Bunker Hill Community College

Final Statistics Exam 2019-05-02 Exam ID 011

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(**gray given bike**) = 0.145

   (b) P(**shovel given red**) = 0.388

   (c) P(**bike**) = 0.316

   (d) P(**cat and red**) = 0.0218

   (e) P(**white**) = 0.14

   (f) P(**bike or red**) = 0.413

2. P(**"shovel" given "not yellow"**) = 0.355

3. $P(61.3 < X < 61.93) = 0.7783$

4. (a) $P(X = 140) = 0.0441$

   (b) $P(127 \leq X \leq 149) = 0.884$

5. $\mathbf{(14.6, 16.9)}$

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

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

   (c) $t^\star = 2.39$

   (d) $SE = 0.052$

   (e) $\mid t_{\mathbf{obs}} \mid = 2.34$

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

   (g) **retain**

7. (a) **LB of p CI = 0.278 or** 27.8%

   (b) **UB of p CI = 0.304 or** 30.4%

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) $SE = 0.041$

(e) $\mid z_{\mathbf{obs}} \mid = 2.1$

(f) p-**value** $= 0.0358$

(g) **reject**

1. In a deck of strange cards, there are 917 cards. Each card has an image and a color. The amounts are shown in the table below.

	gray	green	pink	red	white
bike	42	94	24	89	41
cat	97	64	80	20	68
shovel	43	77	90	69	19

   (a) What is the probability a random card is gray given it is a bike?

   (b) What is the probability a random card is a shovel given it is red?

   (c) What is the probability a random card is a bike?

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

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

   (f) What is the probability a random card is either a bike or red (or both)?

Solution

(a) $P(\text{gray given bike}) = \frac{42}{42+94+24+89+41} = 0.145$

(b) $P(\text{shovel given red}) = \frac{69}{89+20+69} = 0.388$

(c) $P(\text{bike}) = \frac{42+94+24+89+41}{917} = 0.316$

(d) $P(\text{cat and red}) = \frac{20}{917} = 0.0218$

(e) $P(\text{white}) = \frac{41+68+19}{917} = 0.14$

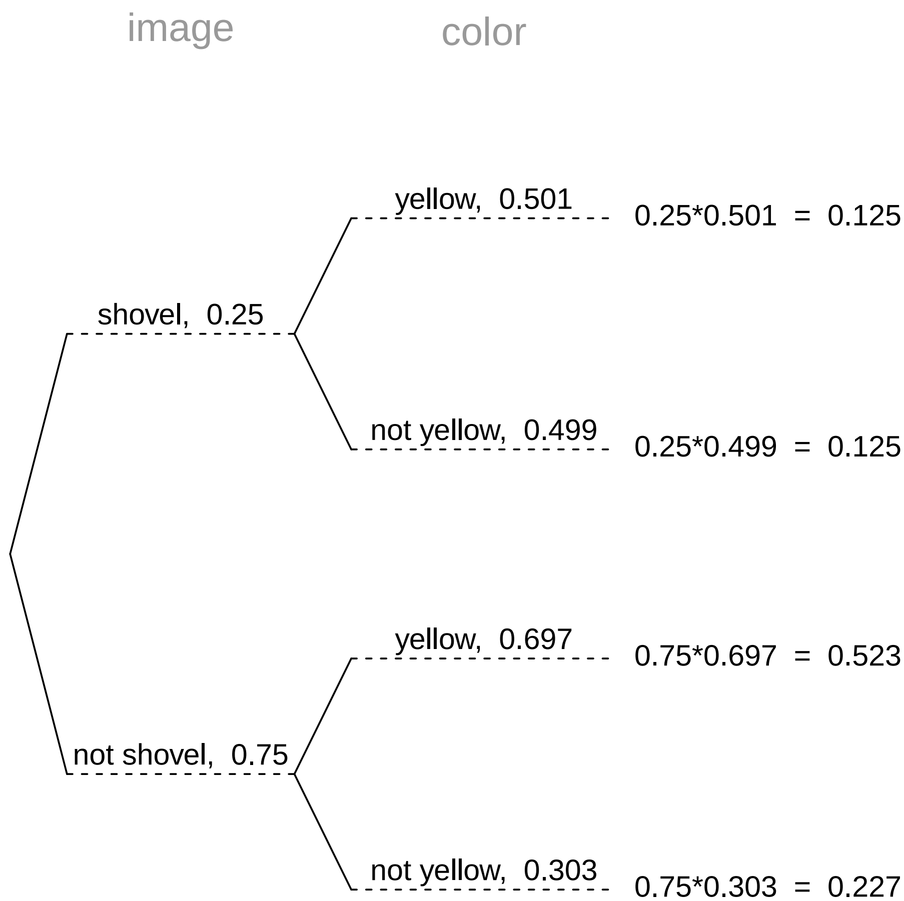
(f) $P(\text{bike or red}) = \frac{42+94+24+89+41+89+20+69-89}{917} = 0.413$

2. In a deck of strange cards, each card has an image and a color. The chance of drawing a shovel is 25%. If a shovel is drawn, there is a 50.1% chance that it is yellow. If a card that is not a shovel is drawn, there is a 69.7% chance that it is yellow.

