Bunker Hill Community College

Final Statistics Exam 2019-05-02 **Exam ID 018**

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(**tree given white**) = 0.129

   (b) P(**white**) = 0.26

   (c) P(**white given tree**) = 0.157

   (d) P(**shovel or teal**) = 0.487

   (e) P(**tree**) = 0.213

   (f) P(**tree and white**) = 0.0334

2. P(**"not cat" given "teal"**) = 0.73

3. $P(74 < X < 74.7) = 0.7567$

4. (a) $P(X = 28) = 0.0905$

   (b) $P(21 \leq X \leq 32) = 0.7656$

5. **(11.7, 12.3)**

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

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

   (c) $t^\star = 2.62$

   (d) $SE = 8.061$

   (e) $\mid t_{\mathbf{obs}} \mid = 2.79$

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

   (g) **reject**

7. (a) **LB of p CI = 0.917 or** 91.7%

   (b) **UB of p CI = 0.921 or** 92.1%

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

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

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

(d) $SE = 0.097$

(e) $\mid z_{\mathbf{obs}} \mid = 2.46$

(f) p**-value** = 0.0138

(g) **reject**

1. In a deck of strange cards, there are 809 cards. Each card has an image and a color. The amounts are shown in the table below.

	green	teal	violet	white
bike	32	48	62	29
cat	76	59	58	83
shovel	15	18	86	71
tree	28	97	20	27

(a) What is the probability a random card is a tree given it is white?

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

(c) What is the probability a random card is white given it is a tree?

(d) What is the probability a random card is either a shovel or teal (or both)?

(e) What is the probability a random card is a tree?

(f) What is the probability a random card is both a tree and white?

Solution

(a) $P(\text{tree given white}) = \frac{27}{29+83+71+27} = 0.129$

(b) $P(\text{white}) = \frac{29+83+71+27}{809} = 0.26$

(c) $P(\text{white given tree}) = \frac{27}{28+97+20+27} = 0.157$

(d) $P(\text{shovel or teal}) = \frac{15+18+86+71+48+59+18+97-18}{809} = 0.487$

(e) $P(\text{tree}) = \frac{28+97+20+27}{809} = 0.213$

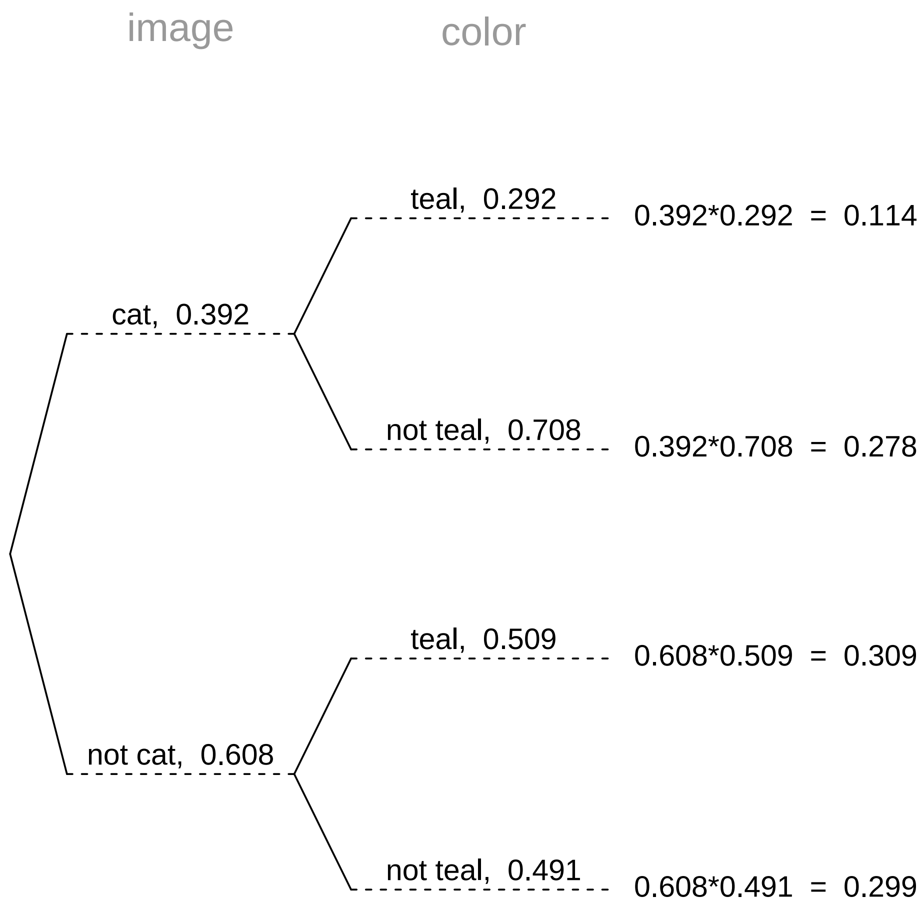
(f) $P(\text{tree and white}) = \frac{27}{809} = 0.0334$

2. In a deck of strange cards, each card has an image and a color. The chance of drawing a cat is 39.2%. If a cat is drawn, there is a 29.2% chance that it is teal. If a card that is not a cat is drawn, there is a 50.9% chance that it is teal.

