 0.988 or** 98.8%

   (b) **UB of p CI = 0.99 or** 99%

5. $n \approx 110000$

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

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

(c) $z^{\star} = 2.81$

(d) $SE = 0.041$

(e) $\mid z_{\mathbf{obs}} \mid = 3$

(f) p-**value** = 0.0026

(g) **reject**

1. **Problem**
   As an ornithologist, you wish to determine the average body mass of *Dendroica pensylvanica*. You randomly capture 28 adults of *Dendroica pensylvanica*, resulting in a sample mean of 12.77 grams and a sample standard deviation of 1.26 grams. You decide to report a 80% 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.

$$\begin{aligned} n &= 28 \\ \bar{x} &= 12.77 \\ s &= 1.26 \\ CL &= 0.8 \end{aligned}$$

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

$$df = n - 1 = 27$$

   Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.8$.

$$t^\star = 1.31$$

   Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.26}{\sqrt{28}} = 0.238$$

   We want to make an inference about the population mean.

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

   Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (12.77 - 1.31 \times 0.238,\ 12.77 - 1.31 \times 0.238) \\ &= (12.5,\ 13.1) \end{aligned}$$

   We are 80% confident that the population mean is between 12.5 and 13.1.

   (a) Lower bound = 12.5

   (b) Upper bound = 13.1

2. **Problem**
A teacher has 6 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
quiz 1:	54.8	54.9	73.8	71.4	80.4	55
quiz 2:	49.4	52	68.5	64.8	74.7	56.5

(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.

$$n = 6$$
$$\alpha = 0.02$$

State the hypotheses.

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

Determine the degrees of freedom.

$$df = n - 1 = 5$$

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

$$t^\star = 3.36$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6
quiz2-quiz1:	-5.4	-2.9	-5.3	-6.6	-5.7	1.5

Find the sample mean.

$$\overline{x_{\text{diff}}} = -4.07$$

Find the sample standard deviation.

$$s_{\text{diff}} = 2.99$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 1.22$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{-4.07 - 0}{1.22} = -3.336$$

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) $t^{\star} = 3.36$

(d) $SE = 0.9967229$

(e) $|t_{\text{obs}}| = 3.336$

(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 0.1. Your data is below. Please use 5 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
85	83
94	79
80	82
126	82
91	74

(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.

$$n_1 = 5$$
$$n_2 = 5$$
$$\alpha = 0.1$$

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.

$$t^\star = 2.02$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 95.2$$
$$s_1 = 18$$
$$\bar{x}_2 = 80$$
$$s_2 = 3.67$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(18)^2}{5} + \frac{(3.67)^2}{5}} \\ &= 8.22 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(80 - 95.2) - (0)}{8.22} \\ &= -1.85 \end{aligned}$$

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.

$$0.1 < p\text{-value} < 0.2$$

Compare p-value and α.

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

We conclude that we should retain the null hypothesis.

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

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

(c) $t^\star = 2.02$

(d) $SE = 8.22$

(e) $|t_{\text{obs}}| = 1.85$

(f) $0.1 < p\text{-value} < 0.2$

(g) retain the null

4. **Problem**
From a very large population, a random sample of 49000 individuals was taken. In that sample, 98.9% were bitter. Determine a 80% 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 $P(|Z| < z^\star) = 0.8$.

$$z^\star = 1.28$$

Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.989)(1-0.989)}{49000}} = 0.000471$$

Calculate the margin of error.

$$ME = z^\star SE = (1.28)(0.000471) = 0.000603$$

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.

$$(0.988\,,\ 0.99)$$

We are 80% confident that the true population proportion is between 98.8% and 99%.

   (a) The lower bound = 0.988, which can also be expressed as 98.8%.
   (b) The upper bound = 0.99, which can also be expressed as 99%.

5. **Problem**
   Your boss wants to know what proportion of a very large population is sweet. She also wants to guarantee that the margin of error of a 80% confidence interval will be less than 0.002 (which is 0.2 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

   Solution
   Determine $z^\star$ such that $P(|Z| < z^\star) = 0.8$.

$$z^\star = 1.28$$

   Determine the maximal standard error.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.002}{1.28} = 0.00156$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.00156)^2} = 102728.4681131$$

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

$$n = 102729$$

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

$$n = 110000$$

6. **Problem**
An experiment is run with a treatment group of size 254 and a control group of size 259. The results are summarized in the table below.

	treatment	control
sick	155	190
not sick	99	69

Using a significance level of 0.005, 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 $P(|Z| > z^\star) = 0.005$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 2.81$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{155}{254} = 0.61$$

$$\hat{p}_2 = \frac{190}{259} = 0.734$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.734 - 0.61 = 0.124$$

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

$$\hat{p} = \frac{155 + 190}{254 + 259} = 0.673$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.673)(0.327)}{254} + \frac{(0.673)(0.327)}{259}}$$

$$= 0.0414$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0414)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.734 - 0.61) - 0}{0.0414} \\ &= 3 \end{aligned}$$

Determine the *p*-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-3) \\ &= 0.0026 \end{aligned}$$

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) $z^\star = 2.81$

(d) $SE = 0.0414$

(e) $|z_{\text{obs}}| = 3$

(f) p-value = 0.0026

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 — Exam ID 011

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 = 18.2**

   (b) **UB = 21.1**

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

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

   (c) $t^\star = 2.76$

   (d) $SE = 1.02$

   (e) $\mid t_{\textbf{obs}} \mid = 2.636$

   (f) $0.04 < p$**-value** < 0.05

   (g) **retain**

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

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

   (c) $t^\star = 1.86$

   (d) $SE = 0.275$

   (e) $\mid t_{\textbf{obs}} \mid = 2.11$

   (f) $0.05 < p$**-value** < 0.1

   (g) **reject**

4. (a) **LB of p CI = 0.429 or** 42.9%

   (b) **UB of p CI = 0.511 or** 51.1%

5. $n \approx 31000$

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

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

(c) $z^{\star} = 2.81$

(d) $SE = 0.083$

(e) $\mid z_{\mathbf{obs}} \mid = 2.91$

(f) p**-value** $= 0.0036$

(g) **reject**

1. **Problem**
   As an ornithologist, you wish to determine the average body mass of *Seiurus noveboracensis*. You randomly capture 31 adults of *Seiurus noveboracensis*, resulting in a sample mean of 19.63 grams and a sample standard deviation of 3.26 grams. You decide to report a 98% 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.

$$\begin{aligned} n &= 31 \\ \bar{x} &= 19.63 \\ s &= 3.26 \\ CL &= 0.98 \end{aligned}$$

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

$$df = n - 1 = 30$$

   Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.98$.

$$t^\star = 2.46$$

   Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{3.26}{\sqrt{31}} = 0.586$$

   We want to make an inference about the population mean.

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

   Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (19.63 - 2.46 \times 0.586,\ 19.63 - 2.46 \times 0.586) \\ &= (18.2,\ 21.1) \end{aligned}$$

   We are 98% confident that the population mean is between 18.2 and 21.1.

   (a) Lower bound = 18.2
   (b) Upper bound = 21.1

2. **Problem**
A teacher has 6 students who have each taken two quizzes. Perform a two-tail test with significance level 0.04 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6
quiz 1:	61.6	62.3	84.1	73.6	69.8	83.4
quiz 2:	67.9	60.5	89.3	78.7	74.6	84.2

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

$$n = 6$$
$$\alpha = 0.04$$

State the hypotheses.

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

Determine the degrees of freedom.

$$df = n - 1 = 5$$

We determine $t^{\star}$ such that $P(|T| > t^{\star}) = 0.04$.

$$t^{\star} = 2.76$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6
quiz2-quiz1:	6.3	-1.8	5.2	5.1	4.8	0.8

Find the sample mean.

$$\overline{x_{\text{diff}}} = 3.4$$

Find the sample standard deviation.

$$s_{\text{diff}} = 3.17$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 1.29$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{3.4 - 0}{1.29} = 2.636$$

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.

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

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) $t^{\star} = 2.76$

(d) $SE = 1.0197905$

(e) $|t_{\text{obs}}| = 2.636$

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

(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 6 cases and a control group with 6 cases. You decide to run a hypothesis test with a significance level of 0.1. Your data is below. Please use 8 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
2.93	4.41
3.5	3.44
2.57	3.12
2.6	2.72
3.03	3.7
3.03	3.75

(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.

$$n_1 = 6$$
$$n_2 = 6$$
$$\alpha = 0.1$$

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.

$$t^\star = 1.86$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 2.94$$
$$s_1 = 0.341$$
$$\bar{x}_2 = 3.52$$
$$s_2 = 0.58$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(0.341)^2}{6} + \frac{(0.58)^2}{6}} \\ &= 0.275 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(3.52 - 2.94) - (0)}{0.275} \\ &= 2.11 \end{aligned}$$

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.

$$0.05 < p\text{-value} < 0.1$$

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) $t^{\star} = 1.86$

(d) $SE = 0.275$

(e) $|t_{\text{obs}}| = 2.11$

(f) $0.05 < p\text{-value} < 0.1$

(g) reject the null

4. **Problem**
   From a very large population, a random sample of 570 individuals was taken. In that sample, 47% were special. 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.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.47)(1-0.47)}{570}} = 0.0209$$

Calculate the margin of error.

$$ME = z^\star SE = (1.96)(0.0209) = 0.041$$

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.

$$(0.429\,,\ 0.511)$$

We are 95% confident that the true population proportion is between 42.9% and 51.1%.

(a) The lower bound = 0.429, which can also be expressed as 42.9%.
(b) The upper bound = 0.511, which can also be expressed as 51.1%.

5. **Problem**

Your boss wants to know what proportion of a very large population is bitter. She also wants to guarantee that the margin of error of a 99.5% confidence interval will be less than 0.008 (which is 0.8 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

Solution

Determine $z^\star$ such that $P(|Z| < z^\star) = 0.995$.

$$z^\star = 2.81$$

Determine the maximal standard error.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.008}{2.81} = 0.00285$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.00285)^2} = 30778.7011388$$

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

$$n = 30779$$

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

$$n = 31000$$

6. **Problem**
An experiment is run with a treatment group of size 72 and a control group of size 70. The results are summarized in the table below.

	treatment	control
green	40	22
not green	32	48

Using a significance level of 0.005, determine whether the treatment causes an effect on the proportion of cases that are green.

(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 $P(|Z| > z^\star) = 0.005$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 2.81$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{40}{72} = 0.556$$

$$\hat{p}_2 = \frac{22}{70} = 0.314$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.314 - 0.556 = -0.242$$

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

$$\hat{p} = \frac{40 + 22}{72 + 70} = 0.437$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.437)(0.563)}{72} + \frac{(0.437)(0.563)}{70}}$$

$$= 0.0833$$

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

$$H_0 : \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0833)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.314 - 0.556) - 0}{0.0833} \\ &= -2.91 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.91) \\ &= 0.0036 \end{aligned}$$

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) $H_0 : \quad p_2 - p_1 = 0$

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

(c) $z^{\star} = 2.81$

(d) $SE = 0.0833$

(e) $|z_{\text{obs}}| = 2.91$

(f) p-value $= 0.0036$

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 012

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 = 12.8**

   (b) **UB = 13.9**

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

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

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

   (d) $SE = 1.02$

   (e) $\mid t_{\textbf{obs}} \mid = 2.645$

   (f) $0.04 < p$**-value** < 0.05

   (g) **reject**

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

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

   (c) $t^{\star} = 2.82$

   (d) $SE = 0.998$

   (e) $\mid t_{\textbf{obs}} \mid = 2.91$

   (f) $0.01 < p$**-value** < 0.02

   (g) **reject**

4. (a) **LB of p CI = 0.927 or** 92.7%

   (b) **UB of p CI = 0.941 or** 94.1%

5. $n \approx 1900$

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

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

(c) $z^{\star} = 2.81$

(d) $SE = 0.039$

(e) $|\ z_{\mathbf{obs}}\ | = 2.98$

(f) p-**value** = 0.0028

(g) **reject**

1. **Problem**
   As an ornithologist, you wish to determine the average body mass of *Vireo philadelphicus*. You randomly capture 35 adults of *Vireo philadelphicus*, resulting in a sample mean of 13.36 grams and a sample standard deviation of 1.92 grams. You decide to report a 90% 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.

$$\begin{aligned} n &= 35 \\ \bar{x} &= 13.36 \\ s &= 1.92 \\ CL &= 0.9 \end{aligned}$$

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

$$df = n - 1 = 34$$

   Determine the critical t value, $t^{\star}$, such that $P(|T| < t^{\star}) = 0.9$.

$$t^{\star} = 1.69$$

   Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.92}{\sqrt{35}} = 0.325$$

   We want to make an inference about the population mean.

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

   Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^{\star}SE,\ \bar{x} + t^{\star}SE) \\ &= (13.36 - 1.69 \times 0.325,\ 13.36 - 1.69 \times 0.325) \\ &= (12.8,\ 13.9) \end{aligned}$$

   We are 90% confident that the population mean is between 12.8 and 13.9.

   (a) Lower bound = 12.8

   (b) Upper bound = 13.9

2. **Problem**
A teacher has 6 students who have each taken two quizzes. Perform a two-tail test with significance level 0.05 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6
quiz 1:	84.6	82.4	89.8	63.3	64.8	50.4
quiz 2:	90.1	81.8	89.6	67.9	69.4	53.5

(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.

$$n = 6$$
$$\alpha = 0.05$$

State the hypotheses.

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

Determine the degrees of freedom.

$$df = n - 1 = 5$$

We determine $t^\star$ such that $P(|T| > t^\star) = 0.05$.

$$t^\star = 2.57$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6
quiz2-quiz1:	5.5	-0.6	-0.2	4.6	4.6	3.1

Find the sample mean.

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

Find the sample standard deviation.

$$s_{\text{diff}} = 2.62$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 1.07$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{2.83 - 0}{1.07} = 2.645$$

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.

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

We conclude that we should reject the null hypothesis.

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

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

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

(d) $SE = 1.0185456$

(e) $|t_{\text{obs}}| = 2.645$

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

(g) reject 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 6 cases and a control group with 6 cases. You decide to run a hypothesis test with a significance level of 0.02. Your data is below. Please use 9 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
12.6	11.6
15.6	11.5
14.8	11
16.3	12.5
12.4	11.2
11.6	8.1

(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.

$$n_1 = 6$$
$$n_2 = 6$$
$$\alpha = 0.02$$

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.

$$t^\star = 2.82$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 13.9$$
$$s_1 = 1.93$$
$$\bar{x}_2 = 11$$
$$s_2 = 1.5$$

Calculate the standard error.

$$SE = \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$$
$$= \sqrt{\frac{(1.93)^2}{6} + \frac{(1.5)^2}{6}}$$
$$= 0.998$$

Determine the test statistic.

$$t_{\text{obs}} = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE}$$
$$= \frac{(11 - 13.9) - (0)}{0.998}$$
$$= -2.91$$

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.

$$0.01 < p\text{-value} < 0.02$$

Compare p-value and α.

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

We conclude that we should reject the null hypothesis.

