Bunker Hill Community College

Final Statistics Exam 2019-05-02 Exam ID 020

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(**cat**) = 0.202

   (b) P(**cat or yellow**) = 0.351

   (c) P(**red**) = 0.286

   (d) P(**pig and yellow**) = 0.0179

   (e) P(**pig given teal**) = 0.108

   (f) P(**indigo given pig**) = 0.337

2. P(**"tree" given "not indigo"**) = 0.104

3. $P(66.65 < X < 66.98) = 0.7677$

4. (a) $P(X = 39) = 0.0852$

   (b) $P(30 \leq X \leq 45) = 0.904$

5. **(9.33, 10.9)**

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

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

   (c) $t^\star = 1.73$

   (d) $SE = 7.13$

   (e) $| \ t_{\mathbf{obs}} \ | = 1.99$

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

   (g) **reject**

7. (a) **LB of p CI = 0.418 or** 41.8%

   (b) **UB of p CI = 0.458 or** 45.8%

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) $SE = 0.074$

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

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

(g) **reject**

1. In a deck of strange cards, there are 1175 cards. Each card has an image and a color. The amounts are shown in the table below.

	indigo	red	teal	yellow
cat	22	94	72	49
dog	26	61	48	16
flower	73	14	93	58
pig	69	78	37	21
shovel	83	89	92	80

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

(b) What is the probability a random card is either a cat or yellow (or both)?

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

(d) What is the probability a random card is both a pig and yellow?

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

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

Solution

(a) $P(\text{cat}) = \frac{22+94+72+49}{1175} = 0.202$

(b) $P(\text{cat or yellow}) = \frac{22+94+72+49+49+16+58+21+80-49}{1175} = 0.351$

(c) $P(\text{red}) = \frac{94+61+14+78+89}{1175} = 0.286$

(d) $P(\text{pig and yellow}) = \frac{21}{1175} = 0.0179$

(e) $P(\text{pig given teal}) = \frac{37}{72+48+93+37+92} = 0.108$

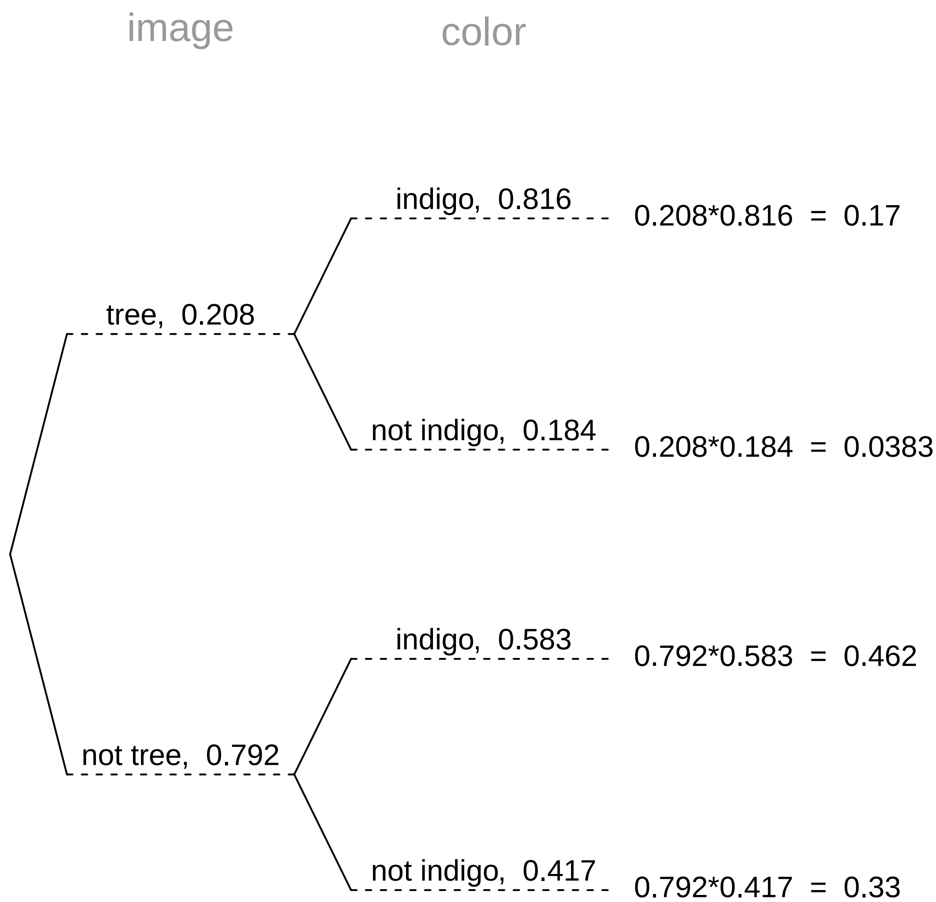
(f) $P(\text{indigo given pig}) = \frac{69}{69+78+37+21} = 0.337$

2. In a deck of strange cards, each card has an image and a color. The chance of drawing a tree is 20.8%. If a tree is drawn, there is a 81.6% chance that it is indigo. If a card that is not a tree is drawn, there is a 58.3% chance that it is indigo.

