STA101 Problem Set 6 - KEY

Summer I, 2021, Duke University

Exercises from the OpenIntro book - 100pts total

Problems include: Chapter 6 exercises 6.12, 6.19, 6.24, 6.41, 6.48 and 1 additional exercise

6.12 (20pts total)

(a) **(2pts for hypothesis, 3pts for conditions, 5pts for p-value, 5pts for correct conclusion)** The hypotheses are as follows:
$H_0 : p = 0.5$ (50% of Americans who decide not to go to college do so because they cannot afford it)
$H_1 : p < 0.5$ (Less than 50% of Americans who decide not to go to college do so because they cannot afford it)

Before calculating the test statistic we should check that the conditions are satisfied.

1. Independence: The sample is representative and we can safely assume that $331 < 10\%$ of all American adults who decide not to go to college, therefore whether or not one person in the sample decided not to go to college because they can't afford it is independent of another.
2. Success-failure: $331 \cdot 0.5 = 165.5 > 10$ and $331 \cdot 0.5 = 165.5 > 10$.

Since the observations are independent and the success-failure condition is met, $\widehat{p}$ is expected to be approximately normal. The test statistic can be calculated as follows:

$$\begin{aligned}
Z &= \frac{\widehat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \\
&= \frac{0.48 - 0.5}{\sqrt{\frac{0.5 \cdot 0.5}{331}}} \\
&= -0.73 \\
\text{p-value} &= \mathbb{P}(\widehat{p} < 0.48 \mid p = 0.5) \\
&= \mathbb{P}(Z < -0.73) \\
&= 0.2327
\end{aligned}$$

Since the p-value is large, we fail to reject H_0. The data do not provide strong evidence that less than half of American adults who decide not to go to college make this decision because they cannot afford college.

(b) **(3pts total for "yes", 2pts for explanation)** Yes, since we failed to reject $H_0 : p = 0.5$.

6.19 (14pts total)

(a) **(2pts for “False", 2pts for explanation)** False. Since $(p_{male} - p_{female})$ is positive, the proportion of males whose favorite color is black is higher than the proportion of females.

(b) **(2pts for “True")** True.

(c) **(2pts for “True")** True.

(d) **(2pts for “True")** True.

(e) **(2pts for “False", 2pts for explanation)** False. To get the 95% confidence interval for $(p_{female} - p_{male})$, all we have to is to swap the bounds of the confidence interval for $(p_{male} - p_{female})$ and take their negatives: (-0.06,-0.02).

6.24 (20pts total)

(a) **(2pts for hypothesis, 3pts for conditions, 5pts for p-value, 5pts for correct conclusion)** The hypotheses are: $H_0 : p_{CA} = p_{OR}$, $H_A : p_{CA} \neq p_{OR}$

1. Independence: Both samples are random, and 11,545 $<$ 10% of all Californians and 4,691 $<$ 10% of all Oregonians, therefore how much one Californian sleeps is independent of how much another Californian sleeps and how much one Oregonian sleeps is independent of how much another Oregonian sleeps. In addition, the two samples are independent of each other.
2. Success-failure:

$$\text{success}_{CA} = n_{CA} \cdot p_{CA} = (11,545)(0.08) = 923.6 \approx 924$$

$$\text{success}_{OR} = n_{OR} \cdot p_{OR} = (4,691)(0.088) = 412.8 \approx 413$$

$$\widehat{p}_{pool} = \frac{\text{success}_{CA} + \text{success}_{OR}}{n_{CA} + n_{OR}} = \frac{924 + 413}{11,545 + 4,691} = \frac{1,337}{16,236} \approx 0.0821$$

$$1 - \widehat{p}_{pool} = 0.918$$

$$(11,545)(0.082) = 946.69 > 10 \quad (11,545)(0.918) = 10,598.31 > 10$$

$$(4,691)(0.082) = 384.662 > 10 \quad (4,691)(0.918) = 4306.338 > 10$$

Since the observations are independent and the success-failure condition is met, $\widehat{p}_{CA} - \widehat{p}_{OR}$ is expected to be approximately normal. Next we calculate the test statistic and the p-value:

$$\begin{aligned} Z &= \frac{(\widehat{p}_{CA} - \widehat{p}_{OR}) - (p_{CA} - p_{OR})}{\sqrt{\frac{\widehat{p}_{pool}(1-\widehat{p}_{pool})}{n_{CA}} + \frac{\widehat{p}_{pool}(1-\widehat{p}_{pool})}{n_{OR}}}} \\ &= \frac{(0.08 - 0.088) - 0}{\sqrt{\frac{0.082 \cdot 0.918}{11,545} + \frac{0.082 \cdot 0.918}{4,691}}} \\ &= -1.68 \end{aligned}$$

p-value $= \mathbb{P}(|\widehat{p}_{CA} - \widehat{p}_{OR}| > 0.008 \mid (p_{CA} - p_{OR}) = 0) = 2\mathbb{P}(|Z| > 1.68) = 2(0.0465) = 0.093$

Since the p-value $> \alpha$ (use $\alpha = 0.05$ since not given), we fail to reject H_0 and conclude that the data do not provide strong evidence that the rate of sleep deprivation is different for the two states.

(b) **(5pts for “Type II")** Type II, since we may have incorrectly failed to reject H_0.

6.41 (12pts total)

(a) **(6pts)** The hypotheses are as follows:
H_0: There is no difference in rates of preferred shipping method and age among Los Angeles residents.
H_1: There is some difference in rates of preferred shipping method and age among Los Angeles residents.

(b) **(6pts for "conditions are not met")** The conditions are not satisfied since some expected counts are below 5.

6.48 (12pts total)

(a) **(2pt for "False", 1pt for explanation)** False. A confidence interval is constructed to estimate the population proportion, not the sample proportion.

(b) **(2pt for "True", 1pt for explanation)** True. This is the correct interpretation of the confidence interval, which can be calculated as $0.46 \pm 0.03 = (0.43, 0.49)$.

(c) **(2pt for "False", 1pt for explanation)** False. The confidence interval does not tell us what we might expect to see in another random sample.

(d) **(2pt for "False", 1pt for explanation)** False. As the confidence level decreases, the margin of error decreases as well.

Problem 6 (22pts total)

(a) **(2pts)** The hypotheses are as follows:
H_0: The distribution of the format of the book used by the students follows the professor's predictions.
H_A: The distribution of the format of the book used by the students does not follow the professor's predictions.

(b) **(4pts)**

$$E_{\text{buy}} = 150 \cdot 0.5 = 75 \quad E_{\text{print}} = 150 \cdot 0.3 = 45 \quad E_{\text{online}} = 150 \cdot 0.2 = 30$$

(c) **(4pts)**

1. Independence: The sample is not random. However, if the professor has reason to believe that the proportions are stable from one term to the next and students are not affecting each other's study habits, independence is probably reasonable.
2. Sample size: All expected counts are at least 5.

(d) **(4pts for χ^2, 2pts for df, and 2pts for p-value)**

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(80-75)^2}{75} + \frac{(40-45)^2}{45} + \frac{(30-30)^2}{30} = 0.889$$

$$df = 2$$

$$\text{p-value} = 0.641$$

(e) **(2pts for "fail to reject H_0, 2pts for interpretation)** Since the p-value is large, we fail to reject H_0. The data do not provide strong evidence indicating the professor's predictions were statistically inaccurate.