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

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

(c) $t^\star = 2.82$

(d) $SE = 0.998$

(e) $|t_{\text{obs}}| = 2.91$

(f) $0.01 < p$-value < 0.02

(g) reject the null

4. **Problem**
   From a very large population, a random sample of 8500 individuals was taken. In that sample, 93.4% were messy. Determine a 99% 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 $P(|Z| < z^\star) = 0.99$.

$$z^\star = 2.58$$

   Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.934)(1-0.934)}{8500}} = 0.00269$$

   Calculate the margin of error.

$$ME = z^\star SE = (2.58)(0.00269) = 0.00694$$

   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.

$$(0.927\,,\;0.941)$$

   We are 99% confident that the true population proportion is between 92.7% and 94.1%.

   (a) The lower bound = 0.927, which can also be expressed as 92.7%.
   (b) The upper bound = 0.941, which can also be expressed as 94.1%.

5. **Problem**
   Your boss wants to know what proportion of a very large population is happy. She also wants to guarantee that the margin of error of a 99% confidence interval will be less than 0.03 (which is 3 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

   Solution
   Determine $z^\star$ such that $P(|Z| < z^\star) = 0.99$.

   $$z^\star = 2.58$$

   Determine the maximal standard error.

   $$ME = z^\star SE$$

   $$SE = \frac{ME}{z^\star} = \frac{0.03}{2.58} = 0.0116$$

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

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

   $$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.0116)^2} = 1857.9072533$$

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

   $$n = 1858$$

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

   $$n = 1900$$

6. **Problem**
An experiment is run with a treatment group of size 138 and a control group of size 186. The results are summarized in the table below.

	treatment	control
sick	128	151
not sick	10	35

Using a significance level of 0.005, 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 $P(|Z| > z^\star) = 0.005$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 2.81$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{128}{138} = 0.928$$

$$\hat{p}_2 = \frac{151}{186} = 0.812$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.812 - 0.928 = -0.116$$

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

$$\hat{p} = \frac{128 + 151}{138 + 186} = 0.861$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.861)(0.139)}{138} + \frac{(0.861)(0.139)}{186}}$$

$$= 0.0389$$

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

$$H_0 : \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0389)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.812 - 0.928) - 0}{0.0389} \\ &= -2.98 \end{aligned}$$

Determine the *p*-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.98) \\ &= 0.0028 \end{aligned}$$

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) $H_0 : \quad p_2 - p_1 = 0$

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

(c) $z^{\star} = 2.81$

(d) $SE = 0.0389$

(e) $|z_{\text{obs}}| = 2.98$

(f) p-value = 0.0028

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 013

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 = 53.7**

   (b) **UB = 57.4**

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

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

   (c) $t^\star = 2.02$

   (d) $SE = 1.14$

   (e) $\mid t_{\textbf{obs}} \mid = 2.063$

   (f) $0.05 < p$**-value** < 0.1

   (g) **reject**

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

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

   (c) $t^\star = 2.36$

   (d) $SE = 30.8$

   (e) $\mid t_{\textbf{obs}} \mid = 2.08$

   (f) $0.05 < p$**-value** < 0.1

   (g) **retain**

4. (a) **LB of p CI = 0.551 or** 55.1%

   (b) **UB of p CI = 0.617 or** 61.7%

5. $n \approx 390$

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) $SE = 0.036$

(e) $\mid z_{\mathbf{obs}} \mid = 2.58$

(f) p-**value** = 0.0098

(g) **reject**

1. **Problem**

   As an ornithologist, you wish to determine the average body mass of *Hylocichla mustelina*. You randomly capture 27 adults of *Hylocichla mustelina*, resulting in a sample mean of 55.53 grams and a sample standard deviation of 7.22 grams. You decide to report a 80% 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.

$$\begin{aligned} n &= 27 \\ \bar{x} &= 55.53 \\ s &= 7.22 \\ CL &= 0.8 \end{aligned}$$

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

$$df = n - 1 = 26$$

   Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.8$.

$$t^\star = 1.31$$

   Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{7.22}{\sqrt{27}} = 1.39$$

   We want to make an inference about the population mean.

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

   Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (55.53 - 1.31 \times 1.39,\ 55.53 - 1.31 \times 1.39) \\ &= (53.7,\ 57.4) \end{aligned}$$

   We are 80% confident that the population mean is between 53.7 and 57.4.

   (a) Lower bound = 53.7

   (b) Upper bound = 57.4

2. **Problem**
A teacher has 6 students who have each taken two quizzes. Perform a two-tail test with significance level 0.1 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6
quiz 1:	57.3	66.2	58.3	55.3	59.6	71.5
quiz 2:	54.1	64.6	53.1	55.6	60.1	69.8

(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.

$$n = 6$$
$$\alpha = 0.1$$

State the hypotheses.

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

Determine the degrees of freedom.

$$df = n - 1 = 5$$

We determine $t^\star$ such that $P(|T| > t^\star) = 0.1$.

$$t^\star = 2.02$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6
quiz2-quiz1:	-3.2	-1.6	-5.2	0.3	0.5	-1.7

Find the sample mean.

$$\overline{x_{\text{diff}}} = -1.82$$

Find the sample standard deviation.

$$s_{\text{diff}} = 2.16$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 0.882$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{-1.82 - 0}{0.882} = -2.063$$

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.

$$0.05 < p\text{-value} < 0.1$$

We conclude that we should reject the null hypothesis.

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

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

(c) $t^{\star} = 2.02$

(d) $SE = 1.1443893$

(e) $|t_{\text{obs}}| = 2.063$

(f) $0.05 < p\text{-value} < 0.1$

(g) reject 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 7 cases and a control group with 7 cases. You decide to run a hypothesis test with a significance level of 0.05. Your data is below. Please use 7 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
270	410
360	420
270	450
500	420
370	420
350	390
320	380

(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.

$$n_1 = 7$$
$$n_2 = 7$$
$$\alpha = 0.05$$

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.

$$t^\star = 2.36$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 349$$
$$s_1 = 78.2$$
$$\bar{x}_2 = 413$$
$$s_2 = 22.9$$

Calculate the standard error.

$$SE = \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$$
$$= \sqrt{\frac{(78.2)^2}{7} + \frac{(22.9)^2}{7}}$$
$$= 30.8$$

Determine the test statistic.

$$t_{\text{obs}} = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE}$$
$$= \frac{(413 - 349) - (0)}{30.8}$$
$$= 2.08$$

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.

$$0.05 < p\text{-value} < 0.1$$

Compare *p*-value and α.

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

We conclude that we should retain the null hypothesis.

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

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

(c) $t^{\star} = 2.36$

(d) $SE = 30.8$

(e) $|t_{\text{obs}}| = 2.08$

(f) $0.05 < p\text{-value} < 0.1$

(g) retain the null

4. **Problem**
   From a very large population, a random sample of 910 individuals was taken. In that sample, 58.4% were bitter. Determine a 96% 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 $P(|Z| < z^{\star}) = 0.96$.

$$z^{\star} = 2.05$$

   Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.584)(1-0.584)}{910}} = 0.0163$$

   Calculate the margin of error.

$$ME = z^{\star} SE = (2.05)(0.0163) = 0.0334$$

   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.

$$(0.551\ ,\ 0.617)$$

   We are 96% confident that the true population proportion is between 55.1% and 61.7%.

   (a) The lower bound = 0.551, which can also be expressed as 55.1%.
   (b) The upper bound = 0.617, which can also be expressed as 61.7%.

5. **Problem**

   Your boss wants to know what proportion of a very large population is floating. She also wants to guarantee that the margin of error of a 95% confidence interval will be less than 0.05 (which is 5 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

   Solution

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

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

   Determine the maximal standard error.

$$ME = z^{\star} SE$$

$$SE = \frac{ME}{z^{\star}} = \frac{0.05}{1.96} = 0.0255$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.0255)^2} = 384.4675125$$

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

$$n = 385$$

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

$$n = 390$$

6. **Problem**
An experiment is run with a treatment group of size 243 and a control group of size 195. The results are summarized in the table below.

	treatment	control
angry	31	43
not angry	212	152

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

(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.

$$\hat{p}_1 = \frac{31}{243} = 0.128$$

$$\hat{p}_2 = \frac{43}{195} = 0.221$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.221 - 0.128 = 0.093$$

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

$$\hat{p} = \frac{31 + 43}{243 + 195} = 0.169$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.169)(0.831)}{243} + \frac{(0.169)(0.831)}{195}}$$

$$= 0.036$$

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

$$H_0 : \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\, 0.036)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.221 - 0.128) - 0}{0.036} \\ &= 2.58 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.58) \\ &= 0.0098 \end{aligned}$$

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) $H_0 : \quad p_2 - p_1 = 0$

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

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

(d) $SE = 0.036$

(e) $|z_{\text{obs}}| = 2.58$

(f) p-value = 0.0098

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 — Exam ID 014

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 = 14.8**

   (b) **UB = 17**

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

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

   (c) $t^{\star} = 2.78$

   (d) $SE = 1.12$

   (e) $\mid t_{\textbf{obs}} \mid = 2.845$

   (f) $0.04 <$ p**-value** < 0.05

   (g) **reject**

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

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

   (c) $t^{\star} = 2.52$

   (d) $SE = 24.3$

   (e) $\mid t_{\textbf{obs}} \mid = 2.47$

   (f) $0.04 <$ p**-value** < 0.05

   (g) **retain**

4. (a) **LB of p CI = 0.736 or** 73.6%

   (b) **UB of p CI = 0.784 or** 78.4%

5. $n \approx 17000$

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

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

(c) $z^{\star} = 2.05$

(d) $SE = 0.027$

(e) $\mid z_{\mathbf{obs}} \mid = 2.25$

(f) p-**value** $= 0.0244$

(g) **reject**

1. **Problem**

   As an ornithologist, you wish to determine the average body mass of *Oporornis formosus*. You randomly capture 22 adults of *Oporornis formosus*, resulting in a sample mean of 15.92 grams and a sample standard deviation of 3.01 grams. You decide to report a 90% 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.

$$\begin{aligned} n &= 22 \\ \bar{x} &= 15.92 \\ s &= 3.01 \\ CL &= 0.9 \end{aligned}$$

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

$$df = n - 1 = 21$$

   Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.9$.

$$t^\star = 1.72$$

   Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{3.01}{\sqrt{22}} = 0.642$$

   We want to make an inference about the population mean.

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

   Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (15.92 - 1.72 \times 0.642,\ 15.92 - 1.72 \times 0.642) \\ &= (14.8,\ 17) \end{aligned}$$

   We are 90% confident that the population mean is between 14.8 and 17.

   (a) Lower bound = 14.8

   (b) Upper bound = 17

2. **Problem**
A teacher has 5 students who have each taken two quizzes. Perform a two-tail test with significance level 0.05 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5
quiz 1:	66.3	53.5	64.2	70.3	60.3
quiz 2:	68.6	61	68.6	69.9	66.7

(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.

$$n = 5$$
$$\alpha = 0.05$$

State the hypotheses.

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

Determine the degrees of freedom.

$$df = n - 1 = 4$$

We determine $t^\star$ such that $P(|T| > t^\star) = 0.05$.

$$t^\star = 2.78$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5
quiz2-quiz1:	2.3	7.5	4.4	-0.4	6.4

Find the sample mean.

$$\overline{x_{\text{diff}}} = 4.04$$

Find the sample standard deviation.

$$s_{\text{diff}} = 3.18$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 1.42$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{4.04 - 0}{1.42} = 2.845$$

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.

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

We conclude that we should reject the null hypothesis.

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

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

(c) $t^{\star} = 2.78$

(d) $SE = 1.1229342$

(e) $|t_{\text{obs}}| = 2.845$

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

(g) reject 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 0.04. Your data is below. Please use 7 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
600	540
550	510
580	520
680	550
570	560

(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.

$$n_1 = 5$$
$$n_2 = 5$$
$$\alpha = 0.04$$

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.

$$t^{\star} = 2.52$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 596$$
$$s_1 = 50.3$$
$$\bar{x}_2 = 536$$
$$s_2 = 20.7$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(50.3)^2}{5} + \frac{(20.7)^2}{5}} \\ &= 24.3 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(536 - 596) - (0)}{24.3} \\ &= -2.47 \end{aligned}$$

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 retain the null hypothesis.

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

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

(c) $t^\star = 2.52$

(d) $SE = 24.3$

(e) $|t_{\text{obs}}| = 2.47$

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

(g) retain the null

4. **Problem**
   From a very large population, a random sample of 850 individuals was taken. In that sample, 76% were tasty. Determine a 90% 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 $P(|Z| < z^{\star}) = 0.9$.

   $$z^{\star} = 1.64$$

   Calculate the standard error.

   $$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.76)(1-0.76)}{850}} = 0.0146$$

   Calculate the margin of error.

   $$ME = z^{\star}SE = (1.64)(0.0146) = 0.0239$$

   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.

   $$(0.736\,,\ 0.784)$$

   We are 90% confident that the true population proportion is between 73.6% and 78.4%.

   (a) The lower bound = 0.736, which can also be expressed as 73.6%.

   (b) The upper bound = 0.784, which can also be expressed as 78.4%.

5. **Problem**
   Your boss wants to know what proportion of a very large population is frigid. She also wants to guarantee that the margin of error of a 96% confidence interval will be less than 0.008 (which is 0.8 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.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.008}{2.05} = 0.0039$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.0039)^2} = 16436.5548981$$

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

$$n = 16437$$

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

$$n = 17000$$

6. **Problem**
An experiment is run with a treatment group of size 233 and a control group of size 213. The results are summarized in the table below.

	treatment	control
omnivorous	206	201
not omnivorous	27	12

Using a significance level of 0.04, determine whether the treatment causes an effect on the proportion of cases that are omnivorous.

(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 $P(|Z| > z^\star) = 0.04$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 2.05$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{206}{233} = 0.884$$

$$\hat{p}_2 = \frac{201}{213} = 0.944$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.944 - 0.884 = 0.06$$

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

$$\hat{p} = \frac{206 + 201}{233 + 213} = 0.913$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.913)(0.087)}{233} + \frac{(0.913)(0.087)}{213}}$$

$$= 0.0267$$

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

$$H_0 : \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0267)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.944 - 0.884) - 0}{0.0267} \\ &= 2.25 \end{aligned}$$

Determine the *p*-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.25) \\ &= 0.0244 \end{aligned}$$

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) $H_0 : \quad p_2 - p_1 = 0$

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

(c) $z^\star = 2.05$

(d) $SE = 0.0267$

(e) $|z_{\text{obs}}| = 2.25$

(f) p-value = 0.0244

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 015

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 = 14.6**

   (b) **UB = 15.4**

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

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

   (c) $t^\star = 2.02$

   (d) $SE = 1$

   (e) $\mid t_{\textbf{obs}} \mid = 1.957$

   (f) $0.1 <$ ***p*-value** < 1

   (g) **retain**

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

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

   (c) $t^\star = 1.83$

   (d) $SE = 0.656$

   (e) $\mid t_{\textbf{obs}} \mid = 1.81$

   (f) $0.1 <$ ***p*-value** < 0.2

   (g) **retain**

4. (a) **LB of p CI = 0.806 or** 80.6%

   (b) **UB of p CI = 0.81 or** 81%

5. $n \approx 1300$

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

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

(c) $z^{\star} = 1.64$

(d) $SE = 0.065$

(e) $\mid z_{\mathbf{obs}} \mid = 2.01$

(f) p-**value** = 0.0444

(g) **reject**

1. **Problem**
   As an ornithologist, you wish to determine the average body mass of *Dendroica castanea*. You randomly capture 35 adults of *Dendroica castanea*, resulting in a sample mean of 15.03 grams and a sample standard deviation of 1.41 grams. You decide to report a 90% 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.

$$\begin{aligned} n &= 35 \\ \bar{x} &= 15.03 \\ s &= 1.41 \\ CL &= 0.9 \end{aligned}$$

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

$$df = n - 1 = 34$$

   Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.9$.

$$t^\star = 1.69$$

   Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.41}{\sqrt{35}} = 0.238$$

   We want to make an inference about the population mean.