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

$$P(\text{"tree" given "not indigo"}) = \frac{0.0383}{0.0383 + 0.33} = 0.104$$

3. In a very large pile of toothpicks, the mean length is 66.81 millimeters and the standard deviation is 1.95 millimeters. If you randomly sample 200 toothpicks, what is the chance the sample mean is between 66.65 and 66.98 millimeters?

Solution

Label the given information.

$$\mu = 66.81$$
$$\sigma = 1.95$$
$$n = 200$$
$$\bar{x}_{\text{lower}} = 66.65$$
$$\bar{x}_{\text{upper}} = 66.98$$

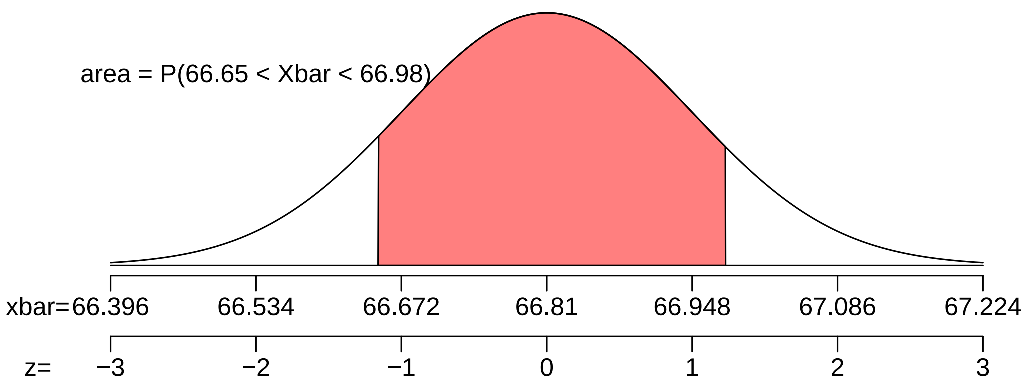
Find the standard error.

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{1.95}{\sqrt{200}} = 0.138$$

Describe the sampling distribution.

$$\bar{X} \sim \mathcal{N}(66.81,\ 0.138)$$

Draw a sketch.



Calculate a z scores.

$$z_{\text{lower}} = \frac{x_{\text{lower}} - \mu}{SE} = \frac{66.65 - 66.81}{0.138} = -1.16$$

$$z_{\text{upper}} = \frac{x_{\text{upper}} - \mu}{SE} = \frac{66.98 - 66.81}{0.138} = 1.23$$

Determine the probability.

$$\begin{aligned} P(66.65 < X < 66.98) &= \Phi(z_{\text{upper}}) - \Phi(z_{\text{lower}}) \\ &= \Phi(1.23) - \Phi(-1.16) \\ &= 0.7677 \end{aligned}$$

4. In a game, there is a 44% chance to win a round. You will play 88 rounds.
   (a) What is the probability of winning exactly 39 rounds?
   (b) What is the probability of winning at least 30 but at most 45 rounds?

Solution
We use the formula for binomial probabilities.

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

$$P(X = 39) = \binom{88}{39}(0.44)^{39}(1-0.44)^{88-39}$$

$$P(X = 39) = \binom{88}{39}(0.44)^{39}(0.56)^{49}$$

$$P(X = 39) = 0.0852$$

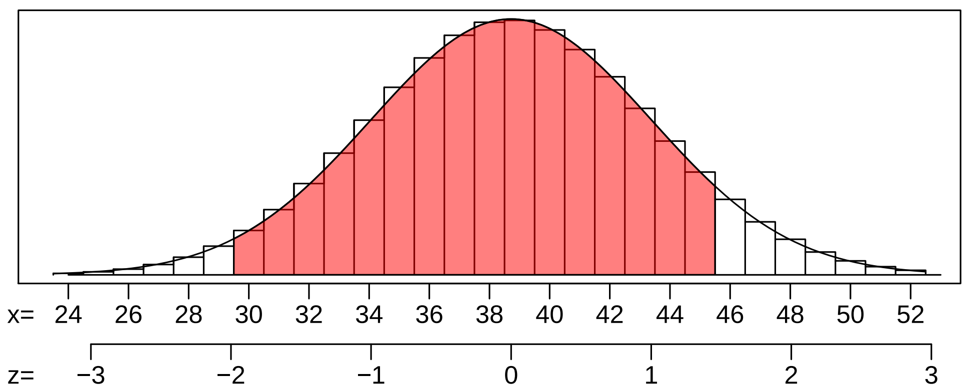
Find the mean.

