Bunker Hill Community College

Final Statistics Exam 2019-05-02 Exam ID 006

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 and yellow**) = 0.0158

   (b) P(**pig or red**) = 0.446

   (c) P(**red**) = 0.294

   (d) P(**cat given violet**) = 0.37

   (e) P(**tree**) = 0.216

   (f) P(**red given tree**) = 0.371

2. P(**"kite" given "red"**) = 0.321

3. $P(59.41 < X < 60.7) = 0.8914$

4. (a) $P(X = 36) = 0.0829$

   (b) $P(28 \leq X \leq 41) = 0.8644$

5. **(14.1, 15.8)**

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

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

   (c) $t^\star = 2$

   (d) $SE = 0.093$

   (e) $\mid t_{\textbf{obs}} \mid = 2.15$

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

   (g) **reject**

7. (a) **LB of p CI = 0.749 or** 74.9%

   (b) **UB of p CI = 0.775 or** 77.5%

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) $SE = 0.091$

   (e) $\mid z_{\mathbf{obs}} \mid = 2.16$

   (f) p**-value** $= 0.0308$

   (g) **reject**

1. In a deck of strange cards, there are 1073 cards. Each card has an image and a color. The amounts are shown in the table below.

	black	red	violet	yellow
bike	90	20	68	27
cat	42	87	97	17
horn	36	60	41	30
pig	48	62	18	98
tree	77	86	38	31

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

(b) What is the probability a random card is either a pig or red (or both)?

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

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

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

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

Solution

(a) $P(\text{cat and yellow}) = \frac{17}{1073} = 0.0158$

(b) $P(\text{pig or red}) = \frac{48+62+18+98+20+87+60+62+86-62}{1073} = 0.446$

(c) $P(\text{red}) = \frac{20+87+60+62+86}{1073} = 0.294$

(d) $P(\text{cat given violet}) = \frac{97}{68+97+41+18+38} = 0.37$

(e) $P(\text{tree}) = \frac{77+86+38+31}{1073} = 0.216$

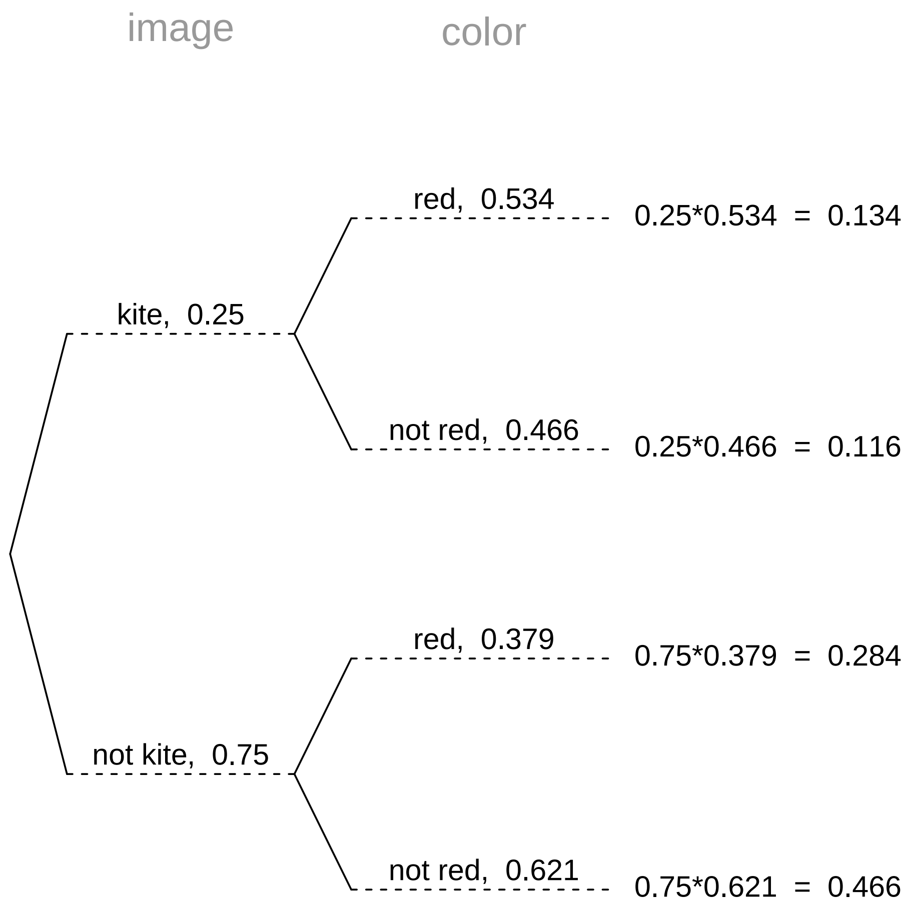
(f) $P(\text{red given tree}) = \frac{86}{77+86+38+31} = 0.371$

2. In a deck of strange cards, each card has an image and a color. The chance of drawing a kite is 25%. If a kite is drawn, there is a 53.4% chance that it is red. If a card that is not a kite is drawn, there is a 37.9% chance that it is red.