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

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.

$$P(\text{"shovel" given "not yellow"}) = \frac{0.125}{0.125 + 0.227} = 0.355$$

3. In a very large pile of toothpicks, the mean length is 61.68 millimeters and the standard deviation is 3.52 millimeters. If you randomly sample 200 toothpicks, what is the chance the sample mean is between 61.3 and 61.93 millimeters?

Solution

Label the given information.

$$\mu = 61.68$$
$$\sigma = 3.52$$
$$n = 200$$
$$\bar{x}_{\text{lower}} = 61.3$$
$$\bar{x}_{\text{upper}} = 61.93$$

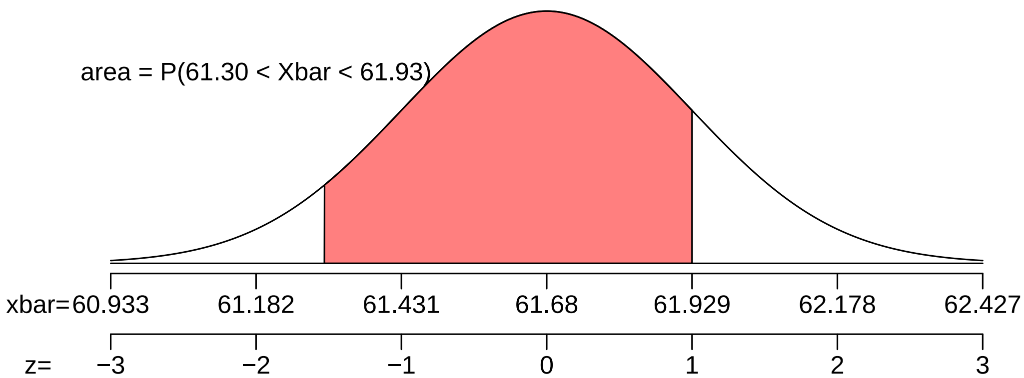
Find the standard error.

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{3.52}{\sqrt{200}} = 0.249$$

Describe the sampling distribution.

$$\bar{X} \sim \mathcal{N}(61.68,\ 0.249)$$

Draw a sketch.



Calculate a z scores.

$$z_{\text{lower}} = \frac{x_{\text{lower}} - \mu}{SE} = \frac{61.3 - 61.68}{0.249} = -1.53$$

$$z_{\text{upper}} = \frac{x_{\text{upper}} - \mu}{SE} = \frac{61.93 - 61.68}{0.249} = 1$$

Determine the probability.

$$\begin{aligned} P(61.3 < X < 61.93) &= \Phi(z_{\text{upper}}) - \Phi(z_{\text{lower}}) \\ &= \Phi(1) - \Phi(-1.53) \\ &= 0.7783 \end{aligned}$$

4. In a game, there is a 69% chance to win a round. You will play 195 rounds.
   (a) What is the probability of winning exactly 140 rounds?
   (b) What is the probability of winning at least 127 but at most 149 rounds?

Solution

We use the formula for binomial probabilities.

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

$$P(X = 140) = \binom{195}{140}(0.69)^{140}(1-0.69)^{195-140}$$

$$P(X = 140) = \binom{195}{140}(0.69)^{140}(0.31)^{55}$$

$$P(X = 140) = 0.0441$$

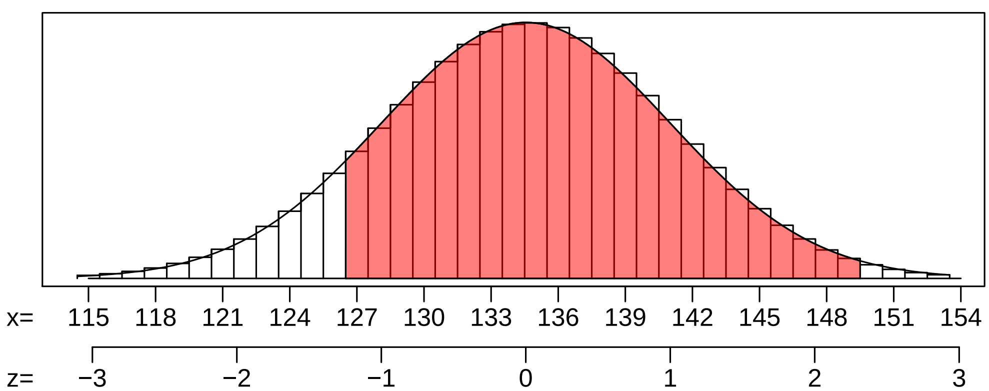
Find the mean.