$$\mu = np = (88)(0.44) = 38.72$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(88)(0.44)(1-0.44)} = 4.6565$$

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



Find the z scores.

$$z_1 = \frac{29.5 - 38.72}{4.6565} = -1.98$$

$$z_2 = \frac{45.5 - 38.72}{4.6565} = 1.46$$

Calculate the probability.

$$P(30 \le X \le 45) = \Phi(1.46) - \Phi(-1.98) = 0.904$$

(a) $P(X = 39) = 0.0852$

(b) $P(30 \le X \le 45) = 0.904$

5. As an ornithologist, you wish to determine the average body mass of *Wilsonia citrina*. You randomly sample 28 adults of *Wilsonia citrina*, resulting in a sample mean of 10.14 grams and a sample standard deviation of 1.98 grams. Determine a 96% confidence interval of the true population mean.

Solution

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

$$\begin{aligned} n &= 28 \\ \bar{x} &= 10.14 \\ s &= 1.98 \\ CL &= 0.96 \end{aligned}$$

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

$$df = n - 1 = 27$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.96$.

$$t^\star = 2.16$$

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.98}{\sqrt{28}} = 0.374$$

We want to make an inference about the population mean.

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

Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (10.14 - 2.16 \times 0.374,\ 10.14 + 2.16 \times 0.374) \\ &= (9.33,\ 10.9) \end{aligned}$$

We are 96% confident that the population mean is between 9.33 and 10.9.

6. A treatment group of size 32 has a mean of 112 and standard deviation of 20.7. A control group of size 12 has a mean of 97.8 and standard deviation of 21.2. 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 19.

   (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.

$$\begin{aligned} n_1 &= 32 \\ \bar{x}_1 &= 112 \\ s_1 &= 20.7 \\ n_2 &= 12 \\ \bar{x}_2 &= 97.8 \\ s_2 &= 21.2 \\ \alpha &= 0.1 \\ df &= 19 \end{aligned}$$

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 $P(|T| > t^\star) = 0.1$ by using a t table.

$$t^\star = 1.73$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(20.7)^2}{32} + \frac{(21.2)^2}{12}} \\ &= 7.13 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(97.8 - 112) - (0)}{7.13} \\ &= -1.99 \end{aligned}$$

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 reject the null hypothesis.

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

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

(c) $t^\star = 1.73$

(d) $SE = 7.13$

(e) $|t_{obs}| = 1.99$

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

(g) reject the null

7. From a very large population, a random sample of 4000 individuals was taken. In that sample, 43.8% were purple. 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.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.438)(1-0.438)}{4000}} = 0.00784$$

Calculate the margin of error.

$$ME = z^{\star} SE = (2.58)(0.00784) = 0.0202$$

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.

$$(0.418\,,\ 0.458)$$

We are 99% confident that the true population proportion is between 41.8% and 45.8%.

(a) The lower bound = 0.418, which can also be expressed as 41.8%.

(b) The upper bound = 0.458, which can also be expressed as 45.8%.

8. An experiment is run with a treatment group of size 50 and a control group of size 84. The results are summarized in the table below.

	treatment	control
sorry	6	23
not sorry	44	61

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

(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.

$$\hat{p}_1 = \frac{6}{50} = 0.12$$
$$\hat{p}_2 = \frac{23}{84} = 0.274$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.274 - 0.12 = 0.154$$

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

$$\hat{p} = \frac{6 + 23}{50 + 84} = 0.216$$

Determine the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}} \\ &= \sqrt{\frac{(0.216)(0.784)}{50} + \frac{(0.216)(0.784)}{84}} \\ &= 0.0735 \end{aligned}$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0735)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.274 - 0.12) - 0}{0.0735} \\ &= 2.1 \end{aligned}$$

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) $H_0: \; p_2 - p_1 = 0$

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

(c) $z^{\star} = 1.96$

(d) $SE = 0.0735$

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

(f) p-value = 0.0358

(g) reject the null