Bunker Hill Community College

Final Statistics Exam 2019-05-02 | Exam ID 009

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 orange**) = 0.163

   (b) P(**wheel and gray**) = 0.0172

   (c) P(**wheel or orange**) = 0.439

   (d) P(**flower**) = 0.202

   (e) P(**orange**) = 0.362

   (f) P(**gray given gem**) = 0.18

2. P(**"gem" given "pink"**) = 0.0891

3. $P(70.39 < X < 70.68) = 0.8111$

4. (a) $P(X = 97) = 0.0554$

   (b) $P(78 \leq X \leq 96) = 0.5916$

5. $\mathbf{(12, 13.1)}$

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

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

   (c) $t^{\star} = 2.51$

   (d) $SE = 5.215$

   (e) $\mid t_{\mathbf{obs}} \mid = 2.21$

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

   (g) **retain**

7. (a) **LB of p CI = 0.663 or** 66.3%

   (b) **UB of p CI = 0.701 or** 70.1%

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

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

(c) $z^\star = 2.33$

(d) $SE = 0.025$

(e) $\mid z_{\mathbf{obs}} \mid = 2.49$

(f) p-**value** = 0.0128

(g) **reject**

1. In a deck of strange cards, there are 816 cards. Each card has an image and a color. The amounts are shown in the table below.

	gray	orange	red
bike	56	99	55
flower	77	64	24
gem	27	40	83
tree	62	48	74
wheel	14	44	49

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

(b) What is the probability a random card is both a wheel and gray?

(c) What is the probability a random card is either a wheel or orange (or both)?

(d) What is the probability a random card is a flower?

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

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

Solution

(a) P(tree given orange) = $\frac{48}{99+64+40+48+44}$ = 0.163

(b) P(wheel and gray) = $\frac{14}{816}$ = 0.0172

(c) P(wheel or orange) = $\frac{14+44+49+99+64+40+48+44-44}{816}$ = 0.439

(d) P(flower) = $\frac{77+64+24}{816}$ = 0.202

(e) P(orange) = $\frac{99+64+40+48+44}{816}$ = 0.362

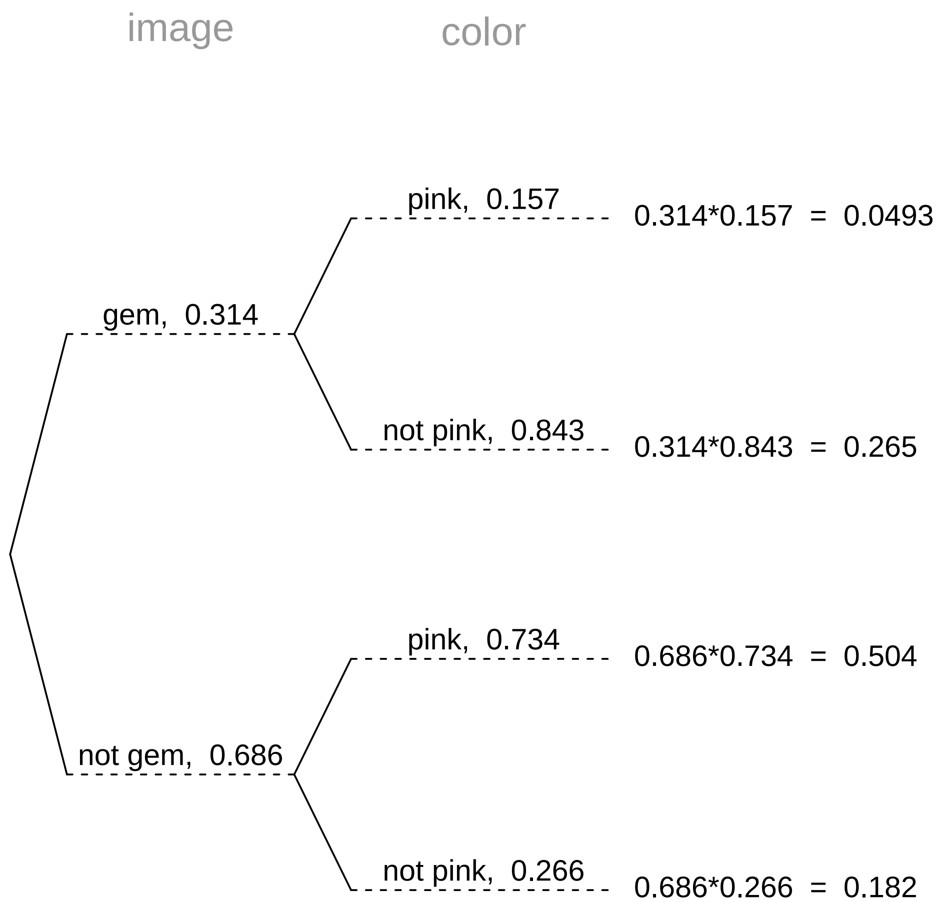
(f) P(gray given gem) = $\frac{27}{27+40+83}$ = 0.18

2. In a deck of strange cards, each card has an image and a color. The chance of drawing a gem is 31.4%. If a gem is drawn, there is a 15.7% chance that it is pink. If a card that is not a gem is drawn, there is a 73.4% chance that it is pink.

