Homework 5 Solution

Statistical Inference, Fall 1401

1- **(15 – each 3)**

a. False, a confidence interval is constructed to estimate the population proportion, not the sample proportion.

b. True, because the boundaries of the confidence interval are calculated as followed:

$$\hat{p} - ME = 82\% - 2\% = 80\% \ , \hat{p} + ME = 82\% + 2\% = 84\%$$

So, we are 95% confident that between 80% and 84% of all Americans think it's the government's responsibility to promote equality between men and women.

c. True, by the definition of a confidence interval.

d. True, because the formulation for the ME and standard error are:

$$ME = z^{*}SE$$

$$SE_{\hat{\mathrm{p}}} = \sqrt{\frac{\hat{\mathrm{p}}(1-\hat{\mathrm{p}})}{\mathrm{n}}}$$

Based on the above formulas, quadrupling the sample size decreases the SE and ME by a factor of $\frac{1}{\sqrt{4}}$, in other words, If the sample size (n) is quadrupled, then $\sqrt{n}$ is doubled and thus the SE is halved. The ME is the product of the SE and z^* and thus if the SE is halved, then the ME is halved too.

e. True, the 95% CI is entirely above 50% which means according to the results obtained from the confidence interval, it can be said that because the confidence interval is more than 50%, then it can be claimed that there is sufficient evidence to conclude that a majority of Americans think it's the government's responsibility to promote equality between men and women.

2- **(11)**

1. Set the hypotheses: $H_0: p = 0.5$

   $H_A: p > 0.5$

2. Set the point estimate: $\hat{p} = 0.53$

3. Check conditions:

a. **Independence**: Because this is a simple random sample that includes fewer than 10% of the population, the observations are independent.

b. **Sample size/skew:** In a one-proportion hypothesis test, the success-failure condition is checked using the null proportion,

$$n\,p = n(1\text{-}\,p) \geq 10$$

With these conditions verified, the normal model may be applied to $\hat{p}$.

4. Next the standard error can be computed. The null value is used again here, because this is a hypothesis test for a single proportion.

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.5(0.5)}{450}} = 0.02357023$$

5. Based on the normal model, the test statistic can be computed as the Z score of the point estimate:

$$\mathrm{Z} = \frac{\hat{p}-p}{SE} \qquad \rightarrow \mathrm{Z} = \frac{0.53-0.5}{0.02357023} = 1.272792$$

6. The p-value, is 0.1015459. Because the p-value is larger than 0.05, we do not reject the null hypothesis, and we do not find convincing evidence to support the campaign manager's claim.

3- **(16)**

a. **(3)** Summary of the study:

	Yes	**No**	**Total**
Nevaripine	26	94	120
Lopinavir	10	110	120
Total	36	204	240

b. **(3)** Set the hypotheses:

$H_0: p_N - p_L = 0$, There is no difference in virologic failure rates between the Nevaripine and Lopinavir groups.

$H_A: p_N - p_L \neq 0$, There is some difference in virologic failure rates between the Nevaripine and Lopinavir groups.

c. **(10)** Check conditions:

1. **Independence**:
   - ✓ Random assignment was used, so the observations in each group are independent. If the patients in the study are representative of those in the general population (something impossible to check with the given information), then we can also confidently generalize the findings to the population.

2. **Sample size/skew**:
   - ✓ the success-failure condition, which we would check using the pooled proportion ($\hat{p}_{\text{pool}} = 0.15$), is satisfied.

   - ✓ $\hat{p}_{\text{pool}} = \frac{\#\text{ of successes1} + \#\text{ of successes2}}{n_1 + n_2} = \frac{26+10}{120+120} = 0.15$

   - ✓ $n_1(1-\hat{p}_{\text{pool}}), n_1(\hat{p}_{\text{pool}}), n_2(1-\hat{p}_{\text{pool}}), n_2(\hat{p}_{\text{pool}}) \geq 10$

3. Next the standard error can be computed:

$$SE = \sqrt{\frac{(\hat{p}_{\text{pool}}(1-\hat{p}_{\text{pool}})}{n_1} + \frac{(\hat{p}_{\text{pool}}(1-\hat{p}_{\text{pool}})}{n_2}}$$

$$SE = \sqrt{\frac{(0.15(1-0.15)}{120} + \frac{(0.15(1-0.15)}{120}} = 0.04609772$$