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

   Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (15.03 - 1.69 \times 0.238,\ 15.03 - 1.69 \times 0.238) \\ &= (14.6,\ 15.4) \end{aligned}$$

   We are 90% confident that the population mean is between 14.6 and 15.4.

   (a) Lower bound = 14.6

   (b) Upper bound = 15.4

2. **Problem**
A teacher has 6 students who have each taken two quizzes. Perform a two-tail test with significance level 0.1 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6
quiz 1:	85	83.7	73.3	76.2	51.4	60.6
quiz 2:	85.5	84.1	71.2	70.4	49.4	58.6

(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.

$$n = 6$$
$$\alpha = 0.1$$

State the hypotheses.

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

Determine the degrees of freedom.

$$df = n - 1 = 5$$

We determine $t^{\star}$ such that $P(|T| > t^{\star}) = 0.1$.

$$t^{\star} = 2.02$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6
quiz2-quiz1:	0.5	0.4	-2.1	-5.8	-2	-2

Find the sample mean.

$$\overline{x_{\text{diff}}} = -1.83$$

Find the sample standard deviation.

$$s_{\text{diff}} = 2.29$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 0.935$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{-1.83 - 0}{0.935} = -1.957$$

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.

$$0.1 < p\text{-value} < 1$$

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) $t^{\star} = 2.02$

(d) $SE = 1.0047092$

(e) $|t_{\text{obs}}| = 1.957$

(f) $0.1 < p\text{-value} < 1$

(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 7 cases and a control group with 7 cases. You decide to run a hypothesis test with a significance level of 0.1. Your data is below. Please use 9 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
4.3	6
7.3	3
6.1	4.1
6.3	5.6
6.4	6.9
7.7	5.7
7.3	5.8

(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.

$$n_1 = 7$$
$$n_2 = 7$$
$$\alpha = 0.1$$

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.

$$t^\star = 1.83$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 6.49$$
$$s_1 = 1.14$$
$$\bar{x}_2 = 5.3$$
$$s_2 = 1.31$$

Calculate the standard error.

$$
\begin{aligned}
SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\
&= \sqrt{\frac{(1.14)^2}{7} + \frac{(1.31)^2}{7}} \\
&= 0.656
\end{aligned}
$$

Determine the test statistic.

$$
\begin{aligned}
t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\
&= \frac{(5.3 - 6.49) - (0)}{0.656} \\
&= -1.81
\end{aligned}
$$

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.

$$0.1 < p\text{-value} < 0.2$$

Compare *p*-value and α.

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

We conclude that we should retain the null hypothesis.

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

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

(c) $t^\star = 1.83$

(d) $SE = 0.656$

(e) $|t_{\text{obs}}| = 1.81$

(f) $0.1 < p\text{-value} < 0.2$

(g) retain the null

4. **Problem**

   From a very large population, a random sample of 97000 individuals was taken. In that sample, 80.8% were messy. 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.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.808)(1-0.808)}{97000}} = 0.00126$$

   Calculate the margin of error.

$$ME = z^{\star} SE = (1.96)(0.00126) = 0.00247$$

   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.

$$(0.806\,,\ 0.81)$$

   We are 95% confident that the true population proportion is between 80.6% and 81%.

   (a) The lower bound = 0.806, which can also be expressed as 80.6%.

   (b) The upper bound = 0.81, which can also be expressed as 81%.

5. **Problem**

   Your boss wants to know what proportion of a very large population is cold. She also wants to guarantee that the margin of error of a 99.5% 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 $P(|Z| < z^\star) = 0.995$.

   $$z^\star = 2.81$$

   Determine the maximal standard error.

   $$ME = z^\star SE$$

   $$SE = \frac{ME}{z^\star} = \frac{0.04}{2.81} = 0.0142$$

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

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

   $$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.0142)^2} = 1239.8333664$$

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

   $$n = 1240$$

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

   $$n = 1300$$

6. **Problem**
An experiment is run with a treatment group of size 111 and a control group of size 106. The results are summarized in the table below.

	treatment	control
special	64	75
not special	47	31

Using a significance level of 0.1, determine whether the treatment causes an effect on the proportion of cases that are special.

(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 $P(|Z| > z^\star) = 0.1$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 1.64$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{64}{111} = 0.577$$

$$\hat{p}_2 = \frac{75}{106} = 0.708$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.708 - 0.577 = 0.131$$

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

$$\hat{p} = \frac{64 + 75}{111 + 106} = 0.641$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.641)(0.359)}{111} + \frac{(0.641)(0.359)}{106}}$$

$$= 0.0651$$

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

$$H_0 : \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0651)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.708 - 0.577) - 0}{0.0651} \\ &= 2.01 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.01) \\ &= 0.0444 \end{aligned}$$

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) $H_0 : \quad p_2 - p_1 = 0$

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

(c) $z^\star = 1.64$

(d) $SE = 0.0651$

(e) $|z_{\text{obs}}| = 2.01$

(f) p-value $= 0.0444$

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 016

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 = 13.1**

   (b) **UB = 13.8**

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

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

   (c) $t^{\star} = 1.94$

   (d) $SE = 1.42$

   (e) $\mid t_{\textbf{obs}} \mid = 1.79$

   (f) $0.1 <$ p**-value** < 1

   (g) **retain**

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

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

   (c) $t^{\star} = 3.5$

   (d) $SE = 0.271$

   (e) $\mid t_{\textbf{obs}} \mid = 3.69$

   (f) $0.005 <$ p**-value** < 0.01

   (g) **reject**

4. (a) **LB of p CI = 0.339 or** 33.9%

   (b) **UB of p CI = 0.437 or** 43.7%

5. $n \approx 160000$

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

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

(c) $z^{\star} = 1.64$

(d) $SE = 0.053$

(e) $| \; z_{\mathbf{obs}} \; | = 1.88$

(f) p-**value** $= 0.0602$

(g) **reject**

1. **Problem**

As an ornithologist, you wish to determine the average body mass of *Vireo philadelphicus*. You randomly capture 35 adults of *Vireo philadelphicus*, resulting in a sample mean of 13.45 grams and a sample standard deviation of 1.63 grams. You decide to report a 80% 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.

$$\begin{aligned} n &= 35 \\ \bar{x} &= 13.45 \\ s &= 1.63 \\ CL &= 0.8 \end{aligned}$$

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

$$df = n - 1 = 34$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.8$.

$$t^\star = 1.31$$

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.63}{\sqrt{35}} = 0.276$$

We want to make an inference about the population mean.

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

Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (13.45 - 1.31 \times 0.276,\ 13.45 - 1.31 \times 0.276) \\ &= (13.1,\ 13.8) \end{aligned}$$

We are 80% confident that the population mean is between 13.1 and 13.8.

(a) Lower bound = 13.1

(b) Upper bound = 13.8

2. **Problem**
A teacher has 7 students who have each taken two quizzes. Perform a two-tail test with significance level 0.1 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6	student7
quiz 1:	63.7	89.5	86.2	59.4	55.3	57.8	52.2
quiz 2:	58.2	89.3	81.7	57.7	55.5	49.8	54.6

(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.

$$n = 7$$
$$\alpha = 0.1$$

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 $P(|T| > t^\star) = 0.1$.

$$t^\star = 1.94$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6	student7
quiz2-quiz1:	-5.5	-0.2	-4.5	-1.7	0.2	-8	2.4

Find the sample mean.

$$\overline{x_{\text{diff}}} = -2.47$$

Find the sample standard deviation.

$$s_{\text{diff}} = 3.66$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 1.38$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{-2.47 - 0}{1.38} = -1.79$$

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.

$$0.1 < p\text{-value} < 1$$

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) $t^{\star} = 1.94$

(d) $SE = 1.4152078$

(e) $|t_{\text{obs}}| = 1.79$

(f) $0.1 < p\text{-value} < 1$

(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 6 cases and a control group with 6 cases. You decide to run a hypothesis test with a significance level of 0.01. Your data is below. Please use 7 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
3.8	2.8
4.3	3.9
4.3	2.6
3.8	2.6
3.9	2.7
4.3	3.8

(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.

$$n_1 = 6$$
$$n_2 = 6$$
$$\alpha = 0.01$$

State the hypotheses.

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

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

$$t^\star = 3.5$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 4.07$$
$$s_1 = 0.258$$
$$\bar{x}_2 = 3.07$$
$$s_2 = 0.612$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(0.258)^2}{6} + \frac{(0.612)^2}{6}} \\ &= 0.271 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(3.07 - 4.07) - (0)}{0.271} \\ &= -3.69 \end{aligned}$$

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.

$$0.005 < p\text{-value} < 0.01$$

Compare p-value and α.

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

We conclude that we should reject the null hypothesis.

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

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

(c) $t^{\star} = 3.5$

(d) $SE = 0.271$

(e) $|t_{\text{obs}}| = 3.69$

(f) $0.005 < p\text{-value} < 0.01$

(g) reject the null

4. **Problem**

   From a very large population, a random sample of 670 individuals was taken. In that sample, 38.8% were messy. Determine a 99% 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 $P(|Z| < z^\star) = 0.99$.

$$z^\star = 2.58$$

   Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.388)(1-0.388)}{670}} = 0.0188$$

   Calculate the margin of error.

$$ME = z^\star SE = (2.58)(0.0188) = 0.0485$$

   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.

$$(0.339\,,\ 0.437)$$

   We are 99% confident that the true population proportion is between 33.9% and 43.7%.

   (a) The lower bound = 0.339, which can also be expressed as 33.9%.

   (b) The upper bound = 0.437, which can also be expressed as 43.7%.

5. **Problem**

   Your boss wants to know what proportion of a very large population is green. She also wants to guarantee that the margin of error of a 98% confidence interval will be less than 0.003 (which is 0.3 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

   Solution

   Determine $z^\star$ such that $P(|Z| < z^\star) = 0.98$.

$$z^\star = 2.33$$

   Determine the maximal standard error.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.003}{2.33} = 0.00129$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.00129)^2} = 150231.3562887$$

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

$$n = 150232$$

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

$$n = 160000$$

6. **Problem**

An experiment is run with a treatment group of size 136 and a control group of size 151. The results are summarized in the table below.

	treatment	control
folksy	44	34
not folksy	92	117

Using a significance level of 0.1, determine whether the treatment causes an effect on the proportion of cases that are folksy.

(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 $P(|Z| > z^\star) = 0.1$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 1.64$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{44}{136} = 0.324$$

$$\hat{p}_2 = \frac{34}{151} = 0.225$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.225 - 0.324 = -0.099$$

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

$$\hat{p} = \frac{44 + 34}{136 + 151} = 0.272$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.272)(0.728)}{136} + \frac{(0.272)(0.728)}{151}}$$

$$= 0.0526$$

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

$$H_0 : \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0526)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.225 - 0.324) - 0}{0.0526} \\ &= -1.88 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-1.88) \\ &= 0.0602 \end{aligned}$$

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) $H_0 : \quad p_2 - p_1 = 0$

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

(c) $z^\star = 1.64$

(d) $SE = 0.0526$

(e) $|z_{\text{obs}}| = 1.88$

(f) p-value $= 0.0602$

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 — Exam ID 017

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 = 9.68**

   (b) **UB = 10.6**

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

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

   (c) $t^\star = 2.45$

   (d) $SE = 1.27$

   (e) $\mid t_{\textbf{obs}} \mid = 2.562$

   (f) $0.04 < p$**-value** < 0.05

   (g) **reject**

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

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

   (c) $t^\star = 3.36$

   (d) $SE = 0.645$

   (e) $\mid t_{\textbf{obs}} \mid = 3.29$

   (f) $0.01 < p$**-value** < 0.02

   (g) **retain**

4. (a) **LB of p CI = 0.208 or** 20.8%

   (b) **UB of p CI = 0.288 or** 28.8%

5. $n \approx 410000$

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

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

(c) $z^\star = 2.81$

(d) $SE = 0.03$

(e) $\mid z_{\mathbf{obs}} \mid = 2.95$

(f) p-**value** = 0.0032

(g) **reject**

1. **Problem**
As an ornithologist, you wish to determine the average body mass of *Vireo griseus*. You randomly capture 33 adults of *Vireo griseus*, resulting in a sample mean of 10.13 grams and a sample standard deviation of 0.848 grams. You decide to report a 99.5% 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.

$$\begin{aligned} n &= 33 \\ \bar{x} &= 10.13 \\ s &= 0.848 \\ CL &= 0.995 \end{aligned}$$

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

$$df = n - 1 = 32$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.995$.

$$t^\star = 3.01$$

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{0.848}{\sqrt{33}} = 0.148$$

We want to make an inference about the population mean.

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

Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (10.13 - 3.01 \times 0.148,\ 10.13 - 3.01 \times 0.148) \\ &= (9.68,\ 10.6) \end{aligned}$$

We are 99.5% confident that the population mean is between 9.68 and 10.6.

   (a) Lower bound = 9.68

   (b) Upper bound = 10.6

2. **Problem**
A teacher has 7 students who have each taken two quizzes. Perform a two-tail test with significance level 0.05 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6	student7
quiz 1:	55	57.2	68	87.8	78.3	75.9	58.2
quiz 2:	48.7	59.3	60.8	85.3	78.1	72.7	52.2

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

$$n = 7$$
$$\alpha = 0.05$$

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 $P(|T| > t^{\star}) = 0.05$.

$$t^{\star} = 2.45$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6	student7
quiz2-quiz1:	-6.3	2.1	-7.2	-2.5	-0.2	-3.2	-6

Find the sample mean.

$$\overline{x_{\text{diff}}} = -3.33$$

Find the sample standard deviation.

$$s_{\text{diff}} = 3.44$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 1.3$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{-3.33 - 0}{1.3} = -2.562$$

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.

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

We conclude that we should reject the null hypothesis.

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

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

(c) $t^{\star} = 2.45$

(d) $SE = 1.269566$

(e) $|t_{\text{obs}}| = 2.562$

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

(g) reject 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 6 cases and a control group with 6 cases. You decide to run a hypothesis test with a significance level of 0.01. Your data is below. Please use 8 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
5.4	7.4
3.3	6.4
6.8	8.5
6.8	8.3
4.5	7.4
6.2	7.7

(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.

$$n_1 = 6$$
$$n_2 = 6$$
$$\alpha = 0.01$$

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.

$$t^\star = 3.36$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 5.5$$
$$s_1 = 1.39$$
$$\bar{x}_2 = 7.62$$
$$s_2 = 0.752$$

Calculate the standard error.

$$SE = \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$$
$$= \sqrt{\frac{(1.39)^2}{6} + \frac{(0.752)^2}{6}}$$
$$= 0.645$$

Determine the test statistic.

$$t_{\text{obs}} = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE}$$
$$= \frac{(7.62 - 5.5) - (0)}{0.645}$$
$$= 3.29$$

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.

$$0.01 < p\text{-value} < 0.02$$

Compare p-value and α.

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

We conclude that we should retain the null hypothesis.

