Bunker Hill Community College

Final Statistics Exam 2019-05-02 — Exam ID 002

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(**pink given bike**) = 0.321

   (b) P(**tree given pink**) = 0.21

   (c) P(**red**) = 0.283

   (d) P(**bike and red**) = 0.112

   (e) P(**bike or red**) = 0.502

   (f) P(**pig**) = 0.208

2. P(**"not tree" given "not yellow"**) = 0.511

3. $P(61.55 < X < 62.3) = 0.8251$

4. (a) $P(X = 30) = 0.0884$

   (b) $P(34 \leq X \leq 39) = 0.2244$

5. **(54.2, 57)**

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

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

   (c) $t^{\star} = 1.99$

   (d) $SE = 20.518$

   (e) $\mid t_{\textbf{obs}} \mid = 1.95$

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

   (g) **retain**

7. (a) **LB of p CI = 0.0182 or** 1.82%

   (b) **UB of p CI = 0.0198 or** 1.98%

8. (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) $SE = 0.166$

(e) $\mid z_{\mathbf{obs}} \mid = 2.14$

(f) p-**value** $= 0.0324$

(g) **reject**

1. In a deck of strange cards, there are 827 cards. Each card has an image and a color. The amounts are shown in the table below.

	gray	pink	red	white
bike	24	88	93	69
pig	87	27	42	16
tree	21	38	47	58
wheel	51	28	52	86

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

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

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

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

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

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

Solution

(a) $P(\text{pink given bike}) = \frac{88}{24+88+93+69} = 0.321$

(b) $P(\text{tree given pink}) = \frac{38}{88+27+38+28} = 0.21$

(c) $P(\text{red}) = \frac{93+42+47+52}{827} = 0.283$

(d) $P(\text{bike and red}) = \frac{93}{827} = 0.112$

(e) $P(\text{bike or red}) = \frac{24+88+93+69+93+42+47+52-93}{827} = 0.502$

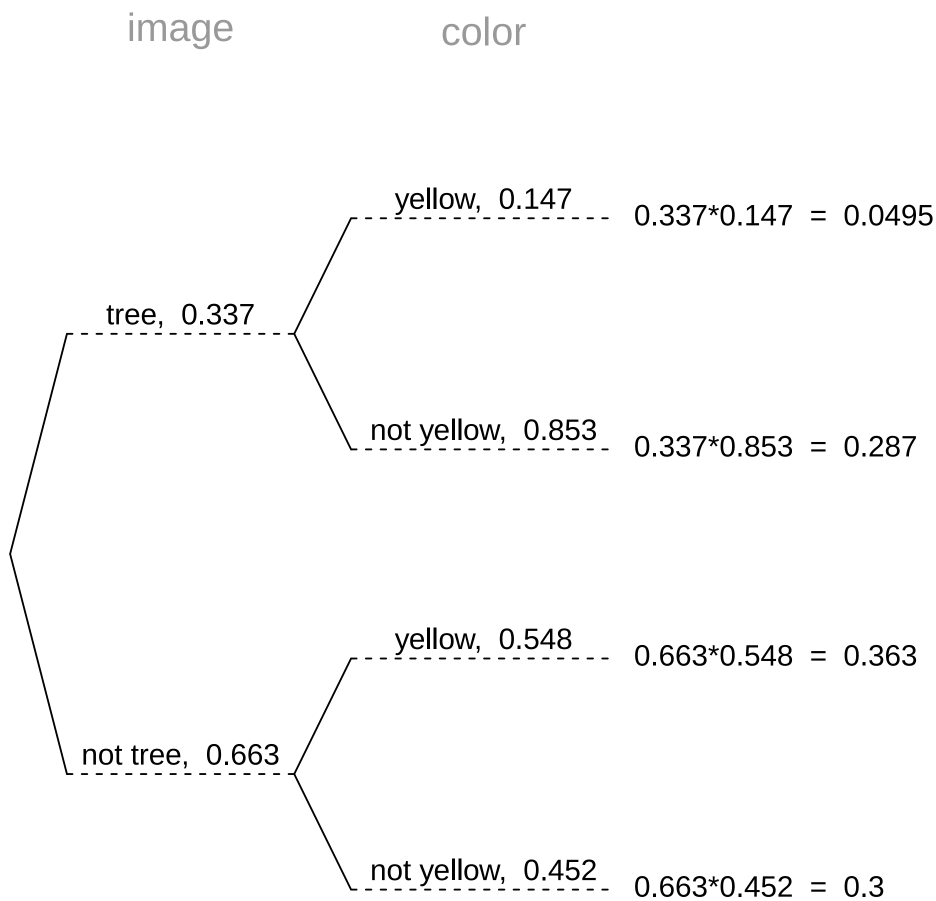
(f) $P(\text{pig}) = \frac{87+27+42+16}{827} = 0.208$

2. In a deck of strange cards, each card has an image and a color. The chance of drawing a tree is 33.7%. If a tree is drawn, there is a 14.7% chance that it is yellow. If a card that is not a tree is drawn, there is a 54.8% chance that it is yellow.

