Bunker Hill Community College

Final Statistics Exam 2019-05-02   **Exam ID 012**

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) $P(\textbf{pig given black}) = 0.117$

   (b) $P(\textbf{pig and green}) = 0.0868$

   (c) $P(\textbf{green}) = 0.299$

   (d) $P(\textbf{horn or green}) = 0.483$

   (e) $P(\textbf{green given bike}) = 0.156$

   (f) $P(\textbf{wheel}) = 0.166$

2. $P(\textbf{"not shovel" given "not yellow"}) = 0.577$

3. $P(60.8 < X < 61.26) = 0.7409$

4. (a) $P(X = 18) = 0.0924$

   (b) $P(22 \leq X \leq 29) = 0.2799$

5. $(13.8, 16)$

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

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

   (c) $t^{\star} = 2.46$

   (d) $SE = 0.756$

   (e) $| t_{\textbf{obs}} | = 2.51$

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

   (g) **reject**

7. (a) **LB of p CI = 0.274 or** 27.4%

   (b) **UB of p CI = 0.31 or** 31%

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

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

      (c) $z^\star = 1.96$

      (d) $SE = 0.066$

      (e) $| \; z_{obs} \; | = 2.08$

      (f) $p\text{-value} = 0.0376$

      (g)   **reject**

1. In a deck of strange cards, there are 1002 cards. Each card has an image and a color. The amounts are shown in the table below.

	black	green	orange
bike	41	20	67
flower	88	85	79
horn	91	71	93
pig	38	87	76
wheel	66	37	63

   (a) What is the probability a random card is a pig given it is black?

   (b) What is the probability a random card is both a pig and green?

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

   (d) What is the probability a random card is either a horn or green (or both)?

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

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

Solution

(a) $P(\text{pig given black}) = \frac{38}{41+88+91+38+66} = 0.117$

(b) $P(\text{pig and green}) = \frac{87}{1002} = 0.0868$

(c) $P(\text{green}) = \frac{20+85+71+87+37}{1002} = 0.299$

(d) $P(\text{horn or green}) = \frac{91+71+93+20+85+71+87+37-71}{1002} = 0.483$

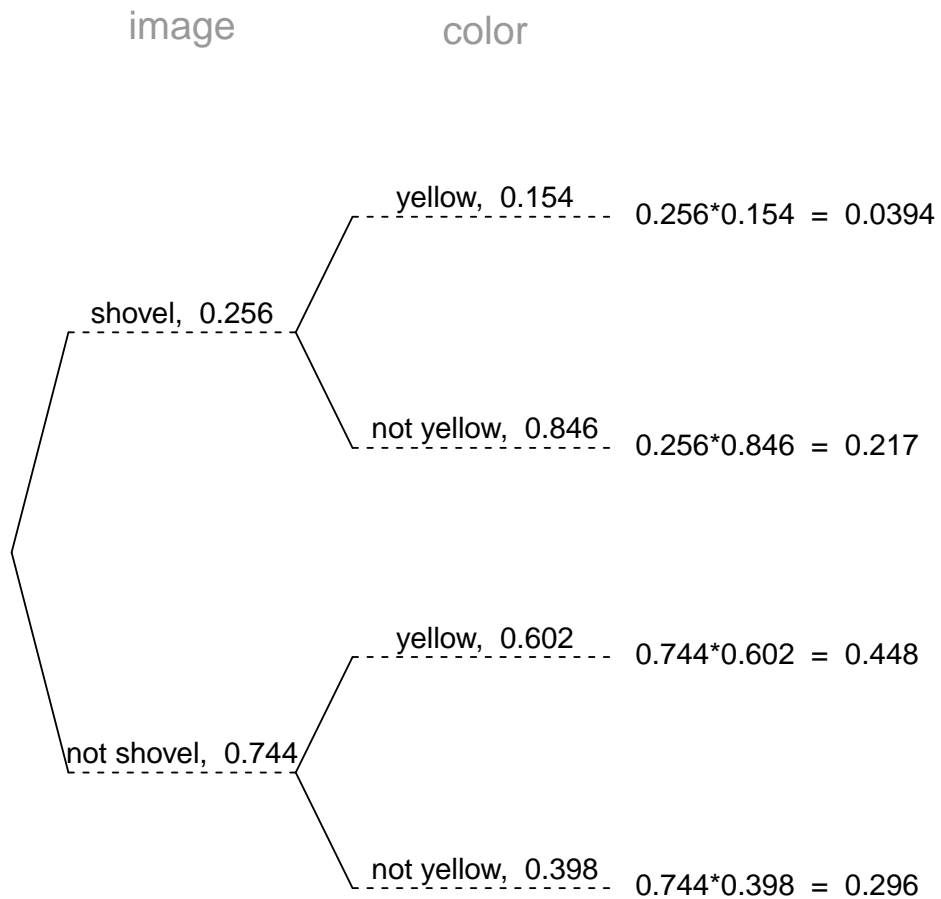(e) $P(\text{green given bike}) = \frac{20}{41+20+67} = 0.156$