$$\mu = np = (195)(0.69) = 134.55$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(195)(0.69)(1-0.69)} = 6.4584$$

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



Find the z scores.

$$z_1 = \frac{126.5 - 134.55}{6.4584} = -1.25$$

$$z_2 = \frac{149.5 - 134.55}{6.4584} = 2.31$$

Calculate the probability.

$$P(127 \leq X \leq 149) = \Phi(2.31) - \Phi(-1.25) = 0.884$$

(a) $P(X = 140) = 0.0441$

(b) $P(127 \leq X \leq 149) = 0.884$

5. As an ornithologist, you wish to determine the average body mass of *Passerina cyanea*. You randomly sample 19 adults of *Passerina cyanea*, resulting in a sample mean of 15.71 grams and a sample standard deviation of 2.39 grams. Determine a 95% confidence interval of the true population mean.

Solution

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

$$\begin{aligned} n &= 19 \\ \bar{x} &= 15.71 \\ s &= 2.39 \\ CL &= 0.95 \end{aligned}$$

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

$$df = n - 1 = 18$$

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

$$t^\star = 2.1$$

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{2.39}{\sqrt{19}} = 0.548$$

We want to make an inference about the population mean.

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

Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (15.71 - 2.1 \times 0.548,\ 15.71 + 2.1 \times 0.548) \\ &= (14.6,\ 16.9) \end{aligned}$$

We are 95% confident that the population mean is between 14.6 and 16.9.

6. A treatment group of size 37 has a mean of 0.948 and standard deviation of 0.18. A control group of size 34 has a mean of 1.07 and standard deviation of 0.251. 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 59.

   (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.

$$\begin{aligned} n_1 &= 37 \\ \bar{x}_1 &= 0.948 \\ s_1 &= 0.18 \\ n_2 &= 34 \\ \bar{x}_2 &= 1.07 \\ s_2 &= 0.251 \\ \alpha &= 0.02 \\ df &= 59 \end{aligned}$$

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.

$$t^\star = 2.39$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(0.18)^2}{37} + \frac{(0.251)^2}{34}} \\ &= 0.052 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(1.07 - 0.948) - (0)}{0.052} \\ &= 2.34 \end{aligned}$$

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.

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

Compare p-value and α.

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

We conclude that we should retain the null hypothesis.

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

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

(c) $t^\star = 2.39$

(d) $SE = 0.052$

(e) $|t_{obs}| = 2.34$

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

(g) retain the null

7. From a very large population, a random sample of 8000 individuals was taken. In that sample, 29.1% were angry. 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.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.291)(1-0.291)}{8000}} = 0.00508$$

Calculate the margin of error.

$$ME = z^\star SE = (2.58)(0.00508) = 0.0131$$

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.

$$(0.278\,,\ 0.304)$$

We are 99% confident that the true population proportion is between 27.8% and 30.4%.

(a) The lower bound = 0.278, which can also be expressed as 27.8%.

(b) The upper bound = 0.304, which can also be expressed as 30.4%.

8. An experiment is run with a treatment group of size 299 and a control group of size 282. The results are summarized in the table below.

	treatment	control
happy	161	176
not happy	138	106

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

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

Determine the sample proportions.

$$\hat{p}_1 = \frac{161}{299} = 0.538$$
$$\hat{p}_2 = \frac{176}{282} = 0.624$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.624 - 0.538 = 0.086$$

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

$$\hat{p} = \frac{161 + 176}{299 + 282} = 0.58$$

Determine the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}} \\ &= \sqrt{\frac{(0.58)(0.42)}{299} + \frac{(0.58)(0.42)}{282}} \\ &= 0.041 \end{aligned}$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0, 0.041)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.624 - 0.538) - 0}{0.041} \\ &= 2.1 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.1) \\ &= 0.0358 \end{aligned}$$

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) $SE = 0.041$

(e) $|z_{\text{obs}}| = 2.1$

(f) p-value = 0.0358

(g) reject the null