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

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.

$$P(\text{"not cat" given "teal"}) = \frac{0.309}{0.309 + 0.114} = 0.73$$

3. In a very large pile of toothpicks, the mean length is 74.2 millimeters and the standard deviation is 3.33 millimeters. If you randomly sample 169 toothpicks, what is the chance the sample mean is between 74 and 74.7 millimeters?

Solution

Label the given information.

$$\mu = 74.2$$
$$\sigma = 3.33$$
$$n = 169$$
$$\bar{x}_{\text{lower}} = 74$$
$$\bar{x}_{\text{upper}} = 74.7$$

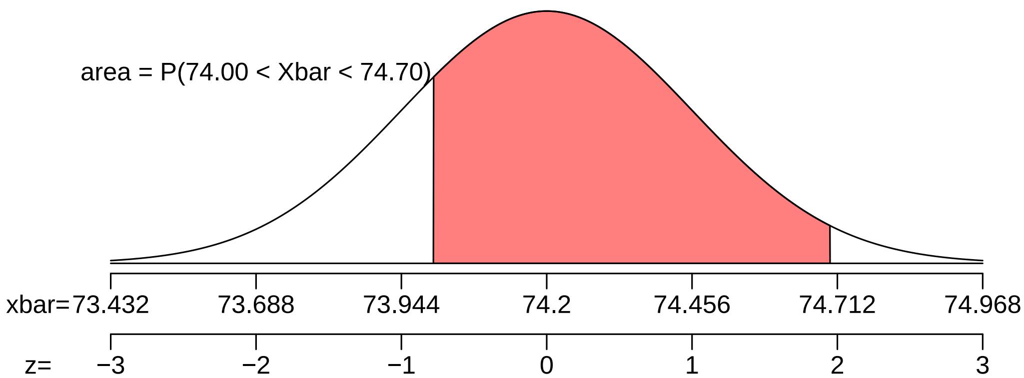
Find the standard error.

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{3.33}{\sqrt{169}} = 0.256$$

Describe the sampling distribution.

$$\bar{X} \sim \mathcal{N}(74.2,\ 0.256)$$

Draw a sketch.



Calculate a z scores.

$$z_{\text{lower}} = \frac{x_{\text{lower}} - \mu}{SE} = \frac{74 - 74.2}{0.256} = -0.78$$

$$z_{\text{upper}} = \frac{x_{\text{upper}} - \mu}{SE} = \frac{74.7 - 74.2}{0.256} = 1.95$$

Determine the probability.

$$\begin{aligned} P(74 < X < 74.7) &= \Phi(z_{\text{upper}}) - \Phi(z_{\text{lower}}) \\ &= \Phi(1.95) - \Phi(-0.78) \\ &= 0.7567 \end{aligned}$$

4. In a game, there is a 31% chance to win a round. You will play 90 rounds.
   (a) What is the probability of winning exactly 28 rounds?
   (b) What is the probability of winning at least 21 but at most 32 rounds?

Solution

We use the formula for binomial probabilities.

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

$$P(X = 28) = \binom{90}{28}(0.31)^{28}(1-0.31)^{90-28}$$

$$P(X = 28) = \binom{90}{28}(0.31)^{28}(0.69)^{62}$$

$$P(X = 28) = 0.0905$$

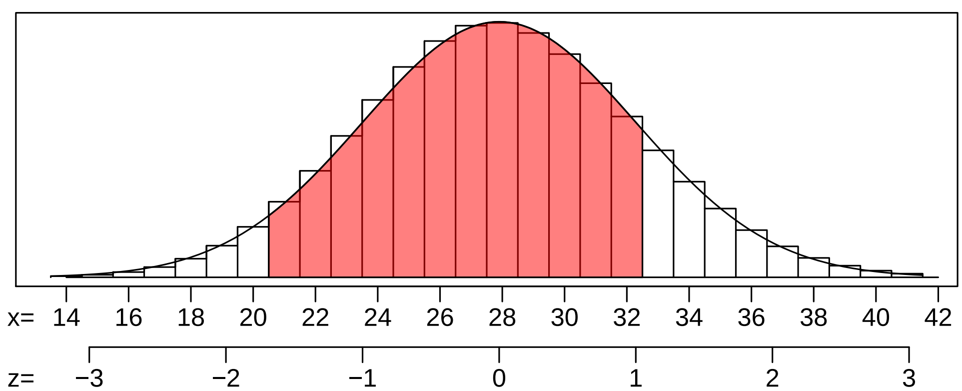
Find the mean.

