Bunker Hill Community College

Final Statistics Exam 2019-05-02   **Exam ID 015**

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) $P(\textbf{flower or white}) = 0.426$

   (b) $P(\textbf{pink given bike}) = 0.109$

   (c) $P(\textbf{orange}) = 0.218$

   (d) $P(\textbf{wheel and blue}) = 0.0372$

   (e) $P(\textbf{bike given blue}) = 0.128$

   (f) $P(\textbf{cat}) = 0.32$

2. $P(\textbf{"not llama" given "yellow"}) = 0.886$

3. $P(60.5 < X < 60.79) = 0.8051$

4. (a) $P(X = 140) = 0.0519$

   (b) $P(144 \leq X \leq 150) = 0.307$

5. $(9.36, 11.3)$

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

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

   (c) $t^\star = 1.69$

   (d) $SE = 8.867$

   (e) $| \ t_{\textbf{obs}} \ | = 1.76$

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

   (g) **reject**

7. (a) **LB of p CI = 0.0974 or** 9.74%

   (b) **UB of p CI = 0.113 or** 11.3%

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

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

(c) $z^\star = 2.05$

(d) $SE = 0.068$

(e) $|\ z_{\mathbf{obs}}\ | = 2.21$

(f) $p\text{-}\mathbf{value} = 0.0272$

(g) **reject**

1. In a deck of strange cards, there are 1128 cards. Each card has an image and a color. The amounts are shown in the table below.

	blue	orange	pink	teal	white
bike	32	28	18	68	19
cat	85	95	52	74	55
flower	91	79	50	71	90
wheel	42	44	94	16	25

(a) What is the probability a random card is either a flower or white (or both)?

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

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

(d) What is the probability a random card is both a wheel and blue?

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

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

Solution

(a) P(flower or white) $= \frac{91+79+50+71+90+19+55+90+25-90}{1128} = 0.426$

(b) P(pink given bike) $= \frac{18}{32+28+18+68+19} = 0.109$

(c) P(orange) $= \frac{28+95+79+44}{1128} = 0.218$

(d) P(wheel and blue) $= \frac{42}{1128} = 0.0372$
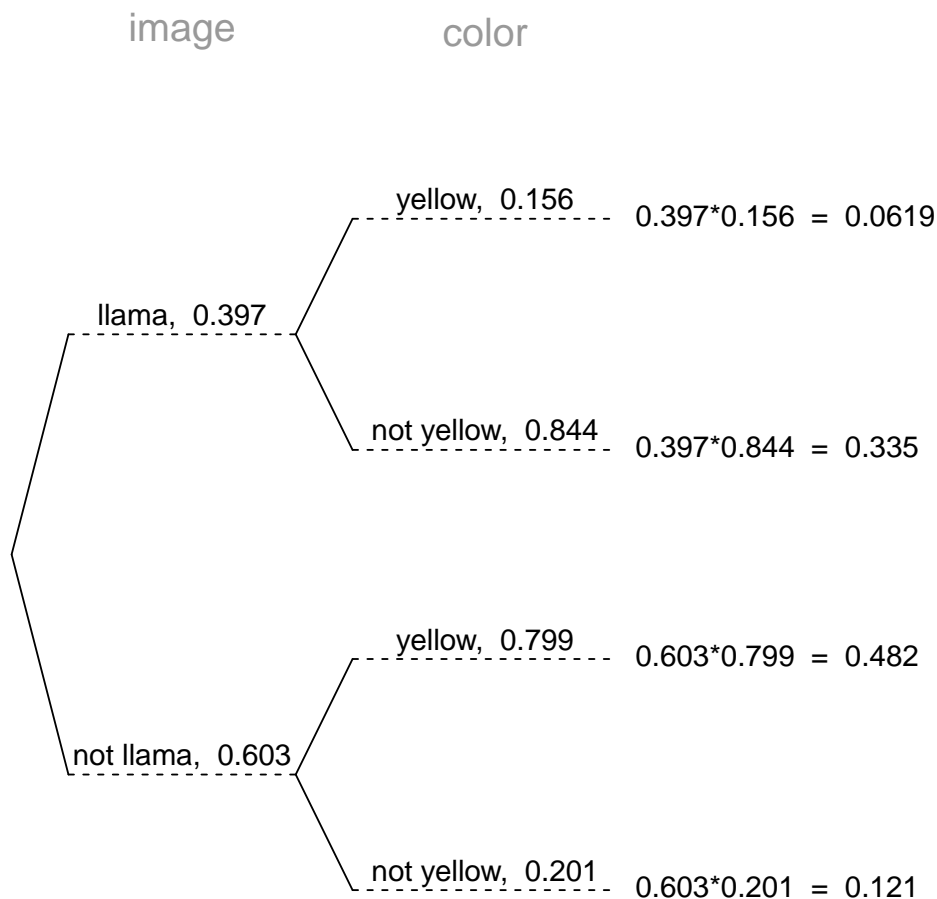
(e) P(bike given blue) $= \frac{32}{32+85+91+42} = 0.128$

(f) P(cat) $= \frac{85+95+52+74+55}{1128} = 0.32$

2. In a deck of strange cards, each card has an image and a color. The chance of drawing a llama is 39.7%. If a llama is drawn, there is a 15.6% chance that it is yellow. If a card that is not a llama is drawn, there is a 79.9% chance that it is yellow.