   Now, someone draws a random card and reveals it is red. 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.

$$P(\text{"kite" given "red"}) = \frac{0.134}{0.134 + 0.284} = 0.321$$

3. In a very large pile of toothpicks, the mean length is 60.11 millimeters and the standard deviation is 3.98 millimeters. If you randomly sample 100 toothpicks, what is the chance the sample mean is between 59.41 and 60.7 millimeters?

Solution

Label the given information.

$$\mu = 60.11$$
$$\sigma = 3.98$$
$$n = 100$$
$$\bar{x}_{\text{lower}} = 59.41$$
$$\bar{x}_{\text{upper}} = 60.7$$

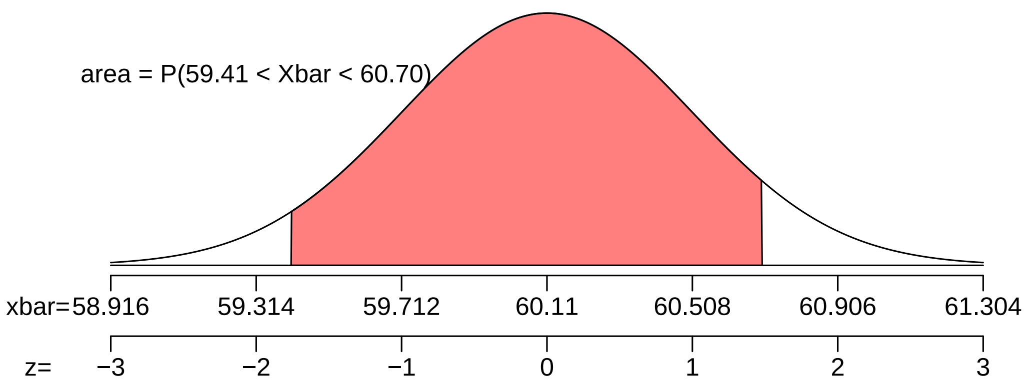
Find the standard error.

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{3.98}{\sqrt{100}} = 0.398$$

Describe the sampling distribution.

$$\bar{X} \sim \mathcal{N}(60.11,\ 0.398)$$

Draw a sketch.



Calculate a z scores.

$$z_{\text{lower}} = \frac{x_{\text{lower}} - \mu}{SE} = \frac{59.41 - 60.11}{0.398} = -1.76$$

$$z_{\text{upper}} = \frac{x_{\text{upper}} - \mu}{SE} = \frac{60.7 - 60.11}{0.398} = 1.48$$

Determine the probability.

$$\begin{aligned} P(59.41 < X < 60.7) &= \Phi(z_{\text{upper}}) - \Phi(z_{\text{lower}}) \\ &= \Phi(1.48) - \Phi(-1.76) \\ &= 0.8914 \end{aligned}$$

4. In a game, there is a 38% chance to win a round. You will play 92 rounds.

   (a) What is the probability of winning exactly 36 rounds?

   (b) What is the probability of winning at least 28 but at most 41 rounds?

Solution

We use the formula for binomial probabilities.

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

$$P(X = 36) = \binom{92}{36}(0.38)^{36}(1 - 0.38)^{92-36}$$

$$P(X = 36) = \binom{92}{36}(0.38)^{36}(0.62)^{56}$$

$$P(X = 36) = 0.0829$$

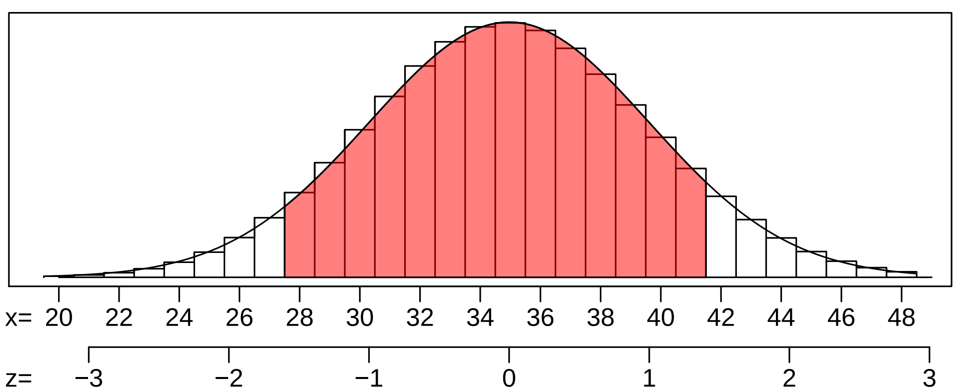
Find the mean.

