Bunker Hill Community College

Final Statistics Exam 2019-05-02 Exam ID 028

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(**flower**) = 0.326

   (b) P(**yellow given bike**) = 0.145

   (c) P(**flower or green**) = 0.515

   (d) P(**cat and green**) = 0.0728

   (e) P(**blue**) = 0.22

   (f) P(**flower given black**) = 0.538

2. P(**"cat" given "not yellow"**) = 0.334

3. $P(62.45 < X < 63.21) = 0.7865$

4. (a) $P(X = 35) = 0.0524$

   (b) $P(31 \leq X \leq 52) = 0.9548$

5. $\mathbf{(9.08, 11)}$

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

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

   (c) $t^{\star} = 2.72$

   (d) $SE = 7.788$

   (e) $\mid t_{\mathbf{obs}} \mid = 2.44$

   (f) $0.01 < p$**-value** < 0.02

   (g) **retain**

7. (a) **LB of p CI = 0.725 or** 72.5%

   (b) **UB of p CI = 0.733 or** 73.3%

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

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

(c) $z^\star = 2.81$

(d) $SE = 0.155$

(e) $|\ z_{\mathbf{obs}}\ | = 2.69$

(f) p-**value** = 0.0072

(g) **retain**

1. In a deck of strange cards, there are 961 cards. Each card has an image and a color. The amounts are shown in the table below.

	black	blue	green	red	yellow
bike	24	51	64	26	28
cat	33	52	70	83	62
dog	22	12	48	44	29
flower	92	96	77	17	31

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

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

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

   (d) What is the probability a random card is both a cat and green?

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

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

Solution

(a) $P(\text{flower}) = \frac{92+96+77+17+31}{961} = 0.326$

(b) $P(\text{yellow given bike}) = \frac{28}{24+51+64+26+28} = 0.145$

(c) $P(\text{flower or green}) = \frac{92+96+77+17+31+64+70+48+77-77}{961} = 0.515$

(d) $P(\text{cat and green}) = \frac{70}{961} = 0.0728$

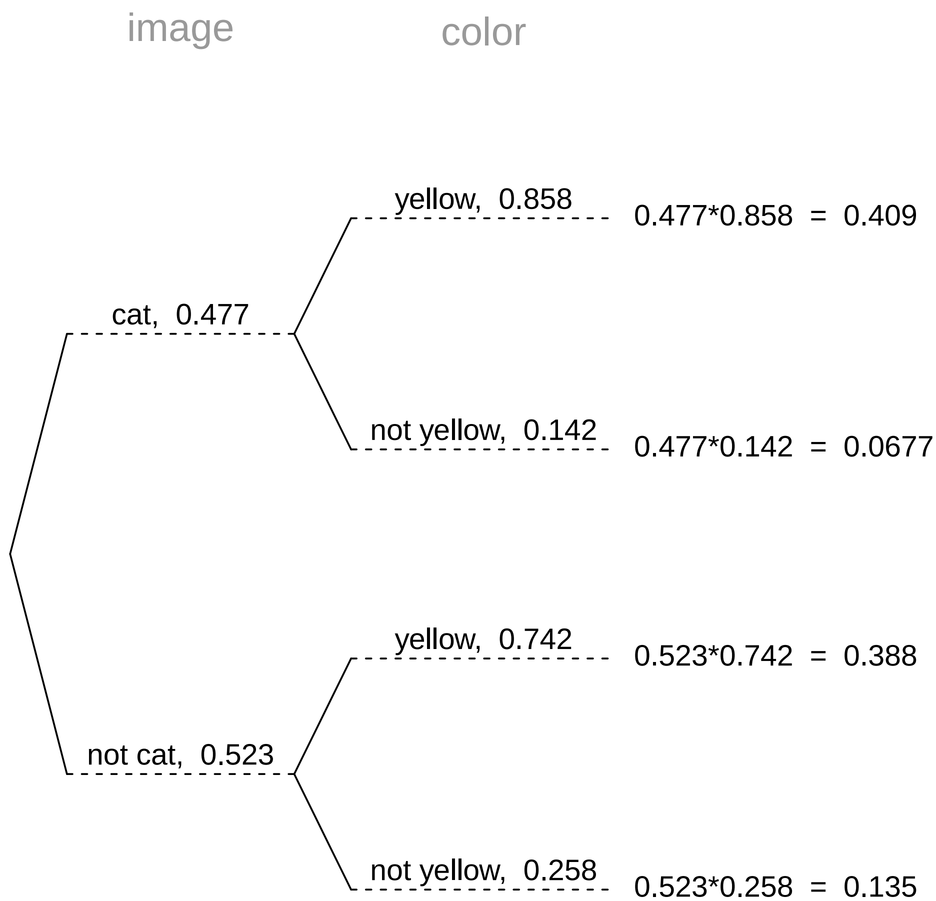
(e) $P(\text{blue}) = \frac{51+52+12+96}{961} = 0.22$

(f) $P(\text{flower given black}) = \frac{92}{24+33+22+92} = 0.538$

2. In a deck of strange cards, each card has an image and a color. The chance of drawing a cat is 47.7%. If a cat is drawn, there is a 85.8% chance that it is yellow. If a card that is not a cat is drawn, there is a 74.2% chance that it is yellow.

