# Bunker Hill Community College

## Final Statistics Exam 2019-05-02 Exam ID 026

**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) $P(\textbf{cat}) = 0.341$

   (b) $P(\textbf{bike given white}) = 0.154$

   (c) $P(\textbf{pink}) = 0.285$

   (d) $P(\textbf{cat and red}) = 0.0825$

   (e) $P(\textbf{bike or red}) = 0.353$

   (f) $P(\textbf{pink given cat}) = 0.135$

2. $P(\textbf{"kite" given "blue"}) = 0.2$

3. $P(73.88 < X < 74.3) = 0.8589$

4. (a) $P(X = 206) = 0.0689$

   (b) $P(200 \leq X \leq 210) = 0.672$

5. $\mathbf{(29.5, 37.3)}$

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

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

   (c) $t^{\star} = 2.14$

   (d) $SE = 80.545$

   (e) $\mid t_{\textbf{obs}} \mid = 1.99$

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

   (g) **retain**

7. (a) **LB of p CI = 0.961 or** 96.1%

   (b) **UB of p CI = 0.965 or** 96.5%

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) $SE = 0.033$

(e) $\mid z_{\mathbf{obs}} \mid = 2.31$

(f) $p$-**value** = 0.0208

(g) **reject**

1. In a deck of strange cards, there are 909 cards. Each card has an image and a color. The amounts are shown in the table below.

| | orange | pink | red | white |
|---|---|---|---|---|
| bike | 23 | 65 | 33 | 34 |
| cat | 94 | 42 | 75 | 99 |
| gem | 29 | 64 | 21 | 50 |
| tree | 84 | 88 | 70 | 38 |

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

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

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

(d) What is the probability a random card is both a cat 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 pink given it is a cat?

**Solution**

(a) $P(\text{cat}) = \frac{94+42+75+99}{909} = 0.341$

(b) $P(\text{bike given white}) = \frac{34}{34+99+50+38} = 0.154$

(c) $P(\text{pink}) = \frac{65+42+64+88}{909} = 0.285$

(d) $P(\text{cat and red}) = \frac{75}{909} = 0.0825$

(e) $P(\text{bike or red}) = \frac{23+65+33+34+33+75+21+70-33}{909} = 0.353$

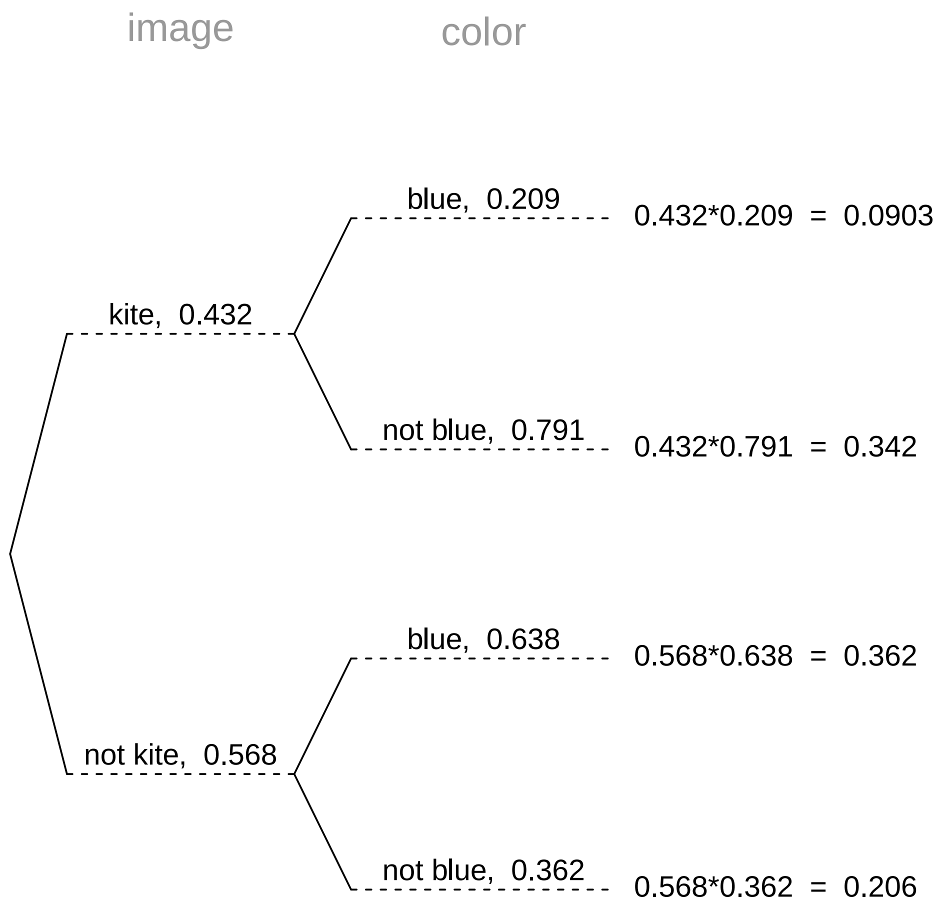
(f) $P(\text{pink given cat}) = \frac{42}{94+42+75+99} = 0.135$

2. In a deck of strange cards, each card has an image and a color. The chance of drawing a kite is 43.2%. If a kite is drawn, there is a 20.9% chance that it is blue. If a card that is not a kite is drawn, there is a 63.8% chance that it is blue.