(f) $P(\text{wheel}) = \frac{66+37+63}{1002} = 0.166$

2. In a deck of strange cards, each card has an image and a color. The chance of drawing a shovel is 25.6%. If a shovel is drawn, there is a 15.4% chance that it is yellow. If a card that is not a shovel is drawn, there is a 60.2% 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 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.

image   color

yellow, 0.154   0.256*0.154 = 0.0394

shovel, 0.256

not yellow, 0.846   0.256*0.846 = 0.217

yellow, 0.602   0.744*0.602 = 0.448

not shovel, 0.744

not yellow, 0.398   0.744*0.398 = 0.296

Determine the appropriate conditional probability.

$$P(\text{"not shovel" given "not yellow"}) = \frac{0.296}{0.296 + 0.217} = 0.577$$

3. In a very large pile of toothpicks, the mean length is 61.04 millimeters and the standard deviation is 2.45 millimeters. If you randomly sample 144 toothpicks, what is the chance the sample mean is between 60.8 and 61.26 millimeters?

Solution

Label the given information.

$$\mu = 61.04$$
$$\sigma = 2.45$$
$$n = 144$$
$$\bar{x}_{\text{lower}} = 60.8$$
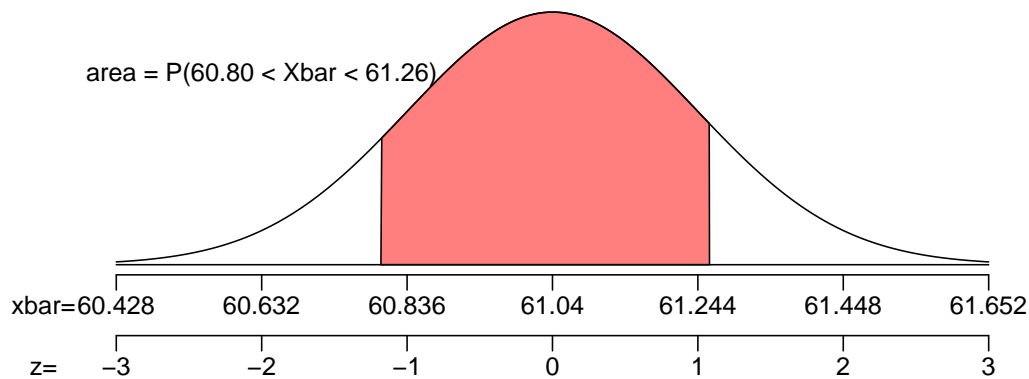$$\bar{x}_{\text{upper}} = 61.26$$

Find the standard error.

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{2.45}{\sqrt{144}} = 0.204$$

Describe the sampling distribution.

$$\bar{X} \sim \mathcal{N}(61.04,\ 0.204)$$

Draw a sketch.

Calculate a z scores.

$$z_{\text{lower}} = \frac{x_{\text{lower}} - \mu}{SE} = \frac{60.8 - 61.04}{0.204} = -1.18$$

$$z_{\text{upper}} = \frac{x_{\text{upper}} - \mu}{SE} = \frac{61.26 - 61.04}{0.204} = 1.08$$

Determine the probability.

$$P(60.8 < X < 61.26) = \Phi(z_{\text{upper}}) - \Phi(z_{\text{lower}})$$
$$= \Phi(1.08) - \Phi(-1.18)$$
$$= 0.7409$$

4. In a game, there is a 18% chance to win a round. You will play 110 rounds.

   (a) What is the probability of winning exactly 18 rounds?

   (b) What is the probability of winning at least 22 but at most 29 rounds?

Solution

We use the formula for binomial probabilities.

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

$$P(X = 18) = \binom{110}{18}(0.18)^{18}(1-0.18)^{110-18}$$

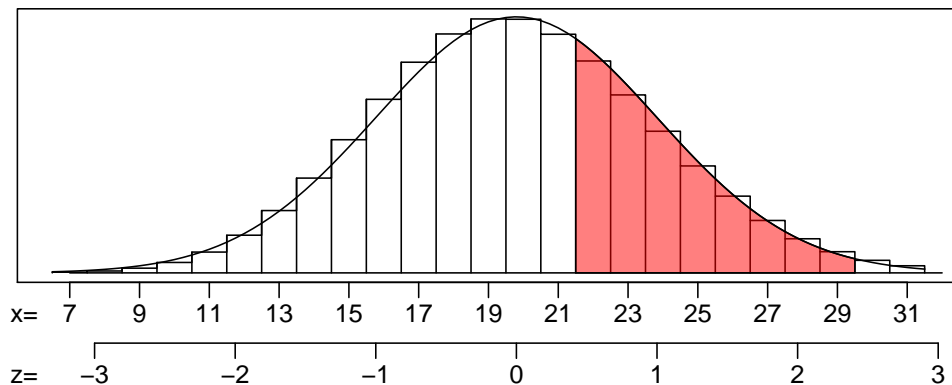$$P(X = 18) = \binom{110}{18}(0.18)^{18}(0.82)^{92}$$

$$P(X = 18) = 0.0924$$

Find the mean.