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

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

(c) $t^\star = 3.36$

(d) $SE = 0.645$

(e) $|t_{\text{obs}}| = 3.29$

(f) $0.01 < p\text{-value} < 0.02$

(g) retain the null

4. **Problem**

From a very large population, a random sample of 480 individuals was taken. In that sample, 24.8% were asleep. Determine a 96% 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 $P(|Z| < z^\star) = 0.96$.

$$z^\star = 2.05$$

Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.248)(1-0.248)}{480}} = 0.0197$$

Calculate the margin of error.

$$ME = z^\star SE = (2.05)(0.0197) = 0.0404$$

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.

$$(0.208\,,\ 0.288)$$

We are 96% confident that the true population proportion is between 20.8% and 28.8%.

(a) The lower bound = 0.208, which can also be expressed as 20.8%.

(b) The upper bound = 0.288, which can also be expressed as 28.8%.

5. **Problem**
   Your boss wants to know what proportion of a very large population is cold. She also wants to guarantee that the margin of error of a 80% confidence interval will be less than 0.001 (which is 0.1 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

   Solution
   Determine $z^{\star}$ such that $P(|Z| < z^{\star}) = 0.8$.

$$z^{\star} = 1.28$$

   Determine the maximal standard error.

$$ME = z^{\star} SE$$

$$SE = \frac{ME}{z^{\star}} = \frac{0.001}{1.28} = 0.000781$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.000781)^2} = 409862.2698828$$

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

$$n = 409863$$

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

$$n = 410000$$

6. **Problem**

An experiment is run with a treatment group of size 200 and a control group of size 150. The results are summarized in the table below.

	treatment	control
angry	9	20
not angry	191	130

Using a significance level of 0.005, determine whether the treatment causes an effect on the proportion of cases that are angry.

(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 $P(|Z| > z^\star) = 0.005$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 2.81$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{9}{200} = 0.045$$

$$\hat{p}_2 = \frac{20}{150} = 0.133$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.133 - 0.045 = 0.088$$

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

$$\hat{p} = \frac{9 + 20}{200 + 150} = 0.0829$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.0829)(0.9171)}{200} + \frac{(0.0829)(0.9171)}{150}}$$

$$= 0.0298$$

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

$$H_0 : \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0298)$$

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.

$$z = \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE}$$
$$= \frac{(0.133 - 0.045) - 0}{0.0298}$$
$$= 2.95$$

Determine the p-value.

$$p\text{-value} = 2 \cdot \Phi(-|z|)$$
$$= 2 \cdot \Phi(-2.95)$$
$$= 0.0032$$

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) $H_0 : \quad p_2 - p_1 = 0$

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

(c) $z^\star = 2.81$

(d) $SE = 0.0298$

(e) $|z_{\text{obs}}| = 2.95$

(f) p-value = 0.0032

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 018

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 = 14.3**

   (b) **UB = 17.9**

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

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

   (c) $t^{\star} = 3.71$

   (d) $SE = 1.68$

   (e) $\mid t_{\textbf{obs}} \mid = 3.799$

   (f) $0.005 < p$**-value** < 0.01

   (g) **reject**

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

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

   (c) $t^{\star} = 2.2$

   (d) $SE = 17.9$

   (e) $\mid t_{\textbf{obs}} \mid = 2.18$

   (f) $0.05 < p$**-value** < 0.1

   (g) **retain**

4. (a) **LB of p CI = 0.596 or** 59.6%

   (b) **UB of p CI = 0.604 or** 60.4%

5. $n \approx 17000$

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

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

(c) $z^{\star} = 2.05$

(d) $SE = 0.022$

(e) $\mid z_{\mathbf{obs}} \mid = 2.28$

(f) p-**value** = 0.0226

(g) **reject**

1. **Problem**
As an ornithologist, you wish to determine the average body mass of *Helmitheros vermivorus*. You randomly capture 19 adults of *Helmitheros vermivorus*, resulting in a sample mean of 16.11 grams and a sample standard deviation of 2.43 grams. You decide to report a 99.5% 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.

$$\begin{aligned} n &= 19 \\ \bar{x} &= 16.11 \\ s &= 2.43 \\ CL &= 0.995 \end{aligned}$$

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

$$df = n - 1 = 18$$

   Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.995$.

$$t^\star = 3.2$$

   Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{2.43}{\sqrt{19}} = 0.557$$

   We want to make an inference about the population mean.

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

   Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (16.11 - 3.2 \times 0.557,\ 16.11 - 3.2 \times 0.557) \\ &= (14.3,\ 17.9) \end{aligned}$$

   We are 99.5% confident that the population mean is between 14.3 and 17.9.

   (a) Lower bound = 14.3

   (b) Upper bound = 17.9

2. **Problem**
A teacher has 7 students who have each taken two quizzes. Perform a two-tail test with significance level 0.01 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6	student7
quiz 1:	56.1	72.1	80	75.5	83	59.2	81
quiz 2:	64.8	70.2	87.9	82.9	85.3	68.2	89.9

(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.

$$n = 7$$
$$\alpha = 0.01$$

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 $P(|T| > t^\star) = 0.01$.

$$t^\star = 3.71$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6	student7
quiz2-quiz1:	8.7	-1.9	7.9	7.4	2.3	9	8.9

Find the sample mean.

$$\overline{x_{\text{diff}}} = 6.04$$

Find the sample standard deviation.

$$s_{\text{diff}} = 4.21$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 1.59$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{6.04 - 0}{1.59} = 3.799$$

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.

$$0.005 < p\text{-value} < 0.01$$

We conclude that we should reject the null hypothesis.

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

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

(c) $t^{\star} = 3.71$

(d) $SE = 1.6798313$

(e) $|t_{\text{obs}}| = 3.799$

(f) $0.005 < p\text{-value} < 0.01$

(g) reject 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 7 cases and a control group with 7 cases. You decide to run a hypothesis test with a significance level of 0.05. Your data is below. Please use 11 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
165	137
172	253
126	206
144	242
172	202
212	179
149	196

   (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.

$$n_1 = 7$$
$$n_2 = 7$$
$$\alpha = 0.05$$

   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.

$$t^\star = 2.2$$

   Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 163$$
$$s_1 = 27.4$$
$$\bar{x}_2 = 202$$
$$s_2 = 38.7$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(27.4)^2}{7} + \frac{(38.7)^2}{7}} \\ &= 17.9 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(202 - 163) - (0)}{17.9} \\ &= 2.18 \end{aligned}$$

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.

$$0.05 < p\text{-value} < 0.1$$

Compare p-value and α.

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

We conclude that we should retain the null hypothesis.

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

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

(c) $t^\star = 2.2$

(d) $SE = 17.9$

(e) $|t_{\text{obs}}| = 2.18$

(f) $0.05 < p\text{-value} < 0.1$

(g) retain the null

4. **Problem**
From a very large population, a random sample of 70000 individuals was taken. In that sample, 60% were sweet. 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.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.6)(1-0.6)}{70000}} = 0.00185$$

Calculate the margin of error.

$$ME = z^\star SE = (1.96)(0.00185) = 0.00363$$

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.

$$(0.596\,,\ 0.604)$$

We are 95% confident that the true population proportion is between 59.6% and 60.4%.

   (a) The lower bound = 0.596, which can also be expressed as 59.6%.
   (b) The upper bound = 0.604, which can also be expressed as 60.4%.

5. **Problem**

   Your boss wants to know what proportion of a very large population is happy. She also wants to guarantee that the margin of error of a 99% confidence interval will be less than 0.01 (which is 1 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

   Solution

   Determine $z^\star$ such that $P(|Z| < z^\star) = 0.99$.

$$z^\star = 2.58$$

   Determine the maximal standard error.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.01}{2.58} = 0.00388$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.00388)^2} = 16606.4406419$$

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

$$n = 16607$$

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

$$n = 17000$$

6. **Problem**
An experiment is run with a treatment group of size 290 and a control group of size 304. The results are summarized in the table below.

	treatment	control
preoccupied	260	288
not preoccupied	30	16

Using a significance level of 0.04, determine whether the treatment causes an effect on the proportion of cases that are preoccupied.

(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 $P(|Z| > z^\star) = 0.04$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 2.05$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{260}{290} = 0.897$$

$$\hat{p}_2 = \frac{288}{304} = 0.947$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.947 - 0.897 = 0.05$$

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

$$\hat{p} = \frac{260 + 288}{290 + 304} = 0.923$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.923)(0.077)}{290} + \frac{(0.923)(0.077)}{304}}$$

$$= 0.0219$$

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

$$H_0 : \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0219)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.947 - 0.897) - 0}{0.0219} \\ &= 2.28 \end{aligned}$$

Determine the *p*-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.28) \\ &= 0.0226 \end{aligned}$$

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) $H_0 : \quad p_2 - p_1 = 0$

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

(c) $z^\star = 2.05$

(d) $SE = 0.0219$

(e) $|z_{\text{obs}}| = 2.28$

(f) p-value = 0.0226

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 019

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 = 29.9**

   (b) **UB = 31.1**

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

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

   (c) $t^{\star} = 2.02$

   (d) $SE = 0.893$

   (e) $\mid t_{\textbf{obs}} \mid = 2.191$

   (f) $0.05 < p$**-value** < 0.1

   (g) **reject**

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

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

   (c) $t^{\star} = 3$

   (d) $SE = 13.6$

   (e) $\mid t_{\textbf{obs}} \mid = 3.16$

   (f) $0.02 < p$**-value** < 0.04

   (g) **reject**

4. (a) **LB of p CI = 0.736 or** 73.6%

   (b) **UB of p CI = 0.828 or** 82.8%

5. $n \approx 4200$

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) $SE = 0.039$

(e) $\mid z_{\mathbf{obs}} \mid = 2.54$

(f) p-**value** $= 0.011$

(g) **reject**

1. **Problem**
   As an ornithologist, you wish to determine the average body mass of *Chaetura pelagica*. You randomly capture 22 adults of *Chaetura pelagica*, resulting in a sample mean of 30.52 grams and a sample standard deviation of 1.36 grams. You decide to report a 95% 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.

$$\begin{aligned} n &= 22 \\ \bar{x} &= 30.52 \\ s &= 1.36 \\ CL &= 0.95 \end{aligned}$$

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

$$df = n - 1 = 21$$

   Determine the critical t value, $t^{\star}$, such that $P(|T| < t^{\star}) = 0.95$.

$$t^{\star} = 2.08$$

   Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.36}{\sqrt{22}} = 0.29$$

   We want to make an inference about the population mean.

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

   Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^{\star} SE,\ \bar{x} + t^{\star} SE) \\ &= (30.52 - 2.08 \times 0.29,\ 30.52 - 2.08 \times 0.29) \\ &= (29.9,\ 31.1) \end{aligned}$$

   We are 95% confident that the population mean is between 29.9 and 31.1.

   (a) Lower bound = 29.9

   (b) Upper bound = 31.1

2. **Problem**
A teacher has 6 students who have each taken two quizzes. Perform a two-tail test with significance level 0.1 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6
quiz 1:	64.9	62.5	63.8	64.5	73.4	59.4
quiz 2:	65.3	59.9	64.3	60.7	72.2	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.

$$n = 6$$
$$\alpha = 0.1$$

State the hypotheses.

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

Determine the degrees of freedom.

$$df = n - 1 = 5$$

We determine $t^{\star}$ such that $P(|T| > t^{\star}) = 0.1$.

$$t^{\star} = 2.02$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6
quiz2-quiz1:	0.4	-2.6	0.5	-3.8	-1.2	-3.9

Find the sample mean.

$$\overline{x_{\text{diff}}} = -1.77$$

Find the sample standard deviation.

$$s_{\text{diff}} = 1.98$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 0.808$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{-1.77 - 0}{0.808} = -2.191$$

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.

$$0.05 < p\text{-value} < 0.1$$

We conclude that we should reject the null hypothesis.

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

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

(c) $t^{\star} = 2.02$

(d) $SE = 0.8931109$

(e) $|t_{\text{obs}}| = 2.191$

(f) $0.05 < p\text{-value} < 0.1$

(g) reject 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 0.04. Your data is below. Please use 4 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
139	199
137	220
126	152
138	160
133	158

(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.

$$n_1 = 5$$
$$n_2 = 5$$
$$\alpha = 0.04$$

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.

$$t^{\star} = 3$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 135$$
$$s_1 = 5.32$$
$$\bar{x}_2 = 178$$
$$s_2 = 30$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(5.32)^2}{5} + \frac{(30)^2}{5}} \\ &= 13.6 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(178 - 135) - (0)}{13.6} \\ &= 3.16 \end{aligned}$$

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.

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

Compare p-value and α.

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

We conclude that we should reject the null hypothesis.

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

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

(c) $t^\star = 3$

(d) $SE = 13.6$

(e) $|t_{\text{obs}}| = 3.16$

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

(g) reject the null

4. **Problem**
   From a very large population, a random sample of 440 individuals was taken. In that sample, 78.2% were super. Determine a 98% 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 $P(|Z| < z^\star) = 0.98$.

$$z^\star = 2.33$$

Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.782)(1-0.782)}{440}} = 0.0197$$

Calculate the margin of error.

$$ME = z^\star SE = (2.33)(0.0197) = 0.0459$$

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.

$$(0.736\,,\ 0.828)$$

We are 98% confident that the true population proportion is between 73.6% and 82.8%.

(a) The lower bound = 0.736, which can also be expressed as 73.6%.
(b) The upper bound = 0.828, which can also be expressed as 82.8%.

5. **Problem**

Your boss wants to know what proportion of a very large population is cold. She also wants to guarantee that the margin of error of a 99% confidence interval will be less than 0.02 (which is 2 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

Solution

Determine $z^\star$ such that $P(|Z| < z^\star) = 0.99$.

$$z^\star = 2.58$$

Determine the maximal standard error.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.02}{2.58} = 0.00775$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.00775)^2} = 4162.3309053$$

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

$$n = 4163$$

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

$$n = 4200$$

6. **Problem**

An experiment is run with a treatment group of size 222 and a control group of size 217. The results are summarized in the table below.

	treatment	control
reclusive	37	58
not reclusive	185	159

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

(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.

$$\hat{p}_1 = \frac{37}{222} = 0.167$$

$$\hat{p}_2 = \frac{58}{217} = 0.267$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.267 - 0.167 = 0.1$$

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

$$\hat{p} = \frac{37 + 58}{222 + 217} = 0.216$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.216)(0.784)}{222} + \frac{(0.216)(0.784)}{217}}$$

$$= 0.0393$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0393)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.267 - 0.167) - 0}{0.0393} \\ &= 2.54 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.54) \\ &= 0.011 \end{aligned}$$

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) $H_0: \quad p_2 - p_1 = 0$

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

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

(d) $SE = 0.0393$

(e) $|z_{\text{obs}}| = 2.54$

(f) p-value = 0.011

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 020

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 = 21.4**

   (b) **UB = 22.9**

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

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

   (c) $t^\star = 2.78$

   (d) $SE = 1.47$

   (e) $\mid t_{\mathbf{obs}} \mid = 2.817$

   (f) $0.04 < p$**-value** < 0.05

   (g) **reject**

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

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

   (c) $t^\star = 1.83$

   (d) $SE = 80.1$

   (e) $\mid t_{\mathbf{obs}} \mid = 1.87$

   (f) $0.05 < p$**-value** < 0.1

   (g) **reject**

4. (a) **LB of p CI = 0.274 or** 27.4%

   (b) **UB of p CI = 0.294 or** 29.4%

5. $n \approx 17000$

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

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

(c) $z^{\star} = 1.28$

(d) $SE = 0.071$

(e) $|\ z_{\mathbf{obs}}\ | = 1.56$

(f) p-**value** = 0.1188

(g) **reject**

1. **Problem**
   As an ornithologist, you wish to determine the average body mass of *Ammodramus maritimus*. You randomly capture 17 adults of *Ammodramus maritimus*, resulting in a sample mean of 22.18 grams and a sample standard deviation of 1.75 grams. You decide to report a 90% 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.

$$\begin{aligned} n &= 17 \\ \bar{x} &= 22.18 \\ s &= 1.75 \\ CL &= 0.9 \end{aligned}$$

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

$$df = n - 1 = 16$$

   Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.9$.

$$t^\star = 1.75$$

   Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.75}{\sqrt{17}} = 0.424$$

   We want to make an inference about the population mean.