   Now, someone draws a random card and reveals it is not yellow. 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.

$$P(\text{"not tree" given "not yellow"}) = \frac{0.3}{0.3 + 0.287} = 0.511$$

3. In a very large pile of toothpicks, the mean length is 61.98 millimeters and the standard deviation is 3.52 millimeters. If you randomly sample 169 toothpicks, what is the chance the sample mean is between 61.55 and 62.3 millimeters?

Solution
Label the given information.

$$\mu = 61.98$$
$$\sigma = 3.52$$
$$n = 169$$
$$\bar{x}_{\text{lower}} = 61.55$$
$$\bar{x}_{\text{upper}} = 62.3$$

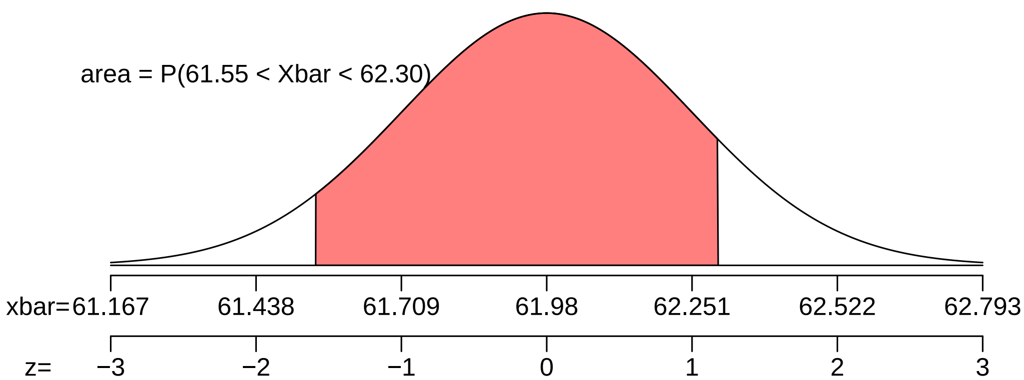
Find the standard error.

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{3.52}{\sqrt{169}} = 0.271$$

Describe the sampling distribution.

$$\bar{X} \sim \mathcal{N}(61.98,\ 0.271)$$

Draw a sketch.



Calculate a z scores.

$$z_{\text{lower}} = \frac{x_{\text{lower}} - \mu}{SE} = \frac{61.55 - 61.98}{0.271} = -1.59$$

$$z_{\text{upper}} = \frac{x_{\text{upper}} - \mu}{SE} = \frac{62.3 - 61.98}{0.271} = 1.18$$

Determine the probability.

$$\begin{aligned} P(61.55 < X < 62.3) &= \Phi(z_{\text{upper}}) - \Phi(z_{\text{lower}}) \\ &= \Phi(1.18) - \Phi(-1.59) \\ &= 0.8251 \end{aligned}$$

4. In a game, there is a 39% chance to win a round. You will play 80 rounds.
   (a) What is the probability of winning exactly 30 rounds?
   (b) What is the probability of winning at least 34 but at most 39 rounds?

Solution

We use the formula for binomial probabilities.

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

$$P(X = 30) = \binom{80}{30}(0.39)^{30}(1-0.39)^{80-30}$$

$$P(X = 30) = \binom{80}{30}(0.39)^{30}(0.61)^{50}$$

$$P(X = 30) = 0.0884$$

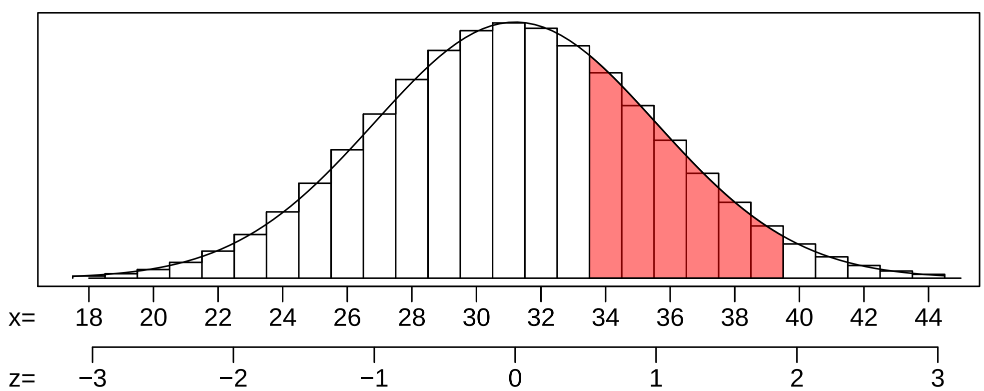
Find the mean.

