Bunker Hill Community College

Final Statistics Exam 2019-05-02 Exam ID 022

Name: ______________________________

This take-home exam is due **Wednesday, May 8**, at the beginning of class.

You may use any notes, textbook, or online tools; however, you may not request help from any other human.

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) P(**cat and yellow**) = 0.0633

   (b) P(**flower**) = 0.396

   (c) P(**tree or yellow**) = 0.359

   (d) P(**yellow given cat**) = 0.177

   (e) P(**flower given yellow**) = 0.273

   (f) P(**violet**) = 0.324

2. P(**"not tree" given "white"**) = 0.787

3. $P(63.62 < X < 64.06) = 0.8622$

4. (a) $P(X = 19) = 0.1167$

   (b) $P(13 \leq X \leq 23) = 0.9463$

5. $\mathbf{(22.4, 24.8)}$

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

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

   (c) $t^{\star} = 1.69$

   (d) $SE = 2.082$

   (e) $\mid t_{\mathbf{obs}} \mid = 1.44$

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

   (g) **retain**

7. (a) **LB of p CI = 0.267 or** 26.7%

   (b) **UB of p CI = 0.275 or** 27.5%

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

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

(c) $z^{\star} = 1.96$

(d) $SE = 0.027$

(e) $\mid z_{\mathbf{obs}} \mid = 2.05$

(f) p-**value** = 0.0404

(g) **reject**

1. In a deck of strange cards, there are 490 cards. Each card has an image and a color. The amounts are shown in the table below.

	orange	violet	yellow
cat	82	62	31
flower	83	87	24
tree	78	10	33

(a) What is the probability a random card is both a cat and yellow?

(b) What is the probability a random card is a flower?

(c) What is the probability a random card is either a tree or yellow (or both)?

(d) What is the probability a random card is yellow given it is a cat?

(e) What is the probability a random card is a flower given it is yellow?

(f) What is the probability a random card is violet?

Solution

(a) $P(\text{cat and yellow}) = \frac{31}{490} = 0.0633$

(b) $P(\text{flower}) = \frac{83+87+24}{490} = 0.396$

(c) $P(\text{tree or yellow}) = \frac{78+10+33+31+24+33-33}{490} = 0.359$

(d) $P(\text{yellow given cat}) = \frac{31}{82+62+31} = 0.177$

(e) $P(\text{flower given yellow}) = \frac{24}{31+24+33} = 0.273$

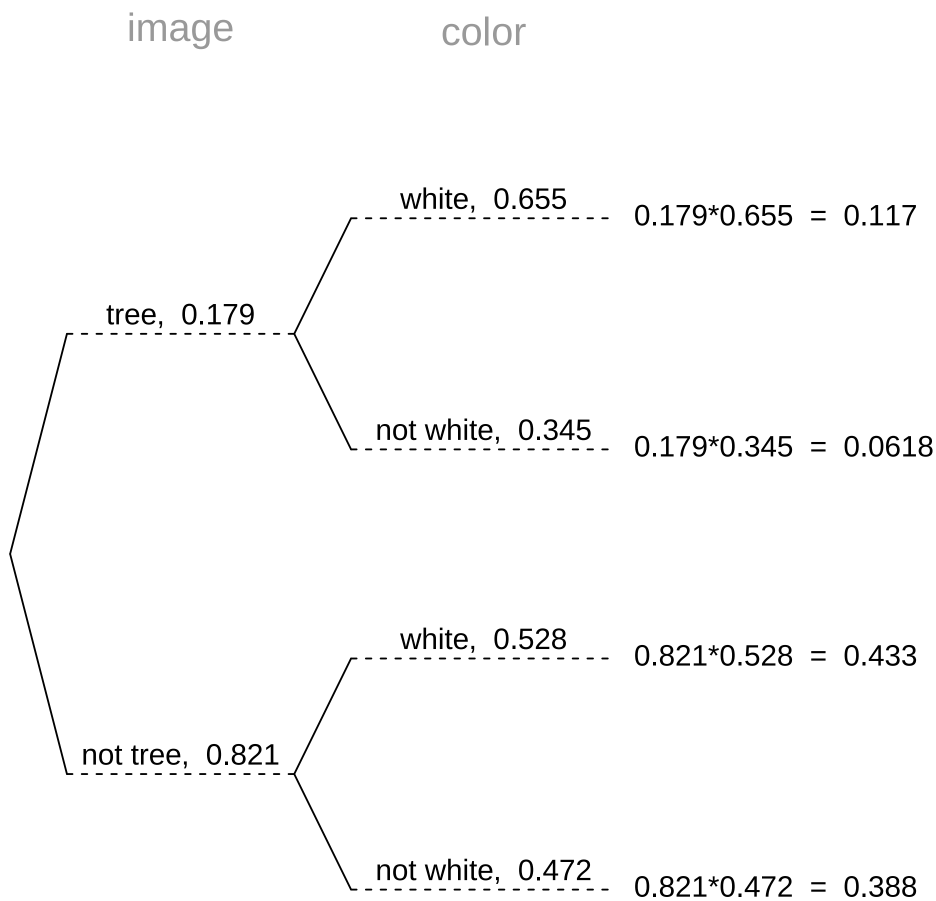
(f) $P(\text{violet}) = \frac{62+87+10}{490} = 0.324$

2. In a deck of strange cards, each card has an image and a color. The chance of drawing a tree is 17.9%. If a tree is drawn, there is a 65.5% chance that it is white. If a card that is not a tree is drawn, there is a 52.8% chance that it is white.