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

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.

$$P(\text{"cat" given "not yellow"}) = \frac{0.0677}{0.0677 + 0.135} = 0.334$$

3. In a very large pile of toothpicks, the mean length is 62.97 millimeters and the standard deviation is 2.96 millimeters. If you randomly sample 121 toothpicks, what is the chance the sample mean is between 62.45 and 63.21 millimeters?

Solution

Label the given information.

$$\mu = 62.97$$
$$\sigma = 2.96$$
$$n = 121$$
$$\bar{x}_{\text{lower}} = 62.45$$
$$\bar{x}_{\text{upper}} = 63.21$$

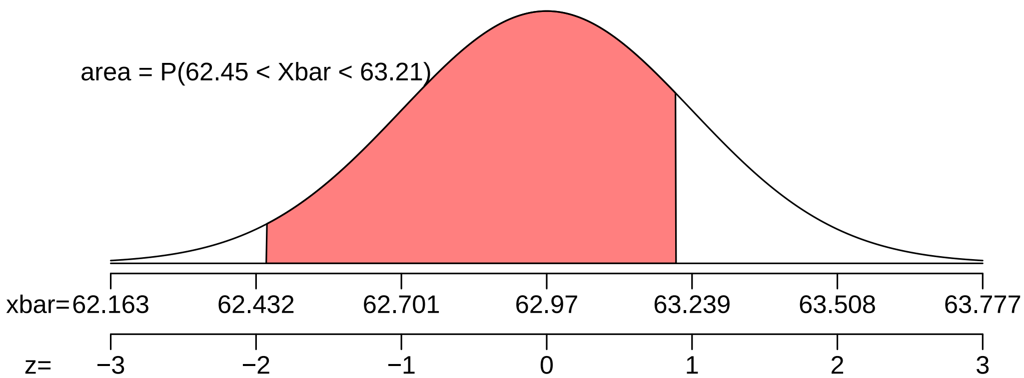
Find the standard error.

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{2.96}{\sqrt{121}} = 0.269$$

Describe the sampling distribution.

$$\bar{X} \sim \mathcal{N}(62.97,\ 0.269)$$

Draw a sketch.



Calculate a z scores.

$$z_{\text{lower}} = \frac{x_{\text{lower}} - \mu}{SE} = \frac{62.45 - 62.97}{0.269} = -1.93$$

$$z_{\text{upper}} = \frac{x_{\text{upper}} - \mu}{SE} = \frac{63.21 - 62.97}{0.269} = 0.89$$

Determine the probability.

$$\begin{aligned} P(62.45 < X < 63.21) &= \Phi(z_{\text{upper}}) - \Phi(z_{\text{lower}}) \\ &= \Phi(0.89) - \Phi(-1.93) \\ &= 0.7865 \end{aligned}$$

4. In a game, there is a 32% chance to win a round. You will play 124 rounds.
   (a) What is the probability of winning exactly 35 rounds?
   (b) What is the probability of winning at least 31 but at most 52 rounds?

Solution
We use the formula for binomial probabilities.

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

$$P(X = 35) = \binom{124}{35}(0.32)^{35}(1-0.32)^{124-35}$$

$$P(X = 35) = \binom{124}{35}(0.32)^{35}(0.68)^{89}$$

$$P(X = 35) = 0.0524$$

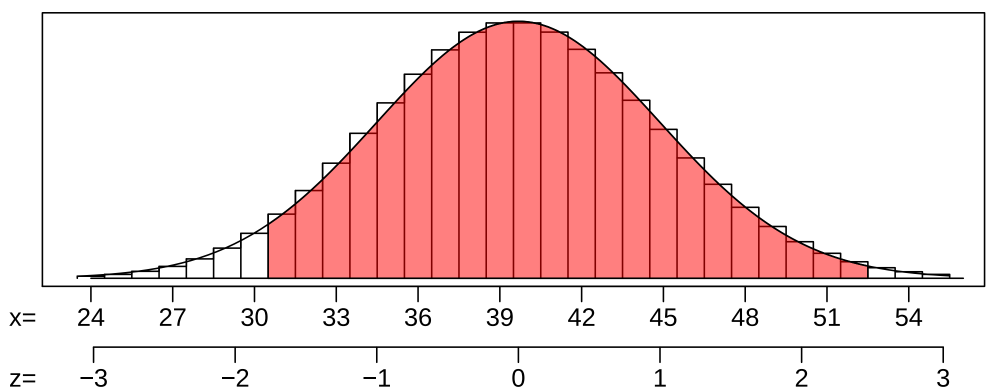
Find the mean.