$$\mu = np = (80)(0.39) = 31.2$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(80)(0.39)(1-0.39)} = 4.3626$$

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



Find the z scores.

$$z_1 = \frac{33.5 - 31.2}{4.3626} = 0.64$$

$$z_2 = \frac{39.5 - 31.2}{4.3626} = 1.79$$

Calculate the probability.

$$P(34 \le X \le 39) = \Phi(1.79) - \Phi(0.64) = 0.2244$$

(a) $P(X = 30) = 0.0884$

(b) $P(34 \le X \le 39) = 0.2244$

5. As an ornithologist, you wish to determine the average body mass of *Hylocichla mustelina*. You randomly sample 29 adults of *Hylocichla mustelina*, resulting in a sample mean of 55.61 grams and a sample standard deviation of 5.63 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.

$$\begin{aligned} n &= 29 \\ \bar{x} &= 55.61 \\ s &= 5.63 \\ CL &= 0.8 \end{aligned}$$

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

$$df = n - 1 = 28$$

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

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

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{5.63}{\sqrt{29}} = 1.05$$

We want to make an inference about the population mean.

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

Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^{\star} SE,\ \bar{x} + t^{\star} SE) \\ &= (55.61 - 1.31 \times 1.05,\ 55.61 + 1.31 \times 1.05) \\ &= (54.2,\ 57) \end{aligned}$$

We are 80% confident that the population mean is between 54.2 and 57.

6. A treatment group of size 32 has a mean of 1060 and standard deviation of 77.5. A control group of size 40 has a mean of 1020 and standard deviation of 96.6. If you decided to use a signficance level of 0.05, 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 69.

   (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.

$$n_1 = 32$$
$$\bar{x}_1 = 1060$$
$$s_1 = 77.5$$
$$n_2 = 40$$
$$\bar{x}_2 = 1020$$
$$s_2 = 96.6$$
$$\alpha = 0.05$$
$$df = 69$$

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 $P(|T| > t^\star) = 0.05$ by using a t table.

$$t^\star = 1.99$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(77.5)^2}{32} + \frac{(96.6)^2}{40}} \\ &= 20.518 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(1020 - 1060) - (0)}{20.518} \\ &= -1.95 \end{aligned}$$

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

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

(c) $t^{\star} = 1.99$

(d) $SE = 20.518$

(e) $|t_{\text{obs}}| = 1.95$

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

(g) retain the null

7. From a very large population, a random sample of 52000 individuals was taken. In that sample, 1.9% were angry. 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.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.019)(1-0.019)}{52000}} = 0.000599$$

Calculate the margin of error.

$$ME = z^\star SE = (1.28)(0.000599) = 0.000767$$

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.

$$(0.0182\,,\ 0.0198)$$

We are 80% confident that the true population proportion is between 1.82% and 1.98%.

(a) The lower bound = 0.0182, which can also be expressed as 1.82%.

(b) The upper bound = 0.0198, which can also be expressed as 1.98%.

8. An experiment is run with a treatment group of size 16 and a control group of size 21. The results are summarized in the table below.

	treatment	control
abysmal	5	14
not abysmal	11	7

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

(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.

$$\hat{p}_1 = \frac{5}{16} = 0.312$$
$$\hat{p}_2 = \frac{14}{21} = 0.667$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.667 - 0.312 = 0.355$$

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

$$\hat{p} = \frac{5+14}{16+21} = 0.514$$

Determine the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}} \\ &= \sqrt{\frac{(0.514)(0.486)}{16} + \frac{(0.514)(0.486)}{21}} \\ &= 0.166 \end{aligned}$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0, 0.166)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.667 - 0.312) - 0}{0.166} \\ &= 2.14 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.14) \\ &= 0.0324 \end{aligned}$$

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) $SE = 0.166$

(e) $|z_{\text{obs}}| = 2.14$

(f) p-value = 0.0324

(g) reject the null