   Now, someone draws a random card and reveals it is white. What is the chance the card is not a tree?

Solution

I'd recommend making a tree. Remember, on the first branch, we put simple probabilities. On the second branches we put conditional probabilites. The results (products) are joint probabilities.



Determine the appropriate conditional probability.

$$P(\text{"not tree" given "white"}) = \frac{0.433}{0.433 + 0.117} = 0.787$$

3. In a very large pile of toothpicks, the mean length is 63.81 millimeters and the standard deviation is 1.77 millimeters. If you randomly sample 150 toothpicks, what is the chance the sample mean is between 63.62 and 64.06 millimeters?

Solution

Label the given information.

$$\mu = 63.81$$
$$\sigma = 1.77$$
$$n = 150$$
$$\bar{x}_{\text{lower}} = 63.62$$
$$\bar{x}_{\text{upper}} = 64.06$$

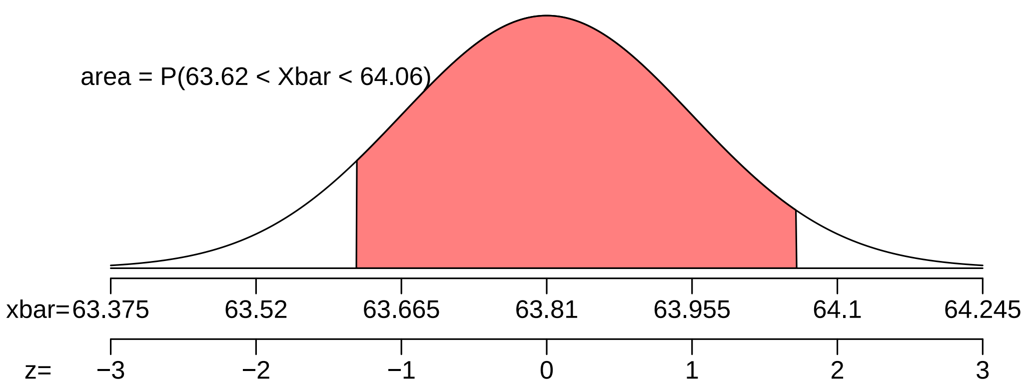
Find the standard error.

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{1.77}{\sqrt{150}} = 0.145$$

Describe the sampling distribution.

$$\bar{X} \sim \mathcal{N}(63.81,\ 0.145)$$

Draw a sketch.



Calculate a z scores.

$$z_{\text{lower}} = \frac{x_{\text{lower}} - \mu}{SE} = \frac{63.62 - 63.81}{0.145} = -1.31$$

$$z_{\text{upper}} = \frac{x_{\text{upper}} - \mu}{SE} = \frac{64.06 - 63.81}{0.145} = 1.72$$

Determine the probability.

$$\begin{aligned} P(63.62 < X < 64.06) &= \Phi(z_{\text{upper}}) - \Phi(z_{\text{lower}}) \\ &= \Phi(1.72) - \Phi(-1.31) \\ &= 0.8622 \end{aligned}$$

4. In a game, there is a 57% chance to win a round. You will play 30 rounds.
   (a) What is the probability of winning exactly 19 rounds?
   (b) What is the probability of winning at least 13 but at most 23 rounds?

Solution
We use the formula for binomial probabilities.

$$P(X = k) = \binom{n}{k} (p)^k (1 - p)^{n-k}$$

$$P(X = 19) = \binom{30}{19} (0.57)^{19} (1 - 0.57)^{30-19}$$

$$P(X = 19) = \binom{30}{19} (0.57)^{19} (0.43)^{11}$$

$$P(X = 19) = 0.1167$$

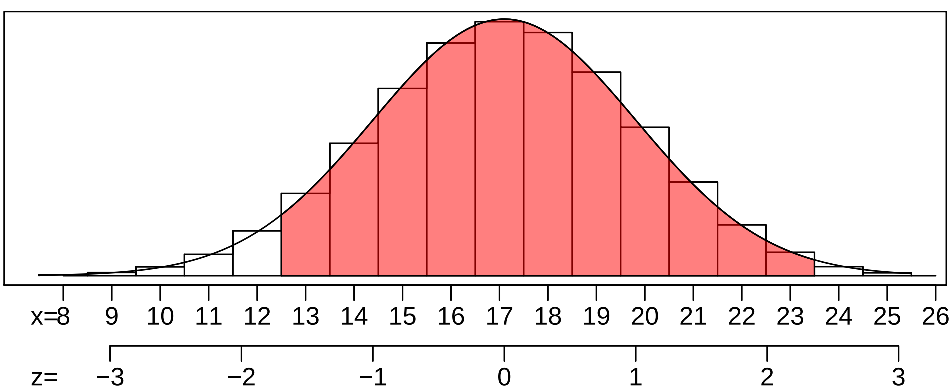
Find the mean.