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

**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.

$$P(\text{"kite" given "blue"}) = \frac{0.0903}{0.0903 + 0.362} = 0.2$$

3. In a very large pile of toothpicks, the mean length is 74.16 millimeters and the standard deviation is 1.73 millimeters. If you randomly sample 196 toothpicks, what is the chance the sample mean is between 73.88 and 74.3 millimeters?

**Solution**

Label the given information.

$$\mu = 74.16$$
$$\sigma = 1.73$$
$$n = 196$$
$$\bar{x}_{\text{lower}} = 73.88$$
$$\bar{x}_{\text{upper}} = 74.3$$

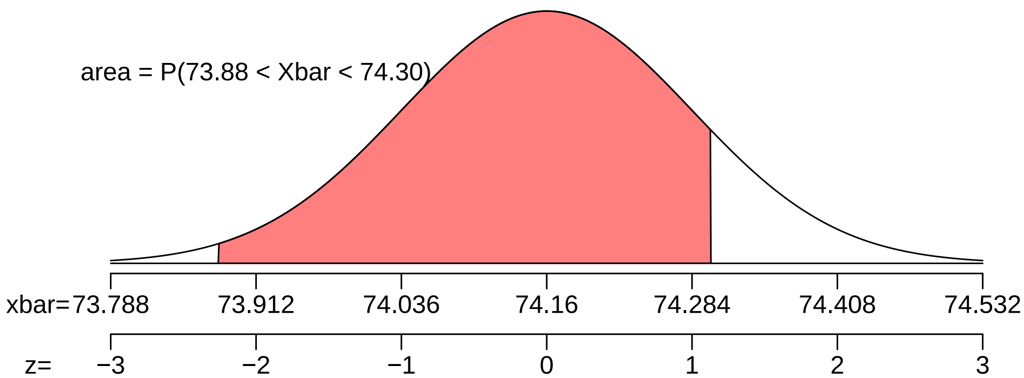
Find the standard error.

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{1.73}{\sqrt{196}} = 0.124$$

Describe the sampling distribution.

$$\bar{X} \sim \mathcal{N}(74.16, 0.124)$$

Draw a sketch.




Calculate a $z$ scores.

$$z_{\text{lower}} = \frac{x_{\text{lower}} - \mu}{SE} = \frac{73.88 - 74.16}{0.124} = -2.26$$

$$z_{\text{upper}} = \frac{x_{\text{upper}} - \mu}{SE} = \frac{74.3 - 74.16}{0.124} = 1.13$$

Determine the probability.

$$\begin{aligned} P(73.88 < X < 74.3) &= \Phi(z_{\text{upper}}) - \Phi(z_{\text{lower}}) \\ &= \Phi(1.13) - \Phi(-2.26) \\ &= 0.8589 \end{aligned}$$

4. In a game, there is a 85% chance to win a round. You will play 240 rounds.
   (a) What is the probability of winning exactly 206 rounds?
   (b) What is the probability of winning at least 200 but at most 210 rounds?

**Solution**

We use the formula for binomial probabilities.

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

$$P(X = 206) = \binom{240}{206}(0.85)^{206}(1-0.85)^{240-206}$$

$$P(X = 206) = \binom{240}{206}(0.85)^{206}(0.15)^{34}$$

$$P(X = 206) = 0.0689$$

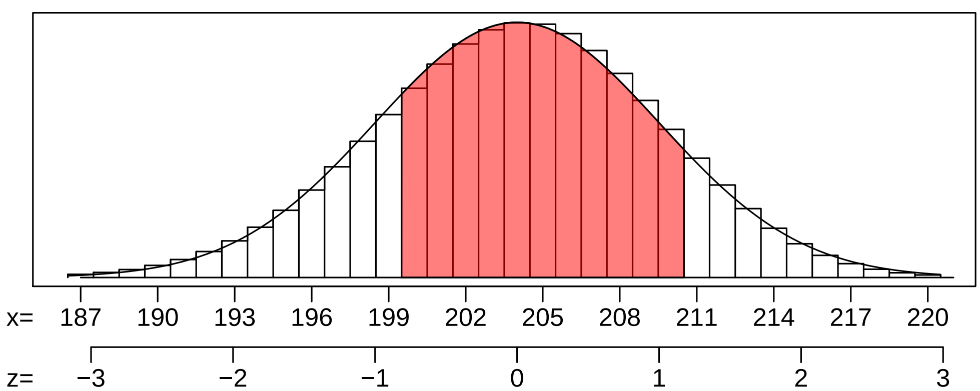
Find the mean.