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

   Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE, \ \bar{x} + t^\star SE) \\ &= (22.18 - 1.75 \times 0.424, \ 22.18 - 1.75 \times 0.424) \\ &= (21.4, \ 22.9) \end{aligned}$$

   We are 90% confident that the population mean is between 21.4 and 22.9.

   (a) Lower bound = 21.4

   (b) Upper bound = 22.9

2. **Problem**
A teacher has 5 students who have each taken two quizzes. Perform a two-tail test with significance level 0.05 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5
quiz 1:	83.5	70.6	71.3	86.4	72.3
quiz 2:	82.7	73.4	74	90.1	77

(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.

$$n = 5$$
$$\alpha = 0.05$$

State the hypotheses.

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

Determine the degrees of freedom.

$$df = n - 1 = 4$$

We determine $t^\star$ such that $P(|T| > t^\star) = 0.05$.

$$t^\star = 2.78$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5
quiz2-quiz1:	-0.8	2.8	2.7	3.7	4.7

Find the sample mean.

$$\overline{x_{\text{diff}}} = 2.62$$

Find the sample standard deviation.

$$s_{\text{diff}} = 2.08$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 0.93$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{2.62 - 0}{0.93} = 2.817$$

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.

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

We conclude that we should reject the null hypothesis.

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

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

(c) $t^{\star} = 2.78$

(d) $SE = 1.4717704$

(e) $|t_{\text{obs}}| = 2.817$

(f) $0.04 < p$-value < 0.05

(g) reject 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 6 cases and a control group with 6 cases. You decide to run a hypothesis test with a significance level of 0.1. Your data is below. Please use 9 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
820	460
430	650
630	530
510	580
830	360
770	510

(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.

$$n_1 = 6$$
$$n_2 = 6$$
$$\alpha = 0.1$$

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.

$$t^\star = 1.83$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 665$$
$$s_1 = 169$$
$$\bar{x}_2 = 515$$
$$s_2 = 99.7$$

Calculate the standard error.

$$SE = \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$$
$$= \sqrt{\frac{(169)^2}{6} + \frac{(99.7)^2}{6}}$$
$$= 80.1$$

Determine the test statistic.

$$t_{\text{obs}} = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE}$$
$$= \frac{(515 - 665) - (0)}{80.1}$$
$$= -1.87$$

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.

$$0.05 < p\text{-value} < 0.1$$

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) $t^\star = 1.83$

(d) $SE = 80.1$

(e) $|t_{\text{obs}}| = 1.87$

(f) $0.05 < p\text{-value} < 0.1$

(g) reject the null

4. **Problem**
From a very large population, a random sample of 7700 individuals was taken. In that sample, 28.4% were glowing. 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.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.284)(1-0.284)}{7700}} = 0.00514$$

Calculate the margin of error.

$$ME = z^\star SE = (1.96)(0.00514) = 0.0101$$

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.

$$(0.274\,,\ 0.294)$$

We are 95% confident that the true population proportion is between 27.4% and 29.4%.

   (a) The lower bound = 0.274, which can also be expressed as 27.4%.
   (b) The upper bound = 0.294, which can also be expressed as 29.4%.

5. **Problem**
   Your boss wants to know what proportion of a very large population is angry. She also wants to guarantee that the margin of error of a 99% confidence interval will be less than 0.01 (which is 1 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

   Solution
   Determine $z^\star$ such that $P(|Z| < z^\star) = 0.99$.

$$z^\star = 2.58$$

   Determine the maximal standard error.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.01}{2.58} = 0.00388$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.00388)^2} = 16606.4406419$$

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

$$n = 16607$$

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

$$n = 17000$$

6. **Problem**
An experiment is run with a treatment group of size 72 and a control group of size 98. The results are summarized in the table below.

	treatment	control
green	55	64
not green	17	34

Using a significance level of 0.2, determine whether the treatment causes an effect on the proportion of cases that are green.

(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 $P(|Z| > z^\star) = 0.2$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 1.28$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{55}{72} = 0.764$$

$$\hat{p}_2 = \frac{64}{98} = 0.653$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.653 - 0.764 = -0.111$$

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

$$\hat{p} = \frac{55 + 64}{72 + 98} = 0.7$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.7)(0.3)}{72} + \frac{(0.7)(0.3)}{98}}$$

$$= 0.0711$$

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

$$H_0 : \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\, 0.0711)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.653 - 0.764) - 0}{0.0711} \\ &= -1.56 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-1.56) \\ &= 0.1188 \end{aligned}$$

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) $H_0 : \quad p_2 - p_1 = 0$

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

(c) $z^\star = 1.28$

(d) $SE = 0.0711$

(e) $|z_{\text{obs}}| = 1.56$

(f) p-value $= 0.1188$

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 021

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 = 34.6**

   (b) **UB = 38.9**

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

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

   (c) $t^\star = 3.71$

   (d) $SE = 1.4$

   (e) $\mid t_{\textbf{obs}} \mid = 3.612$

   (f) $0.01 < p$**-value** < 0.02

   (g) **retain**

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

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

   (c) $t^\star = 3.17$

   (d) $SE = 0.213$

   (e) $\mid t_{\textbf{obs}} \mid = 3.33$

   (f) $0.005 < p$**-value** < 0.01

   (g) **reject**

4. (a) **LB of p CI = 0.929 or** 92.9%

   (b) **UB of p CI = 0.937 or** 93.7%

5. $n \approx 4200$

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

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

(c) $z^{\star} = 1.28$

(d) $SE = 0.082$

(e) $|\ z_{\mathbf{obs}}\ | = 1.52$

(f) p-**value** $= 0.1286$

(g) **reject**

1. **Problem**
As an ornithologist, you wish to determine the average body mass of *Piranga rubra*. You randomly capture 34 adults of *Piranga rubra*, resulting in a sample mean of 36.74 grams and a sample standard deviation of 6.13 grams. You decide to report a 95% 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.

$$n = 34$$
$$\bar{x} = 36.74$$
$$s = 6.13$$
$$CL = 0.95$$

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

$$df = n - 1 = 33$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.95$.

$$t^\star = 2.03$$

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{6.13}{\sqrt{34}} = 1.05$$

We want to make an inference about the population mean.

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

Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (36.74 - 2.03 \times 1.05,\ 36.74 - 2.03 \times 1.05) \\ &= (34.6,\ 38.9) \end{aligned}$$

We are 95% confident that the population mean is between 34.6 and 38.9.

   (a) Lower bound = 34.6

   (b) Upper bound = 38.9

2. **Problem**
A teacher has 7 students who have each taken two quizzes. Perform a two-tail test with significance level 0.01 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6	student7
quiz 1:	87.8	57.4	88.4	83.3	76.5	69.9	56.1
quiz 2:	86.2	50.8	83.4	80	76.7	60.9	52.1

(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.

$$n = 7$$
$$\alpha = 0.01$$

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 $P(|T| > t^{\star}) = 0.01$.

$$t^{\star} = 3.71$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6	student7
quiz2-quiz1:	-1.6	-6.6	-5	-3.3	0.2	-9	-4

Find the sample mean.

$$\overline{x_{\text{diff}}} = -4.19$$

Find the sample standard deviation.

$$s_{\text{diff}} = 3.07$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 1.16$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{-4.19 - 0}{1.16} = -3.612$$

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.

$$0.01 < p\text{-value} < 0.02$$

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) $t^{\star} = 3.71$

(d) $SE = 1.4036283$

(e) $|t_{\text{obs}}| = 3.612$

(f) $0.01 < p\text{-value} < 0.02$

(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 7 cases and a control group with 7 cases. You decide to run a hypothesis test with a significance level of 0.01. Your data is below. Please use 10 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
2.5	1.68
1.59	1.46
2.12	1.13
2.87	1.76
1.51	1.27
1.78	1.29
2.45	1.29

(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.

$$n_1 = 7$$
$$n_2 = 7$$
$$\alpha = 0.01$$

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.

$$t^\star = 3.17$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 2.12$$
$$s_1 = 0.514$$
$$\bar{x}_2 = 1.41$$
$$s_2 = 0.233$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(0.514)^2}{7} + \frac{(0.233)^2}{7}} \\ &= 0.213 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(1.41 - 2.12) - (0)}{0.213} \\ &= -3.33 \end{aligned}$$

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.

$$0.005 < p\text{-value} < 0.01$$

Compare p-value and α.

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

We conclude that we should reject the null hypothesis.

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

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

(c) $t^\star = 3.17$

(d) $SE = 0.213$

(e) $|t_{\text{obs}}| = 3.33$

(f) $0.005 < p\text{-value} < 0.01$

(g) reject the null

4. **Problem**
   From a very large population, a random sample of 18000 individuals was taken. In that sample, 93.3% were blue. Determine a 98% 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 $P(|Z| < z^\star) = 0.98$.

$$z^\star = 2.33$$

Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.933)(1-0.933)}{18000}} = 0.00186$$

Calculate the margin of error.

$$ME = z^\star SE = (2.33)(0.00186) = 0.00433$$

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.

$$(0.929\,,\ 0.937)$$

We are 98% confident that the true population proportion is between 92.9% and 93.7%.

(a) The lower bound = 0.929, which can also be expressed as 92.9%.
(b) The upper bound = 0.937, which can also be expressed as 93.7%.

5. **Problem**
   Your boss wants to know what proportion of a very large population is broken. She also wants to guarantee that the margin of error of a 99% confidence interval will be less than 0.02 (which is 2 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

   Solution
   Determine $z^\star$ such that $P(|Z| < z^\star) = 0.99$.

$$z^\star = 2.58$$

Determine the maximal standard error.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.02}{2.58} = 0.00775$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.00775)^2} = 4162.3309053$$

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

$$n = 4163$$

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

$$n = 4200$$

6. **Problem**
An experiment is run with a treatment group of size 51 and a control group of size 60. The results are summarized in the table below.

	treatment	control
reclusive	9	18
not reclusive	42	42

Using a significance level of 0.2, determine whether the treatment causes an effect on the proportion of cases that are reclusive.

(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 $P(|Z| > z^\star) = 0.2$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 1.28$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{9}{51} = 0.176$$

$$\hat{p}_2 = \frac{18}{60} = 0.3$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.3 - 0.176 = 0.124$$

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

$$\hat{p} = \frac{9 + 18}{51 + 60} = 0.243$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.243)(0.757)}{51} + \frac{(0.243)(0.757)}{60}}$$

$$= 0.0817$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0817)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.3 - 0.176) - 0}{0.0817} \\ &= 1.52 \end{aligned}$$

Determine the *p*-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-1.52) \\ &= 0.1286 \end{aligned}$$

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) $H_0: \quad p_2 - p_1 = 0$

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

(c) $z^\star = 1.28$

(d) $SE = 0.0817$

(e) $|z_{\text{obs}}| = 1.52$

(f) p-value = 0.1286

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 022

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 = 30**

   (b) **UB = 35.1**

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

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

   (c) $t^{\star} = 1.94$

   (d) $SE = 1.74$

   (e) $\mid t_{\textbf{obs}} \mid = 1.861$

   (f) $0.1 <$ p**-value** < 1

   (g) **retain**

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

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

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

   (d) $SE = 0.41$

   (e) $\mid t_{\textbf{obs}} \mid = 2.58$

   (f) $0.02 <$ p**-value** < 0.04

   (g) **reject**

4. (a) **LB of p CI = 0.349 or** 34.9%

   (b) **UB of p CI = 0.373 or** 37.3%

5. $n \approx 790$

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) $SE = 0.045$

(e) $|\ z_{\mathbf{obs}}\ | = 2.48$

(f) p-**value** $= 0.0132$

(g) **reject**

1. **Problem**
   As an ornithologist, you wish to determine the average body mass of *Catharus ustulatus*. You randomly capture 28 adults of *Catharus ustulatus*, resulting in a sample mean of 32.55 grams and a sample standard deviation of 7.85 grams. You decide to report a 90% 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.

$$\begin{aligned} n &= 28 \\ \bar{x} &= 32.55 \\ s &= 7.85 \\ CL &= 0.9 \end{aligned}$$

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

$$df = n - 1 = 27$$

   Determine the critical t value, $t^{\star}$, such that $P(|T| < t^{\star}) = 0.9$.

$$t^{\star} = 1.7$$

   Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{7.85}{\sqrt{28}} = 1.48$$

   We want to make an inference about the population mean.

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

   Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^{\star}SE,\ \bar{x} + t^{\star}SE) \\ &= (32.55 - 1.7 \times 1.48,\ 32.55 - 1.7 \times 1.48) \\ &= (30,\ 35.1) \end{aligned}$$

   We are 90% confident that the population mean is between 30 and 35.1.

   (a) Lower bound = 30

   (b) Upper bound = 35.1

2. **Problem**
A teacher has 7 students who have each taken two quizzes. Perform a two-tail test with significance level 0.1 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6	student7
quiz 1:	69.6	76.4	87.1	64.4	89.6	63.4	56.2
quiz 2:	69.1	73.7	81.4	64.5	87.9	60.7	57.8

(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.

$$n = 7$$
$$\alpha = 0.1$$

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 $P(|T| > t^\star) = 0.1$.

$$t^\star = 1.94$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6	student7
quiz2-quiz1:	-0.5	-2.7	-5.7	0.1	-1.7	-2.7	1.6

Find the sample mean.

$$\overline{x_{\text{diff}}} = -1.66$$

Find the sample standard deviation.

$$s_{\text{diff}} = 2.36$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 0.892$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{-1.66 - 0}{0.892} = -1.861$$

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.

$$0.1 < p\text{-value} < 1$$

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) $t^{\star} = 1.94$

(d) $SE = 1.7425762$

(e) $|t_{\text{obs}}| = 1.861$

(f) $0.1 < p$-value < 1

(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 7 cases and a control group with 7 cases. You decide to run a hypothesis test with a significance level of 0.04. Your data is below. Please use 11 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
6.6	6.9
6.3	5.2
7.2	6.6
7.3	6
7.9	5.4
6.7	4.5
7	7

(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.

$$n_1 = 7$$
$$n_2 = 7$$
$$\alpha = 0.04$$

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.

$$t^\star = 2.33$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 7$$
$$s_1 = 0.529$$
$$\bar{x}_2 = 5.94$$
$$s_2 = 0.948$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(0.529)^2}{7} + \frac{(0.948)^2}{7}} \\ &= 0.41 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(5.94 - 7) - (0)}{0.41} \\ &= -2.58 \end{aligned}$$

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.

