Note # 8: Statistical Inference for Numeric Variables

Problem 1. What does the phrase "95% confidence" in a confidence statement mean?

- **a.** The probability is 0.95 that a randomly chosen individual's value falls within the announced margin of error.
- **b.** 95% of the population falls within the announced margin of error.
- **c.** The results are true for 95% of the population.
- **d.** The results were obtained using a method that gives correct answers in 95% of all samples.

Answer:

d is the correct answer. Confidence intervals describe a population parameter, not individuals, samples, or statistics.

Problem 2. A survey asked the question "What do you think is the ideal number of children for a family to have?" The 519 females who responded had a median of 2, a mean of 3.02, and a standard deviation of 1.93. The 95% confidence interval is (2.85, 3.19). What is the best interpretation of this confidence interval?

- **a.** Ninety-five percent of females want between 2.85 and 3.19 children.
- **b.** We can be 95% confident that the proportion of females who want children is between 2.85 and 3.19.
- **c.** We can be 95% confident that a given female will want between 2.85 and 3.19 children.
- **d.** We can be 95% confident that the mean number of children that females would like to have is between 2.85 and 3.19.

Answer:

d is the correct answer.

i. Confidence level.
ii. Description of the population parameter.
iii. Confidence intervals.

Problem 3. The p-value for a two-sided hypothesis test of the null hypothesis $H_0: \mu = 12$ is 0.07. Which of the following confidence intervals would include the value 12? Select all that apply.

a. 90% Confidence Interval?

b. 95% Confidence Interval?

c. 99% Confidence Interval?

Answer:

b & c are correct answers.

Problem 4. The level of calcium in the blood of healthy young adults follows a normal distribution with $\mu = 10$ milligrams per deciliter and $\sigma = 4$. A clinic measures the blood calcium of 25 healthy pregnant young women at their first visit for prenatal care. The mean of these 25 measurements is $\bar{x} = 9.6$. We want to test the hypotheses $H_0: \mu = 10$; $H_0: \mu < 10$. What does it mean if the p-value is 0.0002?

- **a.** If the true population mean is less than 10, the probability that we get a sample mean of 9.6 is 0.0002.
- **b.** If the true population mean is less than 10, the probability that we get a sample mean of 9.6 or less is 0.0002.
- **c.** If the true population mean is 10, the probability that we get a sample mean of 9.6 is 0.0002.
- **d.** If the true population mean is 10, the probability that we get a sample mean of 9.6 or less is 0.0002.
- **e.** None of the above.

Answer:

d is the correct answer. $P(get\ \bar{x} = 9.6\ or\ less \mid \mu = 10) = 0.0002.$

Problem 5. A student organization at Wittenberg University has 15 freshmen, 18 sophomores, 14 juniors, and 12 seniors. Is there an equal distribution of the 4 classifications in this club?

a. What kind of hypothesis test should be conducted in this scenario?
b. What are the hypotheses?
c. What is the significance level?
d. What is the value of the test statistic?
e. What is the p-value?
f. What is the correct decision?
g. What is the appropriate conclusion/interpretation?
h. Are the assumptions met? Explain.

Answer:

a. $\chi^2 - Goodness\ of\ fit.$

b. $H_0: P_{FR} = P_{SO} = P_{JR} = P_{SR} = 0.25$ vs. $H_A: At\ least\ one\ proportion\ is\ different.$

c. $\alpha = 0.05$

d.

	FR	**SO**	**JR**	**SR**
Obs.	15	18	14	12
Exp.	(59) (0.25) = 14.75	(59) (0.25) = 14.75	(59) (0.25) = 14.75	(59) (0.25) =14.75

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(15-14.75)^2}{14.75} + \frac{(18-14.75)^2}{14.75} + \frac{(14-14.75)^2}{14.75} + \frac{(12-14.75)^2}{14.75}$$
$$= 0.00424 + 0.71610 + 0.03814 + 0.51271 = \mathbf{1.27119}$$

e. $\chi^2 = 1.27119, df = K - 1 = 4 - 1 = 3 \rightarrow \chi^2 < 3.66$

$p - value > 0.30.$

f. $p - value > 0.30 > 0.05 \rightarrow p - value > 0.05, Fail\ to\ reject\ H_0.$

g. The data does not provide statistically significant evidence that the distribution of classifications in this club is unequal.

h.

i. Independence? Maybe not.
ii. Expected counts ≥ 5? All excepted counts = 14.75
iii. $df > 1?\ df = 3$

Problem 6. A news article reports that "Americans have different views on two potentially inconvenient and invasive practices that airports could implement to uncover potential terrorist attacks." This news piece was based on a survey conducted among a random sample of 1,137 adults nationwide, where one of the questions on the survey was "Some airports are now using 'full-body' digital x-ray machines to electronically screen passengers in airport security lines. Do you think these new x-ray machines should or should not be used at airports?" Below is a summary of responses based on party affiliation. Conduct an appropriate hypothesis test to determine if there is a relationship between political affiliation and belief.