$$\mu = np = (92)(0.38) = 34.96$$

Find the standard deviation.

$$\sigma = \sqrt{np(1 - p)} = \sqrt{(92)(0.38)(1 - 0.38)} = 4.6557$$

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



Find the z scores.

$$z_1 = \frac{27.5 - 34.96}{4.6557} = -1.6$$

$$z_2 = \frac{41.5 - 34.96}{4.6557} = 1.4$$

Calculate the probability.

$$P(28 \leq X \leq 41) = \Phi(1.4) - \Phi(-1.6) = 0.8644$$

(a) $P(X = 36) = 0.0829$

(b) $P(28 \leq X \leq 41) = 0.8644$

5. As an ornithologist, you wish to determine the average body mass of *Dendroica castanea*. You randomly sample 24 adults of *Dendroica castanea*, resulting in a sample mean of 14.95 grams and a sample standard deviation of 1.29 grams. Determine a 99.5% confidence interval of the true population mean.

Solution

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

$$\begin{aligned} n &= 24 \\ \bar{x} &= 14.95 \\ s &= 1.29 \\ CL &= 0.995 \end{aligned}$$

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

$$df = n - 1 = 23$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.995$.

$$t^\star = 3.1$$

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{1.29}{\sqrt{24}} = 0.263$$

We want to make an inference about the population mean.

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

Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (14.95 - 3.1 \times 0.263,\ 14.95 + 3.1 \times 0.263) \\ &= (14.1,\ 15.8) \end{aligned}$$

We are 99.5% confident that the population mean is between 14.1 and 15.8.

6. A treatment group of size 22 has a mean of 1.02 and standard deviation of 0.31. A control group of size 40 has a mean of 1.22 and standard deviation of 0.414. If you decided to use a signficance level of 0.05, 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 54.

   (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.

$$\begin{aligned} n_1 &= 22 \\ \bar{x}_1 &= 1.02 \\ s_1 &= 0.31 \\ n_2 &= 40 \\ \bar{x}_2 &= 1.22 \\ s_2 &= 0.414 \\ \alpha &= 0.05 \\ df &= 54 \end{aligned}$$

State the hypotheses.

$$\begin{aligned} H_0 &: \quad \mu_2 - \mu_1 = 0 \\ H_A &: \quad \mu_2 - \mu_1 \neq 0 \end{aligned}$$

We are using a two-tail test. Find $t^{\star}$ such that $P(|T| > t^{\star}) = 0.05$ by using a t table.

$$t^{\star} = 2$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \\ &= \sqrt{\frac{(0.31)^2}{22} + \frac{(0.414)^2}{40}} \\ &= 0.093 \end{aligned}$$

Determine the test statistic.

$$\begin{aligned} t_{\text{obs}} &= \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)_0}{SE} \\ &= \frac{(1.22 - 1.02) - (0)}{0.093} \\ &= 2.15 \end{aligned}$$

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 reject the null hypothesis.

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

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

(c) $t^\star = 2$

(d) $SE = 0.093$

(e) $|t_{\text{obs}}| = 2.15$

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

(g) reject the null

7. From a very large population, a random sample of 7900 individuals was taken. In that sample, 76.2% were shiny. 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.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.762)(1-0.762)}{7900}} = 0.00479$$

Calculate the margin of error.

$$ME = z^\star SE = (2.81)(0.00479) = 0.0135$$

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.

$$(0.749\,,\ 0.775)$$

We are 99.5% confident that the true population proportion is between 74.9% and 77.5%.

(a) The lower bound = 0.749, which can also be expressed as 74.9%.

(b) The upper bound = 0.775, which can also be expressed as 77.5%.

8. An experiment is run with a treatment group of size 39 and a control group of size 83. The results are summarized in the table below.

	treatment	control
omnivorous	18	22
not omnivorous	21	61

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

(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 $P(|Z| > z^\star) = 0.04$.

$$z^\star = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 2.05$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{18}{39} = 0.462$$
$$\hat{p}_2 = \frac{22}{83} = 0.265$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.265 - 0.462 = -0.197$$

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

$$\hat{p} = \frac{18 + 22}{39 + 83} = 0.328$$

Determine the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}} \\ &= \sqrt{\frac{(0.328)(0.672)}{39} + \frac{(0.328)(0.672)}{83}} \\ &= 0.0911 \end{aligned}$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0911)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.265 - 0.462) - 0}{0.0911} \\ &= -2.16 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.16) \\ &= 0.0308 \end{aligned}$$

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) $SE = 0.0911$
(e) $|z_{obs}| = 2.16$
(f) p-value = 0.0308
(g) reject the null