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

Compare p-value and α.

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

We conclude that we should reject the null hypothesis.

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

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

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

(d) $SE = 0.41$

(e) $|t_{\text{obs}}| = 2.58$

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

(g) reject the null

4. **Problem**
   From a very large population, a random sample of 6000 individuals was taken. In that sample, 36.1% were shiny. 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.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.361)(1-0.361)}{6000}} = 0.0062$$

   Calculate the margin of error.

$$ME = z^\star SE = (1.96)(0.0062) = 0.0122$$

   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.

$$(0.349\,,\ 0.373)$$

   We are 95% confident that the true population proportion is between 34.9% and 37.3%.

   (a) The lower bound = 0.349, which can also be expressed as 34.9%.
   (b) The upper bound = 0.373, which can also be expressed as 37.3%.

5. **Problem**
   Your boss wants to know what proportion of a very large population is purple. She also wants to guarantee that the margin of error of a 99.5% confidence interval will be less than 0.05 (which is 5 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

   Solution
   Determine $z^\star$ such that $P(|Z| < z^\star) = 0.995$.

$$z^\star = 2.81$$

Determine the maximal standard error.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.05}{2.81} = 0.0178$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.0178)^2} = 789.0417877$$

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

$$n = 790$$

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

$$n = 790$$

6. **Problem**
An experiment is run with a treatment group of size 222 and a control group of size 241. The results are summarized in the table below.

	treatment	control
reclusive	72	105
not reclusive	150	136

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

(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.

$$\hat{p}_1 = \frac{72}{222} = 0.324$$

$$\hat{p}_2 = \frac{105}{241} = 0.436$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.436 - 0.324 = 0.112$$

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

$$\hat{p} = \frac{72 + 105}{222 + 241} = 0.382$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.382)(0.618)}{222} + \frac{(0.382)(0.618)}{241}}$$

$$= 0.0452$$

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

$$H_0 : \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0452)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.436 - 0.324) - 0}{0.0452} \\ &= 2.48 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.48) \\ &= 0.0132 \end{aligned}$$

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) $H_0 : \quad p_2 - p_1 = 0$

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

(c) $z^\star = 2.33$

(d) $SE = 0.0452$

(e) $|z_{\text{obs}}| = 2.48$

(f) p-value = 0.0132

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 023

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 = 12.3**

   (b) **UB = 13.2**

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

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

   (c) $t^{\star} = 4.32$

   (d) $SE = 1.76$

   (e) $\mid t_{\textbf{obs}} \mid = 4.374$

   (f) $0.004 < p$**-value** < 0.005

   (g) **reject**

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

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

   (c) $t^{\star} = 1.94$

   (d) $SE = 0.095$

   (e) $\mid t_{\textbf{obs}} \mid = 1.9$

   (f) $0.1 < p$**-value** < 0.2

   (g) **retain**

4. (a) **LB of p CI = 0.523 or** 52.3%

   (b) **UB of p CI = 0.549 or** 54.9%

5. $n \approx 130000$

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

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

(c) $z^{\star} = 2.05$

(d) $SE = 0.07$

(e) $\mid z_{\mathbf{obs}} \mid = 2.27$

(f) p-**value** = 0.0232

(g) **reject**

1. **Problem**
   As an ornithologist, you wish to determine the average body mass of *Vermivora peregrina*. You randomly capture 16 adults of *Vermivora peregrina*, resulting in a sample mean of 12.76 grams and a sample standard deviation of 1.06 grams. You decide to report a 90% 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.

$$n = 16$$
$$\bar{x} = 12.76$$
$$s = 1.06$$
$$CL = 0.9$$

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

$$df = n - 1 = 15$$

   Determine the critical t value, $t^{\star}$, such that $P(|T| < t^{\star}) = 0.9$.

$$t^{\star} = 1.75$$

   Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.06}{\sqrt{16}} = 0.265$$

   We want to make an inference about the population mean.

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

   Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^{\star} SE, \ \bar{x} + t^{\star} SE) \\ &= (12.76 - 1.75 \times 0.265, \ 12.76 - 1.75 \times 0.265) \\ &= (12.3, \ 13.2) \end{aligned}$$

   We are 90% confident that the population mean is between 12.3 and 13.2.

   (a) Lower bound = 12.3

   (b) Upper bound = 13.2

2. **Problem**
A teacher has 7 students who have each taken two quizzes. Perform a two-tail test with significance level 0.005 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6	student7
quiz 1:	67.2	52	89.3	74	81.7	50.6	79.2
quiz 2:	77.4	51.5	102	90.2	89	62.1	84.9

(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.

$$n = 7$$
$$\alpha = 0.005$$

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 $P(|T| > t^\star) = 0.005$.

$$t^\star = 4.32$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6	student7
quiz2-quiz1:	10.2	-0.5	12.7	16.2	7.3	11.5	5.7

Find the sample mean.

$$\overline{x_{\text{diff}}} = 9.01$$

Find the sample standard deviation.

$$s_{\text{diff}} = 5.44$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 2.06$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{9.01 - 0}{2.06} = 4.374$$

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.

$$0.004 < p\text{-value} < 0.005$$

We conclude that we should reject the null hypothesis.

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

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

(c) $t^{\star} = 4.32$

(d) $SE = 1.7554899$

(e) $|t_{\text{obs}}| = 4.374$

(f) $0.004 < p\text{-value} < 0.005$

(g) reject 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 0.1. Your data is below. Please use 6 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
1.64	1.54
1.57	1.82
1.43	1.78
1.37	1.78
1.44	1.41

(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.

$$n_1 = 5$$
$$n_2 = 5$$
$$\alpha = 0.1$$

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.

$$t^\star = 1.94$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 1.49$$
$$s_1 = 0.111$$
$$\bar{x}_2 = 1.67$$
$$s_2 = 0.181$$

Calculate the standard error.

$$
\begin{aligned}
SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\
&= \sqrt{\frac{(0.111)^2}{5} + \frac{(0.181)^2}{5}} \\
&= 0.095
\end{aligned}
$$

Determine the test statistic.

$$
\begin{aligned}
t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\
&= \frac{(1.67 - 1.49) - (0)}{0.095} \\
&= 1.9
\end{aligned}
$$

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.

$$0.1 < p\text{-value} < 0.2$$

Compare p-value and α.

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

We conclude that we should retain the null hypothesis.

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

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

(c) $t^\star = 1.94$

(d) $SE = 0.095$

(e) $|t_{\text{obs}}| = 1.9$

(f) $0.1 < p\text{-value} < 0.2$

(g) retain the null

4. **Problem**

   From a very large population, a random sample of 8300 individuals was taken. In that sample, 53.6% were tasty. Determine a 98% 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 $P(|Z| < z^\star) = 0.98$.

$$z^\star = 2.33$$

   Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.536)(1-0.536)}{8300}} = 0.00547$$

   Calculate the margin of error.

$$ME = z^\star SE = (2.33)(0.00547) = 0.0127$$

   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.

$$(0.523\,,\ 0.549)$$

   We are 98% confident that the true population proportion is between 52.3% and 54.9%.

   (a) The lower bound = 0.523, which can also be expressed as 52.3%.

   (b) The upper bound = 0.549, which can also be expressed as 54.9%.

5. **Problem**

   Your boss wants to know what proportion of a very large population is messy. She also wants to guarantee that the margin of error of a 99.5% confidence interval will be less than 0.004 (which is 0.4 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

   Solution

   Determine $z^\star$ such that $P(|Z| < z^\star) = 0.995$.

$$z^\star = 2.81$$

Determine the maximal standard error.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.004}{2.81} = 0.00142$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.00142)^2} = 123983.3366396$$

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

$$n = 123984$$

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

$$n = 130000$$

6. **Problem**
An experiment is run with a treatment group of size 93 and a control group of size 113. The results are summarized in the table below.

	treatment	control
sick	60	55
not sick	33	58

Using a significance level of 0.04, 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 $P(|Z| > z^\star) = 0.04$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 2.05$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{60}{93} = 0.645$$

$$\hat{p}_2 = \frac{55}{113} = 0.487$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.487 - 0.645 = -0.158$$

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

$$\hat{p} = \frac{60 + 55}{93 + 113} = 0.558$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.558)(0.442)}{93} + \frac{(0.558)(0.442)}{113}}$$

$$= 0.0695$$

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

$$H_0 : \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0695)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.487 - 0.645) - 0}{0.0695} \\ &= -2.27 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.27) \\ &= 0.0232 \end{aligned}$$

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) $H_0 : \quad p_2 - p_1 = 0$

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

(c) $z^{\star} = 2.05$

(d) $SE = 0.0695$

(e) $|z_{\text{obs}}| = 2.27$

(f) p-value = 0.0232

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 024

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 = 22.4**

   (b) **UB = 23.7**

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

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

   (c) $t^{\star} = 2.45$

   (d) $SE = 0.909$

   (e) $\mid t_{\textbf{obs}} \mid = 2.254$

   (f) $0.05 < p$**-value** < 0.1

   (g) **retain**

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

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

   (c) $t^{\star} = 2.36$

   (d) $SE = 76.4$

   (e) $\mid t_{\textbf{obs}} \mid = 2.62$

   (f) $0.02 < p$**-value** < 0.04

   (g) **reject**

4. (a) **LB of p CI = 0.719 or** 71.9%

   (b) **UB of p CI = 0.803 or** 80.3%

5. $n \approx 22000$

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

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

(c) $z^{\star} = 2.58$

(d) $SE = 0.023$

(e) $|\ z_{\mathbf{obs}}\ | = 2.79$

(f) p**-value** $= 0.0052$

(g) **reject**

1. **Problem**
   As an ornithologist, you wish to determine the average body mass of *Icterus spurius*. You randomly capture 18 adults of *Icterus spurius*, resulting in a sample mean of 23.05 grams and a sample standard deviation of 1.98 grams. You decide to report a 80% 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.

$$\begin{aligned} n &= 18 \\ \bar{x} &= 23.05 \\ s &= 1.98 \\ CL &= 0.8 \end{aligned}$$

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

$$df = n - 1 = 17$$

   Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.8$.

$$t^\star = 1.33$$

   Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.98}{\sqrt{18}} = 0.467$$

   We want to make an inference about the population mean.

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

   Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE, \ \bar{x} + t^\star SE) \\ &= (23.05 - 1.33 \times 0.467, \ 23.05 - 1.33 \times 0.467) \\ &= (22.4, \ 23.7) \end{aligned}$$

   We are 80% confident that the population mean is between 22.4 and 23.7.

   (a) Lower bound = 22.4

   (b) Upper bound = 23.7

2. **Problem**
A teacher has 7 students who have each taken two quizzes. Perform a two-tail test with significance level 0.05 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6	student7
quiz 1:	66.3	67	76.1	82.8	60	59.7	52.3
quiz 2:	63.4	65.5	75.1	79.7	62.2	58	48.6

(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.

$$n = 7$$
$$\alpha = 0.05$$

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 $P(|T| > t^\star) = 0.05$.

$$t^\star = 2.45$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6	student7
quiz2-quiz1:	-2.9	-1.5	-1	-3.1	2.2	-1.7	-3.7

Find the sample mean.

$$\overline{x_{\text{diff}}} = -1.67$$

Find the sample standard deviation.

$$s_{\text{diff}} = 1.96$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 0.741$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{-1.67 - 0}{0.741} = -2.254$$

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.

$$0.05 < p\text{-value} < 0.1$$

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) $t^{\star} = 2.45$

(d) $SE = 0.9093665$

(e) $|t_{\text{obs}}| = 2.254$

(f) $0.05 < p\text{-value} < 0.1$

(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 6 cases and a control group with 6 cases. You decide to run a hypothesis test with a significance level of 0.05. Your data is below. Please use 7 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
530	750
510	510
490	720
550	580
530	720
520	1050

(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.

$$n_1 = 6$$
$$n_2 = 6$$
$$\alpha = 0.05$$

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.

$$t^{\star} = 2.36$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 522$$
$$s_1 = 20.4$$
$$\bar{x}_2 = 722$$
$$s_2 = 186$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(20.4)^2}{6} + \frac{(186)^2}{6}} \\ &= 76.4 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(722 - 522) - (0)}{76.4} \\ &= 2.62 \end{aligned}$$

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.

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

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) $t^{\star} = 2.36$

(d) $SE = 76.4$

(e) $|t_{\text{obs}}| = 2.62$

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

(g) reject the null

4. **Problem**
From a very large population, a random sample of 440 individuals was taken. In that sample, 76.1% were floating. Determine a 96% 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 $P(|Z| < z^\star) = 0.96$.

$$z^\star = 2.05$$

Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.761)(1-0.761)}{440}} = 0.0203$$

Calculate the margin of error.

$$ME = z^\star SE = (2.05)(0.0203) = 0.0416$$

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.

$$(0.719\,,\ 0.803)$$

We are 96% confident that the true population proportion is between 71.9% and 80.3%.

   (a) The lower bound = 0.719, which can also be expressed as 71.9%.
   (b) The upper bound = 0.803, which can also be expressed as 80.3%.

5. **Problem**
   Your boss wants to know what proportion of a very large population is bitter. She also wants to guarantee that the margin of error of a 96% confidence interval will be less than 0.007 (which is 0.7 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.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.007}{2.05} = 0.00341$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.00341)^2} = 21499.6431059$$

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

$$n = 21500$$

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

$$n = 22000$$

6. **Problem**
An experiment is run with a treatment group of size 272 and a control group of size 306. The results are summarized in the table below.

	treatment	control
cold	33	17
not cold	239	289

Using a significance level of 0.01, determine whether the treatment causes an effect on the proportion of cases that are cold.

(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 $P(|Z| > z^\star) = 0.01$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 2.58$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{33}{272} = 0.121$$

$$\hat{p}_2 = \frac{17}{306} = 0.0556$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.0556 - 0.121 = -0.0654$$

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

$$\hat{p} = \frac{33 + 17}{272 + 306} = 0.0865$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.0865)(0.9135)}{272} + \frac{(0.0865)(0.9135)}{306}}$$

$$= 0.0234$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0234)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.0556 - 0.121) - 0}{0.0234} \\ &= -2.79 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.79) \\ &= 0.0052 \end{aligned}$$

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) $H_0: \quad p_2 - p_1 = 0$

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

(c) $z^\star = 2.58$

(d) $SE = 0.0234$

(e) $|z_{\text{obs}}| = 2.79$

(f) p-value = 0.0052

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 025

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 = 30.8**

   (b) **UB = 37.9**

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

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

   (c) $t^{\star} = 2.13$

   (d) $SE = 2.01$

   (e) $\mid t_{\textbf{obs}} \mid = 1.962$

   (f) $0.1 < p$**-value** < 1

   (g) **retain**

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

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

   (c) $t^{\star} = 3.75$

   (d) $SE = 0.428$

   (e) $\mid t_{\textbf{obs}} \mid = 3.46$

   (f) $0.02 < p$**-value** < 0.04

   (g) **retain**

4. (a) **LB of p CI = 0.101 or** 10.1%

   (b) **UB of p CI = 0.105 or** 10.5%

5. $n \approx 8300$

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) $SE = 0.105$

(e) $|\ z_{\mathbf{obs}}\ | = 2.31$

(f) p-**value** = 0.0208

(g) **retain**

1. **Problem**

As an ornithologist, you wish to determine the average body mass of *Catharus ustulatus*. You randomly capture 17 adults of *Catharus ustulatus*, resulting in a sample mean of 34.37 grams and a sample standard deviation of 5.66 grams. You decide to report a 98% 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.

$$\begin{aligned} n &= 17 \\ \bar{x} &= 34.37 \\ s &= 5.66 \\ CL &= 0.98 \end{aligned}$$

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

$$df = n - 1 = 16$$

Determine the critical t value, $t^{\star}$, such that $P(|T| < t^{\star}) = 0.98$.

$$t^{\star} = 2.58$$

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{5.66}{\sqrt{17}} = 1.37$$

We want to make an inference about the population mean.