$$\mu = np = (124)(0.32) = 39.68$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(124)(0.32)(1-0.32)} = 5.1945$$

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



Find the z scores.

$$z_1 = \frac{30.5 - 39.68}{5.1945} = -1.77$$

$$z_2 = \frac{52.5 - 39.68}{5.1945} = 2.47$$

Calculate the probability.

$$P(31 \leq X \leq 52) = \Phi(2.47) - \Phi(-1.77) = 0.9548$$

(a) $P(X = 35) = 0.0524$

(b) $P(31 \leq X \leq 52) = 0.9548$

5. As an ornithologist, you wish to determine the average body mass of *Cistothorus palustris*. You randomly sample 19 adults of *Cistothorus palustris*, resulting in a sample mean of 10.03 grams and a sample standard deviation of 1.63 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.

$$\begin{aligned} n &= 19 \\ \bar{x} &= 10.03 \\ s &= 1.63 \\ CL &= 0.98 \end{aligned}$$

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

$$df = n - 1 = 18$$

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

$$t^{\star} = 2.55$$

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.63}{\sqrt{19}} = 0.374$$

We want to make an inference about the population mean.

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

Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^{\star} SE,\ \bar{x} + t^{\star} SE) \\ &= (10.03 - 2.55 \times 0.374,\ 10.03 + 2.55 \times 0.374) \\ &= (9.08,\ 11) \end{aligned}$$

We are 98% confident that the population mean is between 9.08 and 11.

6. A treatment group of size 23 has a mean of 122 and standard deviation of 32. A control group of size 36 has a mean of 103 and standard deviation of 24.1. If you decided to use a signficance level of 0.01, 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 37.

   (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.

$$n_1 = 23$$
$$\bar{x}_1 = 122$$
$$s_1 = 32$$
$$n_2 = 36$$
$$\bar{x}_2 = 103$$
$$s_2 = 24.1$$
$$\alpha = 0.01$$
$$df = 37$$

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 $P(|T| > t^\star) = 0.01$ by using a t table.

$$t^\star = 2.72$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(32)^2}{23} + \frac{(24.1)^2}{36}} \\ &= 7.788 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(103 - 122) - (0)}{7.788} \\ &= -2.44 \end{aligned}$$

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.

$$0.01 < p\text{-value} < 0.02$$

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

(c) $t^{\star} = 2.72$

(d) $SE = 7.788$

(e) $|t_{\text{obs}}| = 2.44$

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

(g) retain the null

7. From a very large population, a random sample of 31000 individuals was taken. In that sample, 72.9% were shiny. 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.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.729)(1-0.729)}{31000}} = 0.00252$$

Calculate the margin of error.

$$ME = z^{\star}SE = (1.64)(0.00252) = 0.00413$$

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.

$$(0.725\,,\ 0.733)$$

We are 90% confident that the true population proportion is between 72.5% and 73.3%.

(a) The lower bound = 0.725, which can also be expressed as 72.5%.

(b) The upper bound = 0.733, which can also be expressed as 73.3%.

8. An experiment is run with a treatment group of size 40 and a control group of size 13. The results are summarized in the table below.

	treatment	control
pink	29	4
not pink	11	9

Using a significance level of 0.005, 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 $P(|Z| > z^\star) = 0.005$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 2.81$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{29}{40} = 0.725$$

$$\hat{p}_2 = \frac{4}{13} = 0.308$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.308 - 0.725 = -0.417$$

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

$$\hat{p} = \frac{29 + 4}{40 + 13} = 0.623$$

Determine the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}} \\ &= \sqrt{\frac{(0.623)(0.377)}{40} + \frac{(0.623)(0.377)}{13}} \\ &= 0.155 \end{aligned}$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0, 0.155)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.308 - 0.725) - 0}{0.155} \\ &= -2.69 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.69) \\ &= 0.0072 \end{aligned}$$

Compare the *p*-value to the signficance level.

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

So, we retain the null hypothesis. Thus the difference in proportions is not significant.

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

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

(c) $z^\star = 2.81$

(d) $SE = 0.155$

(e) $|z_{obs}| = 2.69$

(f) p-value = 0.0072

(g) retain the null