	Republican	**Democrat**	**Total**
Should	264	299	563
Should Not	38	55	93
Don't Know/ No Answer	16	15	31
Total	318	369	687

- **a.** What kind of hypothesis test should be conducted in this scenario?
- **b.** What are the hypotheses?
- **c.** What is the significance level?
- **d.** What is the value of the test statistic?
- **e.** What is the p-value?
- **f.** What is the correct decision?
- **g.** What is the appropriate conclusion/interpretation?

Answer:

- **a.** $\chi^2 - Test\ of\ Independence.$
- **b.** $H_0: Political\ party\ \&\ belief\ are\ independent.$

  $H_A: Political\ party\ \&\ belief\ are\ not\ independent.$
- **c.** $\alpha = 0.05$
- **d.** Excepted counts:

	R	**D**
S	$\frac{(563)(318)}{687} = 260.60$	$\frac{(563)(369)}{687} = 302.40$
SN	$\frac{(93)(318)}{687} = 43.05$	$\frac{(93)(369)}{687} = 49.95$
DN	$\frac{(31)(318)}{687} = 14.35$	$\frac{(31)(369)}{687} = 16.65$

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(264-260.60)^2}{260.60} + \frac{(299-302.40)^2}{302.40} + \frac{(38-43.05)^2}{43.05} + \frac{(55-49.95)^2}{49.95} + \frac{(16-14.35)^2}{14.35} + \frac{(15-16.65)^2}{16.65}$$

$$= 0.0444 + 0.0382 + 0.5924 + 0.5106 + 0.1897 + 0.1636 = \mathbf{1.5388}$$

e. $\chi^2 = 1.5388, df = (r-1)(c-1) = (3-1)(2-1) = 2 * 1 = 2 \rightarrow \chi^2 < 2.41$

$p - value > 0.30.$

f. $p - value > 0.30 > 0.05 \rightarrow p - value > 0.05, Fail\ to\ reject\ H_0.$

g. The data does not provide statistically significant evidence that political party and belief on this question are associated.

Problem 7. Nationally, about 66% of high school graduates enroll in higher education. This past year, out of the 200 graduating seniors at the high school in your hometown, 145 enrolled in higher education and 55 did not. You want to perform a statistical inference test to determine if the distribution is the same at the national level and at your school. Test this at the 10% level.

a. What kind of hypothesis test should be conducted in this scenario?

b. What are the hypotheses?

c. What is the significance level?

d. What are the expected counts?

e. What is the value of the test statistic?

f. What is the p-value?

g. What is the correct decision?

h. What is the appropriate conclusion/interpretation?

Answer:

a. $\chi^2 - Goodness\ of\ fit.$

b. H_0: The distribution of college enrollment at this high school follows the same distribution as college enrollment at the national level.

H_A: The distribution of college enrollment at this high school follows doesn't follow the same distribution as college enrollment at the national level.

c. $\alpha = 0.05$

d. Excepted counts:

	Enrolled	**Not enrolled**
Obs.	145	55
Exp.	$(200)(0.66) = 132$	$(200)(0.34) = 68$

e. $\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(145-132)^2}{132} + \frac{(55-68)^2}{68} = 1.280 + 2.485 = 3.765$

$\chi^2 = 3.765, df = k - 1 = 2 - 1 = 1 \rightarrow \ 2.71 < \chi^2 < 3.84$

f. $0.05 < p - value < 0.10 \rightarrow$

g. $p - value < 0.10 < \alpha = 0.10 \rightarrow Reject\ H_0$

h. The data provide statistically significant evidence that distribution of college enrollment for this high school is different from the distribution of college enrollment at the national level.

