STAT 22000: Homework 13

Zixi Li

Problem 1 True or false about retaking road test

(a) is false. 66% of American adults in this sample think licensed drivers should be required to retake their road test once they turn 65, we are 100% confident.

(b) is true. Since we obtain 63% to 69% at the confidence interval 95%, and the sample size is large enough.

(c) is false. The confidence interval is for covering the population proportion, not for covering the proportion of another sample.

(d) is true. Because confidence interval 99% is an extension of 95%, margin of error will be larger.

Problem 2 Babies with low birth rate

(a)

With the following R codes, we know 111 babies had low birth weights, the percentage is 11.1%. A 95% confidence interval can be calculated as

$$\hat{p} \pm z^* \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.111 \pm 1.96\sqrt{\frac{0.111 \times (1-0.111)}{1000}} = (0.0915, 0.1305)$$

```
nc = read.csv("https://www.openintro.org/stat/data/csv/ncbirths.csv")
tally(~lowbirthweight, data=nc)
```

```
## lowbirthweight
##     low not low
##     111     889
```

(b)

Choosing p=0.1305, the sample size can be calculated as

$$n = (\frac{z^*}{m})^2 p(1-p) = (\frac{1.96}{0.015})^2 \times 0.1305 \times (1-0.1305) = 1937.36 = 1938$$

(c)

$H_0 : p = p_0$

$H_A : p > p_0$

Where $p_0 = 0.0817$ is the percentage of the babies born in the U.S. who had a low birth weight, p is the percentage of the babies born in North Carolina who had a low birth weight.

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.111 - 0.0817}{\sqrt{0.0817(1-0.0817)/1000}} = 3.3827$$

p-value is $P(z > 3.3827) = 0.00036 < \alpha = 0.01$, so we can reject H_0.

```
pnorm(3.3827,lower.tail = F)
```

```
## [1] 0.000358885
```

Problem 3 Compare smoking and non-smoking mothers

(a)

Here I make a two-way table cross classify the 1000 babies by whether they had low birth weights and whether the mother smoked using the R commands below

```
nc = read.csv("https://www.openintro.org/stat/data/csv/ncbirths.csv")
tally(~habit+lowbirthweight, data=nc)
```

```
##            lowbirthweight
## habit       low not low
##   nonsmoker  92     781
##   smoker     18     108
##   <NA>        1       0
```

(b)

$H_0 : p_1 = p_2$

$H_A : p_1 > p_2$

Where p_1 is the percentage of babies born to smoking mothers having low birth weights, p_2 is the percentage of babies born to non-smoking mothers having low birth weights.

From the two-way table we can get the pooled estimate of p

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{92 + 18}{92 + 781 + 18 + 108} = 0.1101$$

So the z-statistic for testing $H_0 : p_1 = p_2$ is

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{\frac{18}{18+108} - \frac{92}{92+781}}{\sqrt{0.1101(1-0.1101)(\frac{1}{18+108} + \frac{1}{92+781})}} = 1.256$$

p-value is $P(z > 1.256) = 0.1046 > \alpha = 0.05$, so we fail to reject H_0.

```
pnorm(1.256, lower.tail = F)
```

```
## [1] 0.104558
```

Problem 4 Survey about the daily show

(a)

The 99% CI can be calculated as

$$\hat{p}_1 - \hat{p}_2 \pm z^* \sqrt{\frac{\hat{p}_1\,(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2\,(1-\hat{p}_2)}{n_2}} = 0.33 - 0.22 \pm 2.576 \sqrt{\frac{0.33 * 0.67}{1099} + \frac{0.22 * 0.78}{1110}} = (0.06, 0.16)$$

(b)

(i) is true, because 99% confidence interval suggests $p_1 - p_2$ will be larger than 0, it's a strong evidence.

(ii) is false, because the confidence interval is for covering the population proportion difference, not for covering the proportion difference of another sample.

(iii) is false, because confidence interval 99% is an extension of 95%, 95% will be narrower.

(iv) is true, because the value $p_2 - p_1 \pm z^* SE$ is just a minus of the result in (a).