   Now, someone draws a random card and reveals it is pink. What is the chance the card is a gem?

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.

$$P(\text{"gem" given "pink"}) = \frac{0.0493}{0.0493 + 0.504} = 0.0891$$

3. In a very large pile of toothpicks, the mean length is 70.54 millimeters and the standard deviation is 1.35 millimeters. If you randomly sample 150 toothpicks, what is the chance the sample mean is between 70.39 and 70.68 millimeters?

Solution

Label the given information.

$$\mu = 70.54$$
$$\sigma = 1.35$$
$$n = 150$$
$$\bar{x}_{\text{lower}} = 70.39$$
$$\bar{x}_{\text{upper}} = 70.68$$

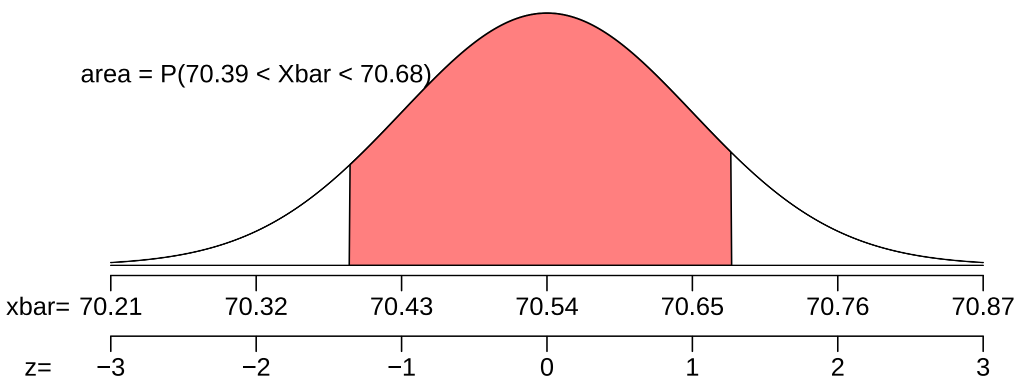
Find the standard error.

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{1.35}{\sqrt{150}} = 0.11$$

Describe the sampling distribution.

$$\bar{X} \sim \mathcal{N}(70.54,\ 0.11)$$

Draw a sketch.



Calculate a z scores.

$$z_{\text{lower}} = \frac{x_{\text{lower}} - \mu}{SE} = \frac{70.39 - 70.54}{0.11} = -1.36$$

$$z_{\text{upper}} = \frac{x_{\text{upper}} - \mu}{SE} = \frac{70.68 - 70.54}{0.11} = 1.27$$

Determine the probability.

$$\begin{aligned} P(70.39 < X < 70.68) &= \Phi(z_{\text{upper}}) - \Phi(z_{\text{lower}}) \\ &= \Phi(1.27) - \Phi(-1.36) \\ &= 0.8111 \end{aligned}$$

4. In a game, there is a 53% chance to win a round. You will play 178 rounds.
   (a) What is the probability of winning exactly 97 rounds?
   (b) What is the probability of winning at least 78 but at most 96 rounds?

Solution

We use the formula for binomial probabilities.

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

$$P(X = 97) = \binom{178}{97}(0.53)^{97}(1-0.53)^{178-97}$$

$$P(X = 97) = \binom{178}{97}(0.53)^{97}(0.47)^{81}$$

$$P(X = 97) = 0.0554$$

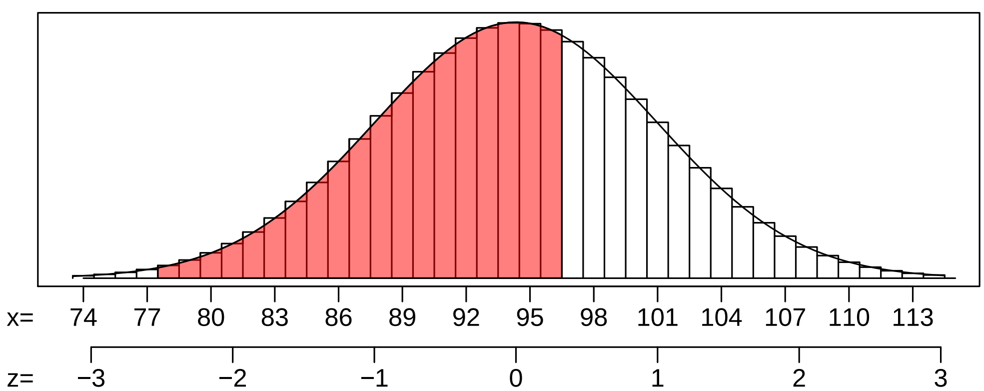
Find the mean.