Problem 8. A researcher is interested in learning about cell phone ownership at a local high school, specifically for freshmen and seniors. They randomly select 500 students and ask them their class year (freshman or senior) and whether they own a cell phone (yes or no). The results are shown in the table below. The researcher wants to perform a statistical inference test to determine if there is an association between class year and cell phone ownership. Test this at the 5% level.

	Owns a Cell Phone	**Doesn't Own a Cell Phone**
Freshman	100	150
Senior	200	50

a. What kind of hypothesis test should be conducted in this scenario?
b. What are the hypotheses?
c. What is the significance level?
d. What are the expected counts?
e. What is the value of the test statistic?
f. What is the p-value?
g. What is the correct decision?
h. What is the appropriate conclusion/interpretation?

Answer:

a. $\chi^2 - Test\ of\ Independence.$

b. H_0: Class year and all phone ownership are independent.

H_A: Class year and all phone ownership are dependent.

c. $\alpha = 0.05$

d. Excepted counts:

	Cell Phone	**No Cell Phone**
FRESHMAN	$\frac{(250)(300)}{500} = 150$	$\frac{(250)(200)}{500} = 100$
Senior	$\frac{(250)(300)}{500} = 150$	$\frac{(250)(200)}{500} = 100$

e. $\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(100-150)^2}{150} + \frac{(150-100)^2}{100} + \frac{(200-150)^2}{150} + \frac{(150-100)^2}{100}$
$= 16.67 + 25 + 16.67 + 25 = \mathbf{83.34}$

f. $\chi^2 = 83.34, df = (r-1)(c-1) = (2-1)(2-1) = 1 * 1 = 1 \rightarrow \ \chi^2 > 10.83$

$p - value < 0.001.$

g. $p - value < 0.001 < 0.05 \;\; \rightarrow \; p - value < 0.05, Reject\ H_0.$

h. The data does provide statistically significant evidence that their class year and cell phone ownership are associated.

Problem 9. A 2013 poll in the State of California surveyed people about taxing sugar-sweetened beverages. The results are presented in the table below and are classified by race/ethnicity as well as a survey response. Does there appear to be a relationship between race/ethnicity and survey response? Test this at the 1% level.

	Favor	**Oppose**	**No Opinion**
White/Non-Hispanic	234	433	43
Hispanic	147	106	19
African American	24	41	6
Asian American	54	48	16

- **a.** What kind of hypothesis test should be conducted in this scenario?
- **b.** What are the hypotheses?
- **c.** What is the significance level?
- **d.** What are the expected counts?
- **e.** What is the value of the test statistic?
- **f.** What is the p-value?
- **g.** What is the correct decision?
- **h.** What is the appropriate conclusion/interpretation?

Answer:

- **a.** $\chi^2 - Test\ of\ Independence.$
- **b.** H_0: Race/ethnicity and survey response are independent.

  H_A: Race/ethnicity and survey response are associated.
- **c.** $\alpha = 0.01$
- **d.** Excepted counts:

	Favor	Oppose	No Opinion
White/Non-Hispanic	$\frac{(710)(459)}{1171} = 278.3$	$\frac{(710)(628)}{1171} = 380.8$	$\frac{(710)(84)}{1171} = 50.9$
Hispanic	$\frac{(272)(459)}{1171} = 106.6$	$\frac{(272)(628)}{1171} = 145.9$	$\frac{(272)(84)}{1171} = 19.5$
African American	$\frac{(71)(459)}{1171} = 27.8$	$\frac{(71)(628)}{1171} = 38.1$	$\frac{(71)(84)}{1171} = 5.1$
Asian American	$\frac{(118)(459)}{1171} = 46.3$	$\frac{(118)(628)}{1171} = 63.2$	$\frac{(118)(84)}{1171} = 8.5$

e. $\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(234-278.3)^2}{278.3} + \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots + \frac{(16-8.5)^2}{8.5}$
$= 7.05 + \cdots\cdots\cdots\cdots\cdots\cdots\cdots + 6.62 = \mathbf{54.13}$

f. $\chi^2 = 54.13, df = (r-1)(c-1) = (4-1)(3-1) = 3 * 2 = 6 \rightarrow \chi^2 > 22.46$
$p - value < 0.001.$

g. $p - value < 0.001 < 0.01 \rightarrow p - value < 0.01, Reject\ H_0.$

h. The data does provide statistically significant evidence that Race/ethnicity and survey response about taxing sugar-sweetened beverages are associated.