$$\mu = np = (90)(0.31) = 27.9$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(90)(0.31)(1-0.31)} = 4.3876$$

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



Find the z scores.

$$z_1 = \frac{20.5 - 27.9}{4.3876} = -1.57$$

$$z_2 = \frac{32.5 - 27.9}{4.3876} = 0.93$$

Calculate the probability.

$$P(21 \leq X \leq 32) = \Phi(0.93) - \Phi(-1.57) = 0.7656$$

(a) $P(X = 28) = 0.0905$

(b) $P(21 \leq X \leq 32) = 0.7656$

5. As an ornithologist, you wish to determine the average body mass of *Dendroica coronata*. You randomly sample 34 adults of *Dendroica coronata*, resulting in a sample mean of 11.97 grams and a sample standard deviation of 1.28 grams. Determine a 80% confidence interval of the true population mean.

Solution

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

$$\begin{aligned} n &= 34 \\ \bar{x} &= 11.97 \\ s &= 1.28 \\ CL &= 0.8 \end{aligned}$$

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

$$df = n - 1 = 33$$

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

$$t^{\star} = 1.31$$

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.28}{\sqrt{34}} = 0.22$$

We want to make an inference about the population mean.

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

Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^{\star} SE,\ \bar{x} + t^{\star} SE) \\ &= (11.97 - 1.31 \times 0.22,\ 11.97 + 1.31 \times 0.22) \\ &= (11.7,\ 12.3) \end{aligned}$$

We are 80% confident that the population mean is between 11.7 and 12.3.

6. A treatment group of size 10 has a mean of 93.5 and standard deviation of 22.3. A control group of size 21 has a mean of 116 and standard deviation of 17.9. If you decided to use a signficance level of 0.02, 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 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.

$$n_1 = 10$$
$$\bar{x}_1 = 93.5$$
$$s_1 = 22.3$$
$$n_2 = 21$$
$$\bar{x}_2 = 116$$
$$s_2 = 17.9$$
$$\alpha = 0.02$$
$$df = 14$$

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 $P(|T| > t^\star) = 0.02$ by using a t table.

$$t^\star = 2.62$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(22.3)^2}{10} + \frac{(17.9)^2}{21}} \\ &= 8.061 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(116 - 93.5) - (0)}{8.061} \\ &= 2.79 \end{aligned}$$

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 reject the null hypothesis.

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

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

(c) $t^\star = 2.62$

(d) $SE = 8.061$

(e) $|t_{obs}| = 2.79$

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

(g) reject the null

7. From a very large population, a random sample of 90000 individuals was taken. In that sample, 91.9% 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.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.919)(1-0.919)}{90000}} = 0.000909$$

Calculate the margin of error.

$$ME = z^{\star} SE = (1.96)(0.000909) = 0.00178$$

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.

$$(0.917\,,\;0.921)$$

We are 95% confident that the true population proportion is between 91.7% and 92.1%.

(a) The lower bound = 0.917, which can also be expressed as 91.7%.

(b) The upper bound = 0.921, which can also be expressed as 92.1%.

8. An experiment is run with a treatment group of size 62 and a control group of size 46. The results are summarized in the table below.

	treatment	control
happy	27	31
not happy	35	15

Using a significance level of 0.02, 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 $P(|Z| > z^\star) = 0.02$.

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

Determine the sample proportions.

$$\hat{p}_1 = \frac{27}{62} = 0.435$$

$$\hat{p}_2 = \frac{31}{46} = 0.674$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.674 - 0.435 = 0.239$$

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

$$\hat{p} = \frac{27 + 31}{62 + 46} = 0.537$$

Determine the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}} \\ &= \sqrt{\frac{(0.537)(0.463)}{62} + \frac{(0.537)(0.463)}{46}} \\ &= 0.097 \end{aligned}$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0, 0.097)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.674 - 0.435) - 0}{0.097} \\ &= 2.46 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.46) \\ &= 0.0138 \end{aligned}$$

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) $SE = 0.097$

(e) $|z_{\text{obs}}| = 2.46$

(f) p-value = 0.0138

(g) reject the null