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

Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^{\star} SE, \; \bar{x} + t^{\star} SE) \\ &= (34.37 - 2.58 \times 1.37, \; 34.37 - 2.58 \times 1.37) \\ &= (30.8, \; 37.9) \end{aligned}$$

We are 98% confident that the population mean is between 30.8 and 37.9.

(a) Lower bound = 30.8

(b) Upper bound = 37.9

2. **Problem**
A teacher has 5 students who have each taken two quizzes. Perform a two-tail test with significance level 0.1 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5
quiz 1:	86.9	78.4	69.9	50.3	60
quiz 2:	90.4	89.2	75.4	48.2	62.8

(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.

$$n = 5$$
$$\alpha = 0.1$$

State the hypotheses.

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

Determine the degrees of freedom.

$$df = n - 1 = 4$$

We determine $t^\star$ such that $P(|T| > t^\star) = 0.1$.

$$t^\star = 2.13$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5
quiz2-quiz1:	3.5	10.8	5.5	-2.1	2.8

Find the sample mean.

$$\overline{x_{\text{diff}}} = 4.1$$

Find the sample standard deviation.

$$s_{\text{diff}} = 4.67$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 2.09$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{4.1 - 0}{2.09} = 1.962$$

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.

$$0.1 < p\text{-value} < 1$$

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) $t^{\star} = 2.13$

(d) $SE = 2.0142354$

(e) $|t_{\text{obs}}| = 1.962$

(f) $0.1 < p\text{-value} < 1$

(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 0.02. Your data is below. Please use 4 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
4.4	5
3.5	5.4
4.1	5.2
4.5	5.3
2.2	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.

$$n_1 = 5$$
$$n_2 = 5$$
$$\alpha = 0.02$$

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.

$$t^\star = 3.75$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 3.74$$
$$s_1 = 0.945$$
$$\bar{x}_2 = 5.22$$
$$s_2 = 0.148$$

Calculate the standard error.

$$SE = \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$$
$$= \sqrt{\frac{(0.945)^2}{5} + \frac{(0.148)^2}{5}}$$
$$= 0.428$$

Determine the test statistic.

$$t_{\text{obs}} = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE}$$
$$= \frac{(5.22 - 3.74) - (0)}{0.428}$$
$$= 3.46$$

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.

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

Compare p-value and α.

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

We conclude that we should retain the null hypothesis.

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

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

(c) $t^\star = 3.75$

(d) $SE = 0.428$

(e) $|t_{\text{obs}}| = 3.46$

(f) $0.02 < p$-value < 0.04

(g) retain the null

4. **Problem**
   From a very large population, a random sample of 82000 individuals was taken. In that sample, 10.3% were tasty. Determine a 96% 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 $P(|Z| < z^{\star}) = 0.96$.

$$z^{\star} = 2.05$$

   Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.103)(1-0.103)}{82000}} = 0.00106$$

   Calculate the margin of error.

$$ME = z^{\star}SE = (2.05)(0.00106) = 0.00217$$

   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.

$$(0.101\ ,\ 0.105)$$

   We are 96% confident that the true population proportion is between 10.1% and 10.5%.

   (a) The lower bound = 0.101, which can also be expressed as 10.1%.

   (b) The upper bound = 0.105, which can also be expressed as 10.5%.

5. **Problem**
   Your boss wants to know what proportion of a very large population is blue. She also wants to guarantee that the margin of error of a 90% confidence interval will be less than 0.009 (which is 0.9 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

   Solution
   Determine $z^\star$ such that $P(|Z| < z^\star) = 0.9$.

$$z^\star = 1.64$$

   Determine the maximal standard error.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.009}{1.64} = 0.00549$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.00549)^2} = 8294.5975627$$

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

$$n = 8295$$

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

$$n = 8300$$

6. **Problem**
   An experiment is run with a treatment group of size 70 and a control group of size 28. The results are summarized in the table below.

	treatment	control
angry	52	14
not angry	18	14

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

(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.

$$\hat{p}_1 = \frac{52}{70} = 0.743$$

$$\hat{p}_2 = \frac{14}{28} = 0.5$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.5 - 0.743 = -0.243$$

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

$$\hat{p} = \frac{52 + 14}{70 + 28} = 0.673$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.673)(0.327)}{70} + \frac{(0.673)(0.327)}{28}}$$

$$= 0.105$$

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

$$H_0 : \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0, 0.105)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.5 - 0.743) - 0}{0.105} \\ &= -2.31 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.31) \\ &= 0.0208 \end{aligned}$$

Compare the p-value to the signficance level.

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

So, we retain the null hypothesis. Thus the difference in proportions is not significant.

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

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

(c) $z^\star = 2.33$

(d) $SE = 0.105$

(e) $|z_{\text{obs}}| = 2.31$

(f) p-value = 0.0208

(g) retain the null

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) $SE = 0.948$

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

   (f) $0.02 <$ p**-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**-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) $SE = 0.9482569$

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

(f) $0.02 < p$-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) $H_0: \quad \mu_2 - \mu_1 = 0$

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

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

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

(c) $z^\star = 2.33$

(d) $SE = 0.0486$

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

(f) p-value = 0.0136

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 027

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 = 34.4**

   (b) **UB = 42.2**

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

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

   (c) $t^{\star} = 3.71$

   (d) $SE = 1.82$

   (e) $\mid t_{\textbf{obs}} \mid = 3.905$

   (f) $0.005 < p$**-value** < 0.01

   (g) **reject**

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

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

   (c) $t^{\star} = 2.61$

   (d) $SE = 0.444$

   (e) $\mid t_{\textbf{obs}} \mid = 2.53$

   (f) $0.04 < p$**-value** < 0.05

   (g) **retain**

4. (a) **LB of p CI = 0.664 or** 66.4%

   (b) **UB of p CI = 0.718 or** 71.8%

5. $n \approx 300$

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

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

(c) $z^{\star} = 2.05$

(d) $SE = 0.085$

(e) $|\ z_{\mathbf{obs}}\ | = 2.29$

(f) p-**value** $= 0.022$

(g) **reject**

1. **Problem**
   As an ornithologist, you wish to determine the average body mass of *Dumetella carolinensis*. You randomly capture 14 adults of *Dumetella carolinensis*, resulting in a sample mean of 38.31 grams and a sample standard deviation of 4.3 grams. You decide to report a 99.5% 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.

$$\begin{aligned} n &= 14 \\ \bar{x} &= 38.31 \\ s &= 4.3 \\ CL &= 0.995 \end{aligned}$$

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

$$df = n - 1 = 13$$

   Determine the critical t value, $t^{\star}$, such that $P(|T| < t^{\star}) = 0.995$.

$$t^{\star} = 3.37$$

   Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{4.3}{\sqrt{14}} = 1.15$$

   We want to make an inference about the population mean.

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

   Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^{\star} SE, \ \bar{x} + t^{\star} SE) \\ &= (38.31 - 3.37 \times 1.15, \ 38.31 - 3.37 \times 1.15) \\ &= (34.4, \ 42.2) \end{aligned}$$

   We are 99.5% confident that the population mean is between 34.4 and 42.2.

   (a) Lower bound = 34.4

   (b) Upper bound = 42.2

2. **Problem**
A teacher has 7 students who have each taken two quizzes. Perform a two-tail test with significance level 0.01 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6	student7
quiz 1:	62.6	64.7	65.1	53.7	61.9	84.6	64.2
quiz 2:	72.5	74.1	68.2	67.7	60.5	95.1	79.1

(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.

$$n = 7$$
$$\alpha = 0.01$$

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 $P(|T| > t^{\star}) = 0.01$.

$$t^{\star} = 3.71$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6	student7
quiz2-quiz1:	9.9	9.4	3.1	14	-1.4	10.5	14.9

Find the sample mean.

$$\overline{x_{\text{diff}}} = 8.63$$

Find the sample standard deviation.

$$s_{\text{diff}} = 5.85$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 2.21$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{8.63 - 0}{2.21} = 3.905$$

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.

$$0.005 < p\text{-value} < 0.01$$

We conclude that we should reject the null hypothesis.

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

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

(c) $t^{\star} = 3.71$

(d) $SE = 1.8190831$

(e) $|t_{\text{obs}}| = 3.905$

(f) $0.005 < p\text{-value} < 0.01$

(g) reject 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 6 cases and a control group with 6 cases. You decide to run a hypothesis test with a significance level of 0.04. Your data is below. Please use 6 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
8.9	7.6
9.2	9.2
9	6.9
8.7	6.4
8.5	8.6
9.2	8.1

(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.

$$n_1 = 6$$
$$n_2 = 6$$
$$\alpha = 0.04$$

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.

$$t^\star = 2.61$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 8.92$$
$$s_1 = 0.279$$
$$\bar{x}_2 = 7.8$$
$$s_2 = 1.05$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(0.279)^2}{6} + \frac{(1.05)^2}{6}} \\ &= 0.444 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(7.8 - 8.92) - (0)}{0.444} \\ &= -2.53 \end{aligned}$$

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 retain the null hypothesis.

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

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

(c) $t^\star = 2.61$

(d) $SE = 0.444$

(e) $|t_{\text{obs}}| = 2.53$

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

(g) retain the null

4. **Problem**
   From a very large population, a random sample of 2000 individuals was taken. In that sample, 69.1% were messy. Determine a 99% 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 $P(|Z| < z^\star) = 0.99$.

$$z^\star = 2.58$$

Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.691)(1-0.691)}{2000}} = 0.0103$$

Calculate the margin of error.

$$ME = z^\star SE = (2.58)(0.0103) = 0.0266$$

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.

$$(0.664\,,\ 0.718)$$

We are 99% confident that the true population proportion is between 66.4% and 71.8%.

(a) The lower bound = 0.664, which can also be expressed as 66.4%.
(b) The upper bound = 0.718, which can also be expressed as 71.8%.

5. **Problem**
   Your boss wants to know what proportion of a very large population is cold. She also wants to guarantee that the margin of error of a 96% confidence interval will be less than 0.06 (which is 6 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.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.06}{2.05} = 0.0293$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.0293)^2} = 291.2089832$$

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

$$n = 292$$

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

$$n = 300$$

6. **Problem**
An experiment is run with a treatment group of size 48 and a control group of size 87. The results are summarized in the table below.

	treatment	control
angry	38	52
not angry	10	35

Using a significance level of 0.04, determine whether the treatment causes an effect on the proportion of cases that are angry.

(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 $P(|Z| > z^{\star}) = 0.04$.

$$z^{\star} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 2.05$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{38}{48} = 0.792$$

$$\hat{p}_2 = \frac{52}{87} = 0.598$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.598 - 0.792 = -0.194$$

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

$$\hat{p} = \frac{38 + 52}{48 + 87} = 0.667$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.667)(0.333)}{48} + \frac{(0.667)(0.333)}{87}}$$

$$= 0.0847$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0847)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.598 - 0.792) - 0}{0.0847} \\ &= -2.29 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.29) \\ &= 0.022 \end{aligned}$$

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) $H_0: \quad p_2 - p_1 = 0$

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

(c) $z^\star = 2.05$

(d) $SE = 0.0847$

(e) $|z_{\text{obs}}| = 2.29$

(f) p-value = 0.022

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 028

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 = 28.4**

   (b) **UB = 29.9**

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

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

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

   (d) $SE = 1.72$

   (e) $\mid t_{\textbf{obs}} \mid = 2.443$

   (f) $0.05 <$ p**-value** < 0.1

   (g) **retain**

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

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

   (c) $t^{\star} = 2.45$

   (d) $SE = 1.5$

   (e) $\mid t_{\textbf{obs}} \mid = 2.26$

   (f) $0.05 <$ p**-value** < 0.1

   (g) **retain**

4. (a) **LB of p CI = 0.703 or** 70.3%

   (b) **UB of p CI = 0.711 or** 71.1%

5. $n \approx 22000$

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

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

(c) $z^{\star} = 2.81$

(d) $SE = 0.07$

(e) $\mid z_{\mathbf{obs}} \mid = 3.13$

(f) p-**value** = 0.0018

(g) **reject**

1. **Problem**
   As an ornithologist, you wish to determine the average body mass of *Catharus guttatus*. You randomly capture 27 adults of *Catharus guttatus*, resulting in a sample mean of 29.12 grams and a sample standard deviation of 1.3 grams. You decide to report a 99.5% 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.

$$\begin{aligned} n &= 27 \\ \bar{x} &= 29.12 \\ s &= 1.3 \\ CL &= 0.995 \end{aligned}$$

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

$$df = n - 1 = 26$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.995$.

$$t^\star = 3.07$$

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.3}{\sqrt{27}} = 0.25$$

We want to make an inference about the population mean.

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

Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (29.12 - 3.07 \times 0.25,\ 29.12 - 3.07 \times 0.25) \\ &= (28.4,\ 29.9) \end{aligned}$$

We are 99.5% confident that the population mean is between 28.4 and 29.9.

(a) Lower bound = 28.4

(b) Upper bound = 29.9

2. **Problem**
A teacher has 6 students who have each taken two quizzes. Perform a two-tail test with significance level 0.05 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6
quiz 1:	77.1	76.9	67.2	53.4	56.3	71.3
quiz 2:	68.2	74	65.4	44.3	51.6	73.2

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

$$n = 6$$
$$\alpha = 0.05$$

State the hypotheses.

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

Determine the degrees of freedom.

$$df = n - 1 = 5$$

We determine $t^{\star}$ such that $P(|T| > t^{\star}) = 0.05$.

$$t^{\star} = 2.57$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6
quiz2-quiz1:	-8.9	-2.9	-1.8	-9.1	-4.7	1.9

Find the sample mean.

$$\overline{x_{\text{diff}}} = -4.25$$

Find the sample standard deviation.

$$s_{\text{diff}} = 4.27$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 1.74$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{-4.25 - 0}{1.74} = -2.443$$

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.

$$0.05 < p\text{-value} < 0.1$$

We conclude that we should retain the null hypothesis.

(a) $H_0: \ \mu_{\text{diff}} = 0$
(b) $H_A: \ \mu_{\text{diff}} \neq 0$
(c) $t^{\star} = 2.57$
(d) $SE = 1.7183285$
(e) $|t_{\text{obs}}| = 2.443$
(f) $0.05 < p\text{-value} < 0.1$
(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 6 cases and a control group with 6 cases. You decide to run a hypothesis test with a significance level of 0.05. Your data is below. Please use 6 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
29.5	20.5
29.3	22.2
25.7	28.8
27.7	26.8
26.3	22.5
29.3	26.7

(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.

$$n_1 = 6$$
$$n_2 = 6$$
$$\alpha = 0.05$$

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.

$$t^\star = 2.45$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 28$$
$$s_1 = 1.67$$
$$\bar{x}_2 = 24.6$$
$$s_2 = 3.28$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(1.67)^2}{6} + \frac{(3.28)^2}{6}} \\ &= 1.5 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(24.6 - 28) - (0)}{1.5} \\ &= -2.26 \end{aligned}$$

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.

$$0.05 < p\text{-value} < 0.1$$

Compare p-value and α.

