Homework 5 Solution

Statistical Inference, Spring 2021

1-

a. The population parameter of interest is the proportion of graduates at a mid-sized university who found a job within one year of completing their undergraduate degree.

The value of the point estimate of this parameter is given by $\hat{P} = \frac{348}{40} = 0.87$

b. The conditions that must be met before constructing a confidence interval include:

The sample size is large.

$$npq = 400(0.87)(1 - 0.87) = 348(0.13) = 45.24 > 10$$

Since, npq > 10, the conditions for constructing confidence interval are met.

c. The 95% confidence interval for the proportion of graduates who found a job within one year of completing their undergraduate degree at this university is given by:

$95\%C.I. = \hat{p} \pm 1.95\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.87 \pm 0.03$

(0.84,0.90)

d. We are 95% confident that the proportion of graduates at a mid-sized university who found a job within one year of completing their undergraduate degree is between 0.84 and 0.90

e. The 99% confidence interval for the proportion of graduates who found a job within one year of completing their undergraduate degree at this university is given by:

$99\%C.I. = \hat{p} \pm 2.575\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.87 \pm 0.04$

(0.83 , 0.91)

f. The width of the 95% confidence interval is 0.06 while the width of the 99% confidence inteval is 0.08.The 99% confidence interval is wider.The wider the width of a confidence interval, the more confident we are that the required proportion is in the interval.

Homework 5 Solution

Statistical Inference, Spring 2021

2-

a. $25\%\ of\ them\ said\ they\ would\ rate\ their\ lives\ poorly\ enough\ to\ be\ con-sidered\ \text{“}suffering\text{”}$

b. Individuals are randomly sampled from the population.
The size of sample is 1000 which is smaller than 10% of the size of population.
Number of success =
25% × 1000 = 2500 > 10
Number of failures =
75% × 1000 = 7500 > 10
Hence the conditions are met.

c. $95\%C.I. = \hat{p} \pm 1.95\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.25 \pm 02535$

(0.224 , 0.275)

d. If the confidence level is increased which will in turn reduce the level of significance but increase the critical value and this will increase the margin of error which will make the confidence interval wider.

e. Looking at the formula for margin of error we see that if the sample size is increased the margin of error will reduce making the confidence level narrower.

3-

a. No.
Respondents only include those who took the SAT between April 25 and April 30 and are willing to respond. This means that several groups of people were excluded:
those who aren’t willing to respond (might be someone who aren’t interested in the topic), those who didn’t not take SAT (might be someone who took ACT or chose not to take a similar kind of exam), and those who took the exam during another period of time (students who take SAT in April tend to be Juniors, so the senior population should be relatively small).

b. $\hat{p} = 0.55$

$$90\%C.I. = \hat{p} \pm 1.95\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.55 \pm 0.0165$$

(0.5335 , 0.5665)

We are 90% confident that the true population of all high school seniors in the US who will participate in a study abroad program in college is between 53.36% and 56.64%.

c. Based on this interval, since both ends of 90% confidence interval are above 50%, it would be appropriate to claim that the majority of high school seniors are fairly certain that they will participate in a study abroad program in college.

4-

a. Null Hypothesis (H_0):Yawning is not contagious i.e. One person will not yawn by looking at other person yawn; the variables are independent

$$P_{ctrl} = P_{trtmnt}$$

Alternate Hypothesis ($\mathrm{H_a}$):Yawning is contagious i.e. One person will tend to yawn by looking at other person yawn. Or the two variables are dependent. Or proportion of yawns in the control group will be less than the proporition of yawns in the treatment group

$$P_{ctrl} < P_{trtmnt}$$

b.

P_{ctrl}= 10/34=0.29

P_{trtmnt}=4/16=0.25

$P_{ctrl} - P_{trtmnt} = 0.04$

c.

$$\hat{P}_{pool} = \frac{total\ successss}{total\ n} = \frac{14}{50} = 0.28$$

$$SE_{pool} = \sqrt{\hat{P}_{pool}(1 - \hat{P}_{pool})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \approx 0.14$$

$$Z = \frac{point\ estimte - null\ value}{SE} = 0.294$$

$$P_{value} = P(Z > 0.294) = 0.61$$

We fail to reject H0 since at a 5% significance level, there is no significant difference between the proportion of people yawning in the control and the treatment group and we conclude that yawning is not contagious.

توجه:در این سوال فرض کردیم شرایط تست برقرار است .

5- Expected values :

Wait time (minutes)	0-15	16-30	31-45	46+
workers	75	55	15	5
Expected :	0.5(150)	0.3(150)	0.1(150)	0.1(150)

a.

H0: The distribution of wait times this year is the same as last year.
Ha: The distribution of wait times this year is different from last year.

b.

Checking conditions:
Condition a. Independence: The workers are randomly sampled, n < 10% of population (150 < population of workers), each worker only falls in one category.
Condition b. Sample size/skew: Each category has at least 5 expected cases.

Chi-square test statistic formula

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

χ^2 test statistic

Wait time (minutes)	0-15	16-30	31-45	46+
$(\text{observed} - \text{expected})$	0	10	0	-10
$(\text{observed} - \text{expected})^2$	0	100	0	100
$\dfrac{(\text{observed} - \text{expected})^2}{\text{expected}}$	0	2.222	0	6.667

$$\chi^2 = 0 + 2.222 + 0 + 6.667$$

$$= 8.889$$

P-value: Since there are 4 categories (length of wait time), we use df = 4 – 1 = 3, so:

tail p	0.05	0.025	0.02
df			
2	5.991	7.378	7.824
3	7.815	9.348	9.837
4	9.488	11.14	11.67

Since our test statistic is between 7.815 and 9.348, the P-value is between 0.025 and 0.05, so the null hypotheses will be rejected.

6-

a. Expected values :

Chi-square test: Breed vs. toy color

	Yellow	Purple	Black	Total
Manx	16	30	14	60
	24	23	13	
Thai	27	23	10	60
	24	23	13	
Toyger	29	16	15	60
	24	23	13	
Total	72	69	39	180

b.

H0: breeds and the color that cats prefer for their toys are independent.
Ha: breeds and the color that cats prefer for their toys are dependent.

c. Calculate P-value :

$$\chi^2 = \frac{(16-24)^2}{24} + \ldots + \frac{(15-13)^2}{13}$$

$$\approx 2.667 + \ldots + 0.308$$

$$\approx 9.421$$

Since there are 3 row variables and 3 coloumn variables, we use df = (3-1)(3-1) = 4, so :

tail p	0.10	0.05	0.025
df			
3	6.251	7.815	9.348
4	7.779	9.488	11.140
5	9.236	11.070	12.830

Since our test statistic is between 7.779 and 9.448, P-value is between 0.05 and 0.10, so we fail to reject H0 and breeds of cats and the color that they prefer for their toys are independent.