$$\mu = np = (178)(0.53) = 94.34$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(178)(0.53)(1-0.53)} = 6.6588$$

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



Find the z scores.

$$z_1 = \frac{77.5 - 94.34}{6.6588} = -2.45$$

$$z_2 = \frac{96.5 - 94.34}{6.6588} = 0.25$$

Calculate the probability.

$$P(78 \leq X \leq 96) = \Phi(0.25) - \Phi(-2.45) = 0.5916$$

(a) $P(X = 97) = 0.0554$

(b) $P(78 \leq X \leq 96) = 0.5916$

5. As an ornithologist, you wish to determine the average body mass of *Vermivora peregrina*. You randomly sample 23 adults of *Vermivora peregrina*, resulting in a sample mean of 12.52 grams and a sample standard deviation of 1.24 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.

$$\begin{aligned} n &= 23 \\ \bar{x} &= 12.52 \\ s &= 1.24 \\ CL &= 0.95 \end{aligned}$$

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

$$df = n - 1 = 22$$

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

$$t^\star = 2.07$$

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.24}{\sqrt{23}} = 0.259$$

We want to make an inference about the population mean.

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

Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (12.52 - 2.07 \times 0.259,\ 12.52 + 2.07 \times 0.259) \\ &= (12,\ 13.1) \end{aligned}$$

We are 95% confident that the population mean is between 12 and 13.1.

6. A treatment group of size 14 has a mean of 98.5 and standard deviation of 16.4. A control group of size 16 has a mean of 110 and standard deviation of 11.3. 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 22.

   (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.

$$\begin{aligned} n_1 &= 14 \\ \bar{x}_1 &= 98.5 \\ s_1 &= 16.4 \\ n_2 &= 16 \\ \bar{x}_2 &= 110 \\ s_2 &= 11.3 \\ \alpha &= 0.02 \\ df &= 22 \end{aligned}$$

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.

$$t^\star = 2.51$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(16.4)^2}{14} + \frac{(11.3)^2}{16}} \\ &= 5.215 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(110 - 98.5) - (0)}{5.215} \\ &= 2.21 \end{aligned}$$

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

(c) $t^{\star} = 2.51$

(d) $SE = 5.215$

(e) $|t_{\text{obs}}| = 2.21$

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

(g) retain the null

7. From a very large population, a random sample of 5000 individuals was taken. In that sample, 68.2% were blue. 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.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.682)(1-0.682)}{5000}} = 0.00659$$

Calculate the margin of error.

$$ME = z^\star SE = (2.81)(0.00659) = 0.0185$$

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.

$$(0.663\,,\ 0.701)$$

We are 99.5% confident that the true population proportion is between 66.3% and 70.1%.

(a) The lower bound = 0.663, which can also be expressed as 66.3%.

(b) The upper bound = 0.701, which can also be expressed as 70.1%.

8. An experiment is run with a treatment group of size 273 and a control group of size 275. The results are summarized in the table below.

	treatment	control
pink	238	257
not pink	35	18

Using a significance level of 0.02, determine whether the treatment causes an effect on the proportion of cases that are pink.

(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.

$$\hat{p}_1 = \frac{238}{273} = 0.872$$
$$\hat{p}_2 = \frac{257}{275} = 0.935$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.935 - 0.872 = 0.063$$

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

$$\hat{p} = \frac{238 + 257}{273 + 275} = 0.903$$

Determine the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}$$
$$= \sqrt{\frac{(0.903)(0.097)}{273} + \frac{(0.903)(0.097)}{275}}$$
$$= 0.0253$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0253)$$

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.

$$z = \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE}$$
$$= \frac{(0.935 - 0.872) - 0}{0.0253}$$
$$= 2.49$$

Determine the p-value.

$$p\text{-value} = 2 \cdot \Phi(-|z|)$$
$$= 2 \cdot \Phi(-2.49)$$
$$= 0.0128$$

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) $SE = 0.0253$

(e) $|z_{\text{obs}}| = 2.49$

(f) p-value = 0.0128

(g) reject the null