$$\mu = np = (240)(0.85) = 204$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(240)(0.85)(1-0.85)} = 5.5317$$

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




Find the $z$ scores.

$$z_1 = \frac{199.5 - 204}{5.5317} = -0.81$$

$$z_2 = \frac{210.5 - 204}{5.5317} = 1.18$$

Calculate the probability.

$$P(200 \le X \le 210) = \Phi(1.18) - \Phi(-0.81) = 0.672$$

(a) $P(X = 206) = 0.0689$

(b) $P(200 \le X \le 210) = 0.672$

5. As an ornithologist, you wish to determine the average body mass of *Dolichonyx orizivorus*. You randomly sample 34 adults of *Dolichonyx orizivorus*, resulting in a sample mean of 33.38 grams and a sample standard deviation of 9.28 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.

$$\begin{aligned} n &= 34 \\ \bar{x} &= 33.38 \\ s &= 9.28 \\ CL &= 0.98 \end{aligned}$$

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

$$df = n - 1 = 33$$

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

$$t^\star = 2.44$$

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{9.28}{\sqrt{34}} = 1.59$$

We want to make an inference about the population mean.

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

Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (33.38 - 2.44 \times 1.59,\ 33.38 + 2.44 \times 1.59) \\ &= (29.5,\ 37.3) \end{aligned}$$

We are 98% confident that the population mean is between 29.5 and 37.3.

6. A treatment group of size 16 has a mean of 1190 and standard deviation of 245. A control group of size 19 has a mean of 1030 and standard deviation of 228. If you decided to use a signficance level of 0.04, 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 31.

   (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 $z_{\text{obs}}$ or $t_{\text{obs}}$. 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.

$$n_1 = 16$$
$$\bar{x}_1 = 1190$$
$$s_1 = 245$$
$$n_2 = 19$$
$$\bar{x}_2 = 1030$$
$$s_2 = 228$$
$$\alpha = 0.04$$
$$df = 31$$

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 $P(|T| > t^\star) = 0.04$ by using a $t$ table.

$$t^\star = 2.14$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(245)^2}{16} + \frac{(228)^2}{19}} \\ &= 80.545 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(1030 - 1190) - (0)}{80.545} \\ &= -1.99 \end{aligned}$$

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 $\alpha$.

$$p\text{-value} > \alpha$$

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) $t^\star = 2.14$

(d) $SE = 80.545$

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

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

(g) retain the null

7. From a very large population, a random sample of 20000 individuals was taken. In that sample, 96.3% were angry. Determine a 90% 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 $P(|Z| < z^\star) = 0.9$.

$$z^\star = 1.64$$

Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.963)(1-0.963)}{20000}} = 0.00133$$

Calculate the margin of error.

$$ME = z^\star SE = (1.64)(0.00133) = 0.00218$$

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.

$$(0.961\ ,\ 0.965)$$

We are 90% confident that the true population proportion is between 96.1% and 96.5%.

(a) The lower bound = 0.961, which can also be expressed as 96.1%.

(b) The upper bound = 0.965, which can also be expressed as 96.5%.

8. An experiment is run with a treatment group of size 201 and a control group of size 226. The results are summarized in the table below.

| | treatment | control |
|---|---|---|
| happy | 36 | 23 |
| not happy | 165 | 203 |

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

(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 $z_{\text{obs}}$ or $t_{\text{obs}}$. 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.

$$\hat{p}_1 = \frac{36}{201} = 0.179$$
$$\hat{p}_2 = \frac{23}{226} = 0.102$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.102 - 0.179 = -0.077$$

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

$$\hat{p} = \frac{36 + 23}{201 + 226} = 0.138$$

Determine the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}} \\ &= \sqrt{\frac{(0.138)(0.862)}{201} + \frac{(0.138)(0.862)}{226}} \\ &= 0.0334 \end{aligned}$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0334)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.102 - 0.179) - 0}{0.0334} \\ &= -2.31 \end{aligned}$$

Determine the $p$-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.31) \\ &= 0.0208 \end{aligned}$$

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) $SE = 0.0334$

(e) $|z_{obs}| = 2.31$

(f) $p$-value = 0.0208

(g) reject the null