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

We conclude that we should retain the null hypothesis.

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

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

(c) $t^{\star} = 2.45$

(d) $SE = 1.5$

(e) $|t_{\text{obs}}| = 2.26$

(f) $0.05 < p\text{-value} < 0.1$

(g) retain the null

4. **Problem**
   From a very large population, a random sample of 55000 individuals was taken. In that sample, 70.7% were shiny. 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.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.707)(1-0.707)}{55000}} = 0.00194$$

Calculate the margin of error.

$$ME = z^\star SE = (1.96)(0.00194) = 0.0038$$

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.

$$(0.703\,,\ 0.711)$$

We are 95% confident that the true population proportion is between 70.3% and 71.1%.

(a) The lower bound = 0.703, which can also be expressed as 70.3%.

(b) The upper bound = 0.711, which can also be expressed as 71.1%.

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 0.007 (which is 0.7 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.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.007}{2.05} = 0.00341$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.00341)^2} = 21499.6431059$$

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

$$n = 21500$$

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

$$n = 22000$$

6. **Problem**
   An experiment is run with a treatment group of size 82 and a control group of size 117. The results are summarized in the table below.

	treatment	control
sick	62	63
not sick	20	54

Using a significance level of 0.005, 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 $P(|Z| > z^\star) = 0.005$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 2.81$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{62}{82} = 0.756$$

$$\hat{p}_2 = \frac{63}{117} = 0.538$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.538 - 0.756 = -0.218$$

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

$$\hat{p} = \frac{62 + 63}{82 + 117} = 0.628$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.628)(0.372)}{82} + \frac{(0.628)(0.372)}{117}}$$

$$= 0.0696$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0696)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.538 - 0.756) - 0}{0.0696} \\ &= -3.13 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-3.13) \\ &= 0.0018 \end{aligned}$$

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) $H_0: \quad p_2 - p_1 = 0$

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

(c) $z^\star = 2.81$

(d) $SE = 0.0696$

(e) $|z_{\text{obs}}| = 3.13$

(f) p-value = 0.0018

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 029

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 = 14.2**

   (b) **UB = 15.3**

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

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

   (c) $t^\star = 2.61$

   (d) $SE = 1.12$

   (e) $\mid t_{\textbf{obs}} \mid = 2.649$

   (f) $0.02 < p$**-value** < 0.04

   (g) **reject**

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

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

   (c) $t^\star = 3.36$

   (d) $SE = 0.162$

   (e) $\mid t_{\textbf{obs}} \mid = 3.2$

   (f) $0.02 < p$**-value** < 0.04

   (g) **retain**

4. (a) **LB of p CI = 0.444 or** 44.4%

   (b) **UB of p CI = 0.586 or** 58.6%

5. $n \approx 1700$

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

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

(c) $z^{\star} = 2.81$

(d) $SE = 0.116$

(e) $|\ z_{\mathbf{obs}}\ | = 3.15$

(f) p-**value** = 0.0016

(g) **reject**

1. **Problem**
   As an ornithologist, you wish to determine the average body mass of *Dendroica castanea*. You randomly capture 33 adults of *Dendroica castanea*, resulting in a sample mean of 14.77 grams and a sample standard deviation of 1.26 grams. You decide to report a 98% 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.

$$\begin{aligned} n &= 33 \\ \bar{x} &= 14.77 \\ s &= 1.26 \\ CL &= 0.98 \end{aligned}$$

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

$$df = n - 1 = 32$$

   Determine the critical t value, $t^{\star}$, such that $P(|T| < t^{\star}) = 0.98$.

$$t^{\star} = 2.45$$

   Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.26}{\sqrt{33}} = 0.219$$

   We want to make an inference about the population mean.

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

   Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^{\star} SE,\ \bar{x} + t^{\star} SE) \\ &= (14.77 - 2.45 \times 0.219,\ 14.77 - 2.45 \times 0.219) \\ &= (14.2,\ 15.3) \end{aligned}$$

   We are 98% confident that the population mean is between 14.2 and 15.3.

   (a) Lower bound = 14.2

   (b) Upper bound = 15.3

2. **Problem**
A teacher has 7 students who have each taken two quizzes. Perform a two-tail test with significance level 0.04 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6	student7
quiz 1:	72.8	89.7	60.8	54.9	84.1	71.8	61.4
quiz 2:	75.6	97.1	59.3	58.6	89.1	74.8	61.6

(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.

$$n = 7$$
$$\alpha = 0.04$$

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 $P(|T| > t^{\star}) = 0.04$.

$$t^{\star} = 2.61$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6	student7
quiz2-quiz1:	2.8	7.4	-1.5	3.7	5	3	0.2

Find the sample mean.

$$\overline{x_{\text{diff}}} = 2.94$$

Find the sample standard deviation.

$$s_{\text{diff}} = 2.94$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 1.11$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{2.94 - 0}{1.11} = 2.649$$

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 reject the null hypothesis.

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

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

(c) $t^\star = 2.61$

(d) $SE = 1.1150041$

(e) $|t_{\text{obs}}| = 2.649$

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

(g) reject 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 0.02. Your data is below. Please use 5 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
2.84	3.24
2.55	3.34
2.14	2.99
2.47	2.77
2.82	3.08

   (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.

$$n_1 = 5$$
$$n_2 = 5$$
$$\alpha = 0.02$$

   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.

$$t^\star = 3.36$$

   Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 2.56$$
$$s_1 = 0.287$$
$$\bar{x}_2 = 3.08$$
$$s_2 = 0.222$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(0.287)^2}{5} + \frac{(0.222)^2}{5}} \\ &= 0.162 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(3.08 - 2.56) - (0)}{0.162} \\ &= 3.2 \end{aligned}$$

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.

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

Compare p-value and α.

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

We conclude that we should retain the null hypothesis.

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

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

(c) $t^\star = 3.36$

(d) $SE = 0.162$

(e) $|t_{\text{obs}}| = 3.2$

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

(g) retain the null

4. **Problem**
   From a very large population, a random sample of 390 individuals was taken. In that sample, 51.5% were floating. Determine a 99.5% 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 $P(|Z| < z^{\star}) = 0.995$.

$$z^{\star} = 2.81$$

   Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.515)(1-0.515)}{390}} = 0.0253$$

   Calculate the margin of error.

$$ME = z^{\star} SE = (2.81)(0.0253) = 0.0711$$

   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.

$$(0.444\,,\ 0.586)$$

   We are 99.5% confident that the true population proportion is between 44.4% and 58.6%.

   (a) The lower bound = 0.444, which can also be expressed as 44.4%.
   (b) The upper bound = 0.586, which can also be expressed as 58.6%.

5. **Problem**

   Your boss wants to know what proportion of a very large population is messy. She also wants to guarantee that the margin of error of a 90% confidence interval will be less than 0.02 (which is 2 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

   Solution

   Determine $z^\star$ such that $P(|Z| < z^\star) = 0.9$.

$$z^\star = 1.64$$

   Determine the maximal standard error.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.02}{1.64} = 0.0122$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.0122)^2} = 1679.6560064$$

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

$$n = 1680$$

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

$$n = 1700$$

6. **Problem**
An experiment is run with a treatment group of size 40 and a control group of size 34. The results are summarized in the table below.

	treatment	control
preoccupied	24	8
not preoccupied	16	26

Using a significance level of 0.005, determine whether the treatment causes an effect on the proportion of cases that are preoccupied.

(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 $P(|Z| > z^\star) = 0.005$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 2.81$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{24}{40} = 0.6$$

$$\hat{p}_2 = \frac{8}{34} = 0.235$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.235 - 0.6 = -0.365$$

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

$$\hat{p} = \frac{24 + 8}{40 + 34} = 0.432$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.432)(0.568)}{40} + \frac{(0.432)(0.568)}{34}}$$

$$= 0.116$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0, 0.116)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.235 - 0.6) - 0}{0.116} \\ &= -3.15 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-3.15) \\ &= 0.0016 \end{aligned}$$

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) $H_0: \quad p_2 - p_1 = 0$

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

(c) $z^\star = 2.81$

(d) $SE = 0.116$

(e) $|z_{\text{obs}}| = 3.15$

(f) p-value = 0.0016

(g) reject the null

Bunker Hill Community College

Third Statistics Exam 2019-04-25 Exam ID 030

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 = 14.8**

   (b) **UB = 15.6**

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

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

   (c) $t^{\star} = 2.45$

   (d) $SE = 1.78$

   (e) $\mid t_{\textbf{obs}} \mid = 2.527$

   (f) $0.04 < p$**-value** < 0.05

   (g) **reject**

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

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

   (c) $t^{\star} = 1.94$

   (d) $SE = 0.41$

   (e) $\mid t_{\textbf{obs}} \mid = 2$

   (f) $0.05 < p$**-value** < 0.1

   (g) **reject**

4. (a) **LB of p CI = 0.128 or** 12.8%

   (b) **UB of p CI = 0.156 or** 15.6%

5. $n \approx 2500$

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

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

(c) $z^{\star} = 1.64$

(d) $SE = 0.089$

(e) $\mid z_{\mathbf{obs}} \mid = 1.89$

(f) p-**value** = 0.0588

(g) **reject**

1. **Problem**
   As an ornithologist, you wish to determine the average body mass of *Dendroica castanea*. You randomly capture 31 adults of *Dendroica castanea*, resulting in a sample mean of 15.24 grams and a sample standard deviation of 1.11 grams. You decide to report a 95% 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.

$$\begin{aligned} n &= 31 \\ \bar{x} &= 15.24 \\ s &= 1.11 \\ CL &= 0.95 \end{aligned}$$

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

$$df = n - 1 = 30$$

   Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.95$.

$$t^\star = 2.04$$

   Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.11}{\sqrt{31}} = 0.199$$

   We want to make an inference about the population mean.

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

   Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (15.24 - 2.04 \times 0.199,\ 15.24 - 2.04 \times 0.199) \\ &= (14.8,\ 15.6) \end{aligned}$$

   We are 95% confident that the population mean is between 14.8 and 15.6.

   (a) Lower bound = 14.8

   (b) Upper bound = 15.6

2. **Problem**
A teacher has 7 students who have each taken two quizzes. Perform a two-tail test with significance level 0.05 to determine whether students' performance changed on average.

	student1	student2	student3	student4	student5	student6	student7
quiz 1:	89.4	55.1	81.6	87.4	69.5	86	76.5
quiz 2:	88	49.7	72.5	80.1	72.7	84	65.6

(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.

$$n = 7$$
$$\alpha = 0.05$$

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 $P(|T| > t^{\star}) = 0.05$.

$$t^{\star} = 2.45$$

Subtract each student's scores to get the differences.

	student1	student2	student3	student4	student5	student6	student7
quiz2-quiz1:	-1.4	-5.4	-9.1	-7.3	3.2	-2	-10.9

Find the sample mean.

$$\overline{x_{\text{diff}}} = -4.7$$

Find the sample standard deviation.

$$s_{\text{diff}} = 4.93$$

Determine the standard error.

$$SE = \frac{s_{\text{diff}}}{\sqrt{n}} = 1.86$$

Calculate the observed t score.

$$t_{\text{obs}} = \frac{\overline{x_{\text{diff}}} - (\mu_{\text{diff}})_0}{SE} = \frac{-4.7 - 0}{1.86} = -2.527$$

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.

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

We conclude that we should reject the null hypothesis.

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

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

(c) $t^{\star} = 2.45$

(d) $SE = 1.7762917$

(e) $|t_{\text{obs}}| = 2.527$

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

(g) reject 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 6 cases and a control group with 6 cases. You decide to run a hypothesis test with a significance level of 0.1. Your data is below. Please use 6 for the degrees of freedom (calculated with the Welch-Satterthwaite equation).

treatment	control
3.44	4.07
3.77	1.78
3.11	2.51
3.31	2.54
3.28	1.43
2.53	2.16

(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.

$$n_1 = 6$$
$$n_2 = 6$$
$$\alpha = 0.1$$

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.

$$t^\star = 1.94$$

Determine the sample statistics. Use a calculator!

$$\bar{x}_1 = 3.24$$
$$s_1 = 0.412$$
$$\bar{x}_2 = 2.42$$
$$s_2 = 0.917$$

Calculate the standard error.

$$
\begin{aligned}
SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\
&= \sqrt{\frac{(0.412)^2}{6} + \frac{(0.917)^2}{6}} \\
&= 0.41
\end{aligned}
$$

Determine the test statistic.

$$
\begin{aligned}
t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\
&= \frac{(2.42 - 3.24) - (0)}{0.41} \\
&= -2
\end{aligned}
$$

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.

$$0.05 < p\text{-value} < 0.1$$

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) $t^\star = 1.94$

(d) $SE = 0.41$

(e) $|t_{\text{obs}}| = 2$

(f) $0.05 < p\text{-value} < 0.1$

(g) reject the null

4. **Problem**
From a very large population, a random sample of 2400 individuals was taken. In that sample, 14.2% were tasty. 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.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.142)(1-0.142)}{2400}} = 0.00712$$

Calculate the margin of error.

$$ME = z^\star SE = (1.96)(0.00712) = 0.014$$

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.

$$(0.128\,,\ 0.156)$$

We are 95% confident that the true population proportion is between 12.8% and 15.6%.

   (a) The lower bound = 0.128, which can also be expressed as 12.8%.
   (b) The upper bound = 0.156, which can also be expressed as 15.6%.

5. **Problem**

   Your boss wants to know what proportion of a very large population is bitter. She also wants to guarantee that the margin of error of a 95% confidence interval will be less than 0.02 (which is 2 percentage points). How large of a sample is needed? Please round up, using only 2 significant digits.

   Solution

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

   $$z^\star = 1.96$$

   Determine the maximal standard error.

   $$ME = z^\star SE$$

   $$SE = \frac{ME}{z^\star} = \frac{0.02}{1.96} = 0.0102$$

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

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

   $$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.0102)^2} = 2402.9219531$$

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

   $$n = 2403$$

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

   $$n = 2500$$

6. **Problem**
An experiment is run with a treatment group of size 42 and a control group of size 77. The results are summarized in the table below.

	treatment	control
omnivorous	18	20
not omnivorous	24	57

Using a significance level of 0.1, determine whether the treatment causes an effect on the proportion of cases that are omnivorous.

(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 $P(|Z| > z^\star) = 0.1$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 1.64$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{18}{42} = 0.429$$

$$\hat{p}_2 = \frac{20}{77} = 0.26$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.26 - 0.429 = -0.169$$

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

$$\hat{p} = \frac{18 + 20}{42 + 77} = 0.319$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.319)(0.681)}{42} + \frac{(0.319)(0.681)}{77}}$$

$$= 0.0894$$

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

$$H_0 : \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0894)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.26 - 0.429) - 0}{0.0894} \\ &= -1.89 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-1.89) \\ &= 0.0588 \end{aligned}$$

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) $H_0 : \quad p_2 - p_1 = 0$

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

(c) $z^{\star} = 1.64$

(d) $SE = 0.0894$

(e) $|z_{\text{obs}}| = 1.89$

(f) p-value = 0.0588

(g) reject the null