$$\mu = np = (30)(0.57) = 17.1$$

Find the standard deviation.

$$\sigma = \sqrt{np(1 - p)} = \sqrt{(30)(0.57)(1 - 0.57)} = 2.7116$$

Make a sketch, specifically try to picture whether you need to add or subtract 0.5 for the continuity correction.



Find the z scores.

$$z_1 = \frac{12.5 - 17.1}{2.7116} = -1.7$$

$$z_2 = \frac{23.5 - 17.1}{2.7116} = 2.36$$

Calculate the probability.

$$P(13 \leq X \leq 23) = \Phi(2.36) - \Phi(-1.7) = 0.9463$$

(a) $P(X = 19) = 0.1167$

(b) $P(13 \leq X \leq 23) = 0.9463$

5. As an ornithologist, you wish to determine the average body mass of *Zonotrichia albicollis*. You randomly sample 14 adults of *Zonotrichia albicollis*, resulting in a sample mean of 23.61 grams and a sample standard deviation of 1.67 grams. Determine a 98% confidence interval of the true population mean.

Solution

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

$$n = 14$$
$$\bar{x} = 23.61$$
$$s = 1.67$$
$$CL = 0.98$$

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 $P(|T| < t^\star) = 0.98$.

$$t^\star = 2.65$$

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.67}{\sqrt{14}} = 0.446$$

We want to make an inference about the population mean.

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

Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (23.61 - 2.65 \times 0.446,\ 23.61 + 2.65 \times 0.446) \\ &= (22.4,\ 24.8) \end{aligned}$$

We are 98% confident that the population mean is between 22.4 and 24.8.

6. A treatment group of size 20 has a mean of 103 and standard deviation of 7.85. A control group of size 32 has a mean of 100 and standard deviation of 6.33. If you decided to use a signficance level of 0.1, is there sufficient evidence to conclude the treatment causes an effect?

   By using the Welch-Satterthwaite equation, I've calculated the degrees of freedom should be 34.

   (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.

$$n_1 = 20$$
$$\bar{x}_1 = 103$$
$$s_1 = 7.85$$
$$n_2 = 32$$
$$\bar{x}_2 = 100$$
$$s_2 = 6.33$$
$$\alpha = 0.1$$
$$df = 34$$

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 $P(|T| > t^\star) = 0.1$ by using a t table.

$$t^\star = 1.69$$

Calculate the standard error.

$$SE = \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} = \sqrt{\frac{(7.85)^2}{20} + \frac{(6.33)^2}{32}} = 2.082$$

Determine the test statistic.

$$t_{\text{obs}} = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} = \frac{(100 - 103) - (0)}{2.082} = -1.44$$

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) $t^\star = 1.69$

(d) $SE = 2.082$

(e) $|t_{obs}| = 1.44$

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

(g) retain the null

7. From a very large population, a random sample of 53000 individuals was taken. In that sample, 27.1% were blue. 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.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.271)(1-0.271)}{53000}} = 0.00193$$

Calculate the margin of error.

$$ME = z^{\star} SE = (2.05)(0.00193) = 0.00396$$

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.

$$(0.267\,,\ 0.275)$$

We are 96% confident that the true population proportion is between 26.7% and 27.5%.

(a) The lower bound = 0.267, which can also be expressed as 26.7%.

(b) The upper bound = 0.275, which can also be expressed as 27.5%.

8. An experiment is run with a treatment group of size 257 and a control group of size 293. The results are summarized in the table below.

	treatment	control
reclusive	221	268
not reclusive	36	25

Using a significance level of 0.05, 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 $P(|Z| > z^\star) = 0.05$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 1.96$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{221}{257} = 0.86$$
$$\hat{p}_2 = \frac{268}{293} = 0.915$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.915 - 0.86 = 0.055$$

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

$$\hat{p} = \frac{221 + 268}{257 + 293} = 0.889$$

Determine the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}} \\ &= \sqrt{\frac{(0.889)(0.111)}{257} + \frac{(0.889)(0.111)}{293}} \\ &= 0.0268 \end{aligned}$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\, 0.0268)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.915 - 0.86) - 0}{0.0268} \\ &= 2.05 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.05) \\ &= 0.0404 \end{aligned}$$

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) $z^\star = 1.96$

(d) $SE = 0.0268$

(e) $|z_{\text{obs}}| = 2.05$

(f) p-value = 0.0404

(g) reject the null