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

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.

image   color

llama, 0.397

yellow, 0.156 — 0.397*0.156 = 0.0619

not yellow, 0.844 — 0.397*0.844 = 0.335

not llama, 0.603

yellow, 0.799 — 0.603*0.799 = 0.482

not yellow, 0.201 — 0.603*0.201 = 0.121

Determine the appropriate conditional probability.

$$P(\text{"not llama" given "yellow"}) = \frac{0.482}{0.482 + 0.0619} = 0.886$$

3. In a very large pile of toothpicks, the mean length is 60.6 millimeters and the standard deviation is 1.35 millimeters. If you randomly sample 175 toothpicks, what is the chance the sample mean is between 60.5 and 60.79 millimeters?

Solution

Label the given information.

$$\mu = 60.6$$
$$\sigma = 1.35$$
$$n = 175$$
$$\bar{x}_{\text{lower}} = 60.5$$
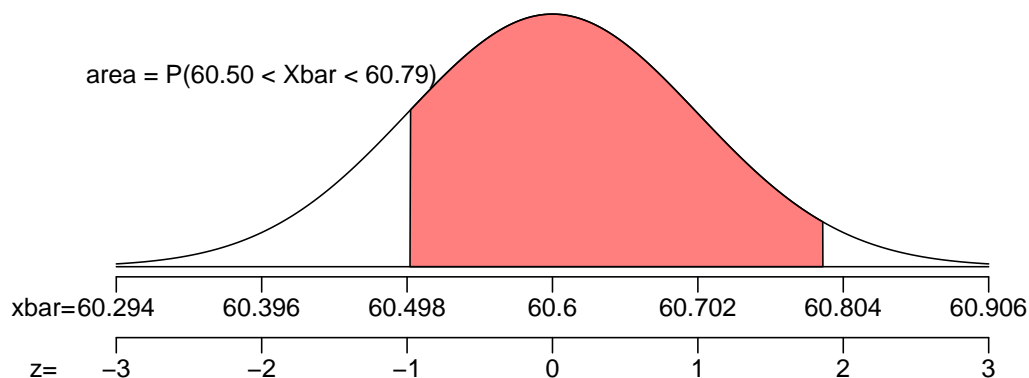$$\bar{x}_{\text{upper}} = 60.79$$

Find the standard error.

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{1.35}{\sqrt{175}} = 0.102$$

Describe the sampling distribution.

$$\bar{X} \sim \mathcal{N}(60.6, \ 0.102)$$

Draw a sketch.

Calculate a z scores.

$$z_{\text{lower}} = \frac{x_{\text{lower}} - \mu}{SE} = \frac{60.5 - 60.6}{0.102} = -0.98$$

$$z_{\text{upper}} = \frac{x_{\text{upper}} - \mu}{SE} = \frac{60.79 - 60.6}{0.102} = 1.86$$

Determine the probability.

$$P(60.5 < X < 60.79) = \Phi(z_{\text{upper}}) - \Phi(z_{\text{lower}})$$
$$= \Phi(1.86) - \Phi(-0.98)$$
$$= 0.8051$$

4. In a game, there is a 70% chance to win a round. You will play 205 rounds.

    (a) What is the probability of winning exactly 140 rounds?

    (b) What is the probability of winning at least 144 but at most 150 rounds?

Solution

We use the formula for binomial probabilities.

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

$$P(X = 140) = \binom{205}{140}(0.7)^{140}(1-0.7)^{205-140}$$

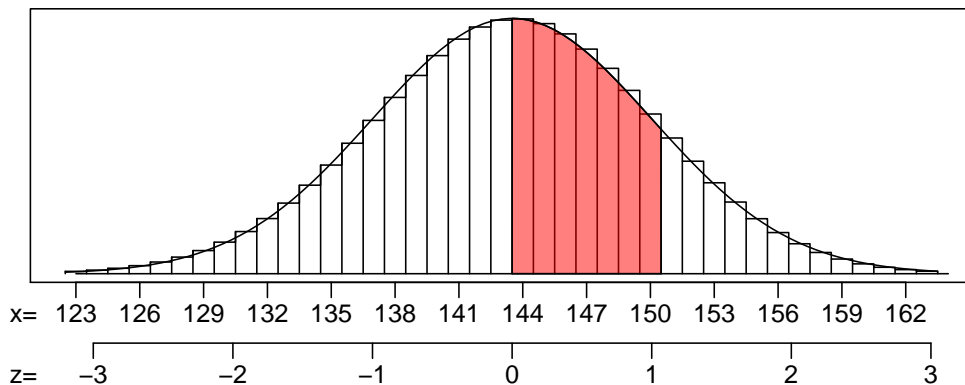$$P(X = 140) = \binom{205}{140}(0.7)^{140}(0.3)^{65}$$

$$P(X = 140) = 0.0519$$

Find the mean.

