Homework №17

Author: David Oniani
Instructor: Dr. Eric Westlund

April 14, 2019

23.1 We start with the **PLAN** part as the **STATE** part is the description of the problem itself.

PLAN
Let p_{F} be the proportion of females who have used internet to search for health information and p_{M} be the the proportion of males who have used internet to search for health. Then we must find a 95% confidence interval for the difference in these proportions.

SOLVE

We have, $\hat{p_{\mathrm{F}}} = \dfrac{811}{1308} \approx 0.6200$ and $\hat{p_{\mathrm{M}}} = \dfrac{520}{1084} \approx 0.4797$. Then the standard error (SE) is

$$\mathrm{SE} = \sqrt{\frac{0.62 \times (1 - 0.62)}{1308} + \frac{0.4797 \times (1 - 0.4797)}{1084}} \approx 0.0203.$$

Now, from table C, we know that z^* value for 95% is 1.96 and the corresponding confidence interval is $(0.6200 - 0.4797) \pm 1.96 \times 0.0203$ which is (approximately) the confidence interval 0.1403 ± 0.0398. This confidence interval is approximately from 10% to 18% (or more precisely, from 10.05% to 18.01%).

CONCLUDE
We can be 95% confident that from 10% to 18% more women have looked for health information on the internet (than men).

23.2 We start with the **PLAN** part as the **STATE** part is the description of the problem itself.

PLAN
Let p_{N} be the proportion of ninth-graders who answered "yes" on the question whether they smoke marijuana or not. Let p_{T} be the proportion of twelfth-graders who answered "yes" on the question whether they smoke marijuana or not. Then we must find a 99% confidence interval for the difference in these proportions.

SOLVE

We have, $\hat{p_{\mathrm{N}}} = \dfrac{620}{3500} \approx 0.1771$ and $\hat{p_{\mathrm{T}}} = \dfrac{969}{3497} \approx 0.2771$. Then the standard error

(SE) is

$$\text{SE} = \sqrt{\frac{0.1771 \times (1 - 1771)}{3500} + \frac{0.2771 \times (1 - 0.2771)}{3497}} \approx 0.01.$$

Now, from table C, we know that z^* value for 99% is 2.576 and the corresponding confidence interval is $(0.1771 - 0.2771) \pm 2.576 \times 0.01$ which is the confidence interval -0.1 ± 0.02576. This confidence interval is approximately from -12.6% to -7.4% (or more precisely, from -12.576% to -7.424%).

CONCLUDE

We can be 99% confident that the difference between the proportions of all ninth-graders and twelfth graders who smoked marijuana at least once in the 30 days prior to the survey is between -12.6% and -7.4%.

23.4 (a) This is an observational study. It is an observational study as people were not assigned to a city.

(b) We start with the **PLAN** part as the **STATE** part is the description of the problem itself.

PLAN

Let p_{N} be the proportion of the belted drivers in New York. Let p_{B} be the proportion of the belted drivers in Boston. Then we must find a 95% confidence interval for the difference in these proportions.

SOLVE

We have, $\hat{p_{\text{N}}} = \frac{183}{220} \approx 0.832$ and $\hat{p_{\text{B}}} = \frac{68}{117} \approx 0.581$. Then the standard error (SE) is

$$\text{SE} = \sqrt{\frac{0.832 \times (1 - 0.832)}{220} + \frac{0.581 \times (1 - 0.581)}{117}} \approx 0.052.$$

Now, from table C, we know that z^* value for 95% is 1.96 and the corresponding confidence interval is $(0.832 - 0.581) \pm 1.96 \times 0.52$ which is the confidence interval 0.251 ± 0.102. This confidence interval is from 14.9% to 35.3%.

CONCLUDE

According to our results, we have a good evidence that a smaller proportion of female drivers wear seat belts in Boston than in New York.

23.5 We start with the **PLAN** part as the **STATE** part is the description of the problem itself.

PLAN

Let p_{SK} be the proportion of the injured skiers who wear helmets. Let p_{SN} be the proportion of the injured snowboarders who wear helmets. Then our null hypothesis is $\mathrm{H}_0 : p_{\mathrm{SK}} = p_{\mathrm{SN}}$ and the alternative hypothesis is $\mathrm{H}_{\mathrm{a}} : p_{\mathrm{SK}} < p_{\mathrm{SN}}$.