$$\mu = np = (110)(0.18) = 19.8$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(110)(0.18)(1-0.18)} = 4.0294$$

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

Find the z scores.

$$z_1 = \frac{21.5 - 19.8}{4.0294} = 0.55$$

$$z_2 = \frac{29.5 - 19.8}{4.0294} = 2.28$$

Calculate the probability.

$$P(22 \leq X \leq 29) = \Phi(2.28) - \Phi(0.55) = 0.2799$$

(a) $P(X = 18) = 0.0924$

(b) $P(22 \leq X \leq 29) = 0.2799$

5. As an ornithologist, you wish to determine the average body mass of *Passerina cyanea*. You randomly sample 24 adults of *Passerina cyanea*, resulting in a sample mean of 14.92 grams and a sample standard deviation of 2.16 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.

$$n = 24$$
$$\bar{x} = 14.92$$
$$s = 2.16$$
$$CL = 0.98$$

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

$$df = n - 1 = 23$$

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

$$t^{\star} = 2.5$$

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{2.16}{\sqrt{24}} = 0.441$$

We want to make an inference about the population mean.

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

Determine the bounds.

$$
\begin{aligned}
CI &= (\bar{x} - t^{\star} SE,\ \bar{x} + t^{\star} SE) \\
&= (14.92 - 2.5 \times 0.441,\ 14.92 + 2.5 \times 0.441) \\
&= (13.8,\ 16)
\end{aligned}
$$

We are 98% confident that the population mean is between 13.8 and 16.

6. A treatment group of size 17 has a mean of 12.1 and standard deviation of 1.9. A control group of size 17 has a mean of 10.2 and standard deviation of 2.47. 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 30.

(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.

$$n_1 = 17$$
$$\bar{x}_1 = 12.1$$
$$s_1 = 1.9$$
$$n_2 = 17$$
$$\bar{x}_2 = 10.2$$
$$s_2 = 2.47$$
$$\alpha = 0.02$$
$$df = 30$$

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.

$$t^\star = 2.46$$

Calculate the standard error.

$$SE = \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$$
$$= \sqrt{\frac{(1.9)^2}{17} + \frac{(2.47)^2}{17}}$$
$$= 0.756$$

Determine the test statistic.

$$t_{obs} = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE}$$
$$= \frac{(10.2 - 12.1) - (0)}{0.756}$$
$$= -2.51$$

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

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

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) $t^* = 2.46$

(d) $SE = 0.756$

(e) $|t_{obs}| = 2.51$

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

(g) reject the null

7. From a very large population, a random sample of 5300 individuals was taken. In that sample, 29.2% were super. Determine a 99.5% 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 $P(|Z| < z^\star) = 0.995$.

$$z^\star = 2.81$$

Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = \sqrt{\frac{(0.292)(1 - 0.292)}{5300}} = 0.00625$$

Calculate the margin of error.

$$ME = z^\star SE = (2.81)(0.00625) = 0.0176$$

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.

$$(0.274 \, , \, 0.31)$$

We are 99.5% confident that the true population proportion is between 27.4% and 31%.

(a) The lower bound = 0.274, which can also be expressed as 27.4%.

(b) The upper bound = 0.31, which can also be expressed as 31%.

8. An experiment is run with a treatment group of size 117 and a control group of size 71. The results are summarized in the table below.

	treatment	control
sick	92	46
not sick	25	25

Using a significance level of 0.05, determine whether the treatment causes an effect on the proportion of cases that are sick.

(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.

$$\hat{p}_1 = \frac{92}{117} = 0.786$$

$$\hat{p}_2 = \frac{46}{71} = 0.648$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.648 - 0.786 = -0.138$$

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

$$\hat{p} = \frac{92 + 46}{117 + 71} = 0.734$$

Determine the standard error.

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

$$= \sqrt{\frac{(0.734)(0.266)}{117} + \frac{(0.734)(0.266)}{71}}$$

$$= 0.0665$$

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

$$H_0 : \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0665)$$

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.

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

$$= \frac{(0.648 - 0.786) - 0}{0.0665}$$

$$= -2.08$$

Determine the p-value.

$$p\text{-value} = 2 \cdot \Phi(-|z|)$$
$$= 2 \cdot \Phi(-2.08)$$
$$= 0.0376$$

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) $SE = 0.0665$

(e) $|z_{\text{obs}}| = 2.08$

 (f) $p\text{-value} = 0.0376$

(g) reject the null