$$\mu = np = (205)(0.7) = 143.5$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(205)(0.7)(1-0.7)} = 6.5612$$

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

Find the z scores.

$$z_1 = \frac{143.5 - 143.5}{6.5612} = 0.08$$

$$z_2 = \frac{150.5 - 143.5}{6.5612} = 0.99$$

Calculate the probability.

$$P(144 \leq X \leq 150) = \Phi(0.99) - \Phi(0.08) = 0.307$$

(a) $P(X = 140) = 0.0519$

(b) $P(144 \leq X \leq 150) = 0.307$

5. As an ornithologist, you wish to determine the average body mass of *Dendroica dominica*. You randomly sample 25 adults of *Dendroica dominica*, resulting in a sample mean of 10.35 grams and a sample standard deviation of 1.99 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.

$$n = 25$$
$$\bar{x} = 10.35$$
$$s = 1.99$$
$$CL = 0.98$$

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

$$df = n - 1 = 24$$

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

$$t^\star = 2.49$$

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.99}{\sqrt{25}} = 0.398$$

We want to make an inference about the population mean.

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

Determine the bounds.

$$
\begin{aligned}
CI &= (\bar{x} - t^\star SE, \ \bar{x} + t^\star SE) \\
&= (10.35 - 2.49 \times 0.398, \ 10.35 + 2.49 \times 0.398) \\
&= (9.36, \ 11.3)
\end{aligned}
$$

We are 98% confident that the population mean is between 9.36 and 11.3.

6. A treatment group of size 20 has a mean of 94.4 and standard deviation of 25. A control group of size 19 has a mean of 110 and standard deviation of 30. 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 35.

   (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.

$$n_1 = 20$$
$$\bar{x}_1 = 94.4$$
$$s_1 = 25$$
$$n_2 = 19$$
$$\bar{x}_2 = 110$$
$$s_2 = 30$$
$$\alpha = 0.1$$
$$df = 35$$

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.

$$SE = \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$$
$$= \sqrt{\frac{(25)^2}{20} + \frac{(30)^2}{19}}$$
$$= 8.867$$

Determine the test statistic.

$$t_{obs} = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE}$$
$$= \frac{(110 - 94.4) - (0)}{8.867}$$
$$= 1.76$$

Compare $|t_{obs}|$ and $t^\star$.

$$|t_{obs}| > t^\star$$

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

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

(c) $t^\star = 1.69$

(d) $SE = 8.867$

(e) $|t_{obs}| = 1.76$

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

(g) reject the null

7. From a very large population, a random sample of 8900 individuals was taken. In that sample, 10.5% were purple. 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.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.105)(1-0.105)}{8900}} = 0.00325$$

Calculate the margin of error.

$$ME = z^\star SE = (2.33)(0.00325) = 0.00757$$

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.

$$(0.0974\ ,\ 0.113)$$

We are 98% confident that the true population proportion is between 9.74% and 11.3%.

(a) The lower bound = 0.0974, which can also be expressed as 9.74%.

(b) The upper bound = 0.113, which can also be expressed as 11.3%.

8. An experiment is run with a treatment group of size 99 and a control group of size 118. The results are summarized in the table below.

	treatment	control
green	61	55
not green	38	63

Using a significance level of 0.04, 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 $P(|Z| > z^\star) = 0.04$.

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

Determine the sample proportions.

$$\hat{p}_1 = \frac{61}{99} = 0.616$$

$$\hat{p}_2 = \frac{55}{118} = 0.466$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.466 - 0.616 = -0.15$$

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

$$\hat{p} = \frac{61 + 55}{99 + 118} = 0.535$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.535)(0.465)}{99} + \frac{(0.535)(0.465)}{118}}$$

$$= 0.068$$

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

$$H_0 : \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0, 0.068)$$

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.

$$Z = \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE}$$

$$= \frac{(0.466 - 0.616) - 0}{0.068}$$

$$= -2.21$$

Determine the p-value.

$$p\text{-value} = 2 \cdot \Phi(-|z|)$$
$$= 2 \cdot \Phi(-2.21)$$
$$= 0.0272$$

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) $SE = 0.068$

(e) $|z_{obs}| = 2.21$

(f) $p\text{-value} = 0.0272$

(g) reject the null