SOLVE

$p_{\mathrm{SK}} = \dfrac{96}{578} \approx 0.1661$. $p_{\mathrm{SN}} = \dfrac{656}{2992} \approx 0.1661$. $\hat{p} = \dfrac{96+656}{578+2992} \approx 0.2106$. Then, the standard error is

$$\mathrm{SE} = \sqrt{0.2106 \times (1 - 0.2106) \times \left(\frac{1}{578} + \frac{1}{2992}\right)} \approx 0.01853.$$

Now, $z = \dfrac{0.1661 - 0.2193}{0.01851} = -2.87$ and therefore, $P = 0.0021$.

CONCLUDE

Since $P = 0.0021 < 0.01$, we have a strong evidence that skiers and snowboarders with head injuries are less liekly to use helmets than those without head injuries.

23.18 (a) Let p_{S} be the proportion of subjects experiencing the primary outcome for the sibutramine group. Let p_{P} be the proportion of subjects experiencing the primary outcome for the placebo group. Then, we have that

$$p_{\mathrm{S}} = \frac{561}{4906} \approx 0.1143.$$

$$p_{\mathrm{P}} = \frac{490}{4898} \approx 0.1000.$$

(b) It is appropriate to use the large-sample confidence interval for comparing the proportions of sibutramine and placebo subjects who experienced the primary outcome. This is since both the number of successes and the number of failures are greater than or equal to 10.

(c) For 95%, our z-value is 1.96. Thus, the confidence interval is

$$(0.1143 - 0.1000) \pm 1.96 \times \sqrt{\frac{0.1143 \times (1 - 0.1143)}{4906} + \frac{0.1000 \times (1 - 0.1000)}{4898}}.$$

which is approximately a confidence interval from 0.0021 to 0.0265.

23.20 (a) We start with the **PLAN** part as the **STATE** part is the description of the problem itself.

STATE THE HYPOTHESIS

Let p_{T} be the proportion of treatment group. Let p_{P} be the proportion of the placebo group. Then our null hypothesis is $\mathrm{H}_0 : p_{\mathrm{T}} = p_{\mathrm{P}}$ and the alternative hypothesis is $\mathrm{H_a} : p_{\mathrm{T}} \neq p_{\mathrm{P}}$.

FIND THE TEST STATISTIC

$p_{\mathrm{T}} = \dfrac{561}{4906} \approx 0.1143$. $p_{\mathrm{P}} = \dfrac{490}{4898} \approx 0.1000$. $p = \dfrac{561+490}{4906+4898} \approx 0.1072$. Then, the test statistic is

$$z = \frac{0.1143 - 0.1000}{\sqrt{0.1072 \times (1-0.1072)} \times \sqrt{\dfrac{1}{4906} + \dfrac{1}{4898}}} \approx 2.29.$$

FIND THE P-VALUE

Now, using Table A, we get that our P-value is $P = 2 \times 0.011 = 0.022$.

STATE THE CONCLUSION

Since $P = 0.022 < 0.05$, we reject the null hypothesis H_0 and support the claim of the difference in proportions.

(b) This is important since the improvements could possible be due to the "placebo effect" and if no placebo groups is present, we do not know for sure whether the treatment is effective or not.

20.35 (a) This is an experiment since the subjects were assigned to the groups.

(b) Let $p_{\mathrm{HL+}}$ be the proportion for the HL+ group. Let p_{C} be the proportion for the control group. Then our null hypothesis is $\mathrm{H}_0 : p_{\mathrm{HL+}} = p_{\mathrm{C}}$ and the alternative hypothesis is $\mathrm{H_a} : p_{\mathrm{HL+}} < p_{\mathrm{C}}$.

Now, $\hat{p_{\mathrm{HL+}}} = \dfrac{49}{49+67} \approx 0.4224$, $\hat{p_{\mathrm{C}}} = \dfrac{49}{49+47} \approx 0.5104$, and $\hat{p} = \dfrac{49+49}{116+96} \approx 0.4623$.

From these values, we get that the standard error is

$$\mathrm{SE} = \sqrt{0.4623 \times (1-0.4623) \times (\frac{1}{116} + \frac{1}{96})} \approx 0.0688.$$

Then $z = \dfrac{0.4224 - 0.5104}{0.0688} \approx -1.28$ and $P = 0.1003