4. point estimate = $\hat{p}_N - \hat{p}_L = \frac{26}{120} - \frac{10}{120} = 0.1333334$

5. P-value?

Point estimate = 0.1333334

Null value = 0

SE = 0.04609772

$$Z = \frac{0.1333334-0}{0.04609772} = 2.892408$$

P-value= P($|Z|$ > 2.892408) =0.003823011

6. Since the p-value is low, we reject H_0. There is strong evidence of a difference in virologic failure rates between the Nevaripine and Lopinavir groups.

4- **(9)**

1. Check conditions:

a. Independence:

- ✓ First the conditions must be verified. Because each group is a simple random sample from less than 10% of the population, the observations are independent, both within the samples and between the samples.

b. Sample size/skew:

- ✓ The success-failure condition also holds for each sample::
- ✓ $n_1(P_1), n_1(1-P_1)$, n_2(1-P_2), $n_2(P_2)$ ≥ 10

2. Because all conditions are met, the normal model can be used for the point estimate of the difference in support, where P_1 corresponds to the original ordering and P_2, to the reversed ordering:

$$\text{point estimate} = \hat{p}_1 - \hat{p}_2 = 0.47 - 0.34 = 0.13$$

3. Next the standard error can be computed:

$$SE = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$SE = \sqrt{\frac{0.47(1-0.47)}{771} + \frac{0.34(1-0.34)}{732}} = 0.025$$

4. For a 90% confidence interval, we use z* = 1.65:

Point estimate ± z*SE = 0.13 ± 1.65*0.025 = (0.09,0.17)

5. We are 90% confident that the approval rating for the 2010 healthcare law changes between 9% and 17% due to the ordering of the two statements in the survey question.

5- **(9.5)**

Use a chi-squared goodness of fit test.

a. Set the hypotheses:

H_0: Each option is equally likely.
H_A: Some options are preferred over others.

b. Check conditions:

1. **Independence**: ok

2. Sample size/skew:

- ✓ The Total sample size: 99. Expected counts: $(\frac{1}{3})$ * 99 = 33 for each option. These are all above 5, so conditions are satisfied.

d. Df = 3 − 1 = 2

e. $\chi^2 = \sum_{i=1}^{k} \frac{(O-E)^2}{E}$

$$\chi^2 = \frac{(43-33)^2}{33} + \frac{(21-33)^2}{33} + \frac{(35-33)^2}{33} = 7.52 \rightarrow \text{p-value} = 0.023$$

Since the p-value is less than 5%, we reject H_0. The data provide convincing evidence that some options are preferred over others.

6- **(11.5)**

a. Set the hypotheses:

H_0: Gender and condiments are independent

H_A: Gender and condiments are not independent

b. We need to expected counts table :

$$\text{Expected count} = \frac{\text{(row total)} * \text{(column total)}}{\text{table total}}$$

$E_{\text{row 1, col 1}} = \frac{48*40}{100} = 19.2$, $E_{\text{row 1, col 2}} = \frac{48*42}{100} = 20.16, \ldots$

	Ketchup	**Mustard**	**Relish**	**Total**
Male	15 (19.2)	23 (20.16)	10 (8.64)	48
Female	25 (20.8)	19 (21.84)	8 (9.36)	52
Total	40	42	18	100

c. None of the expected counts in the table are less than 5. Therefore, we can proceed with the Chi-Square test.

d. The test statistic is:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O-E)^2}{E}$$

$$\chi^2 = \frac{(15-19.2)^2}{19.2} + \frac{(23-20.16)^2}{20.16} + \frac{(10-8.64)^2}{8.64} + \frac{(25-20.8)^2}{20.8} + \frac{(19-21.84)^2}{21.84} + \frac{(8-9.36)^2}{9.36} = 2.947891$$

e. Df = (2-1) × (3-1) = 2 → p-value = 0.2290201

f. With a p-value greater than 5%, fail to reject H_0, we can conclude that there is not enough evidence in the data to suggest that gender and preferred condiment are related.

7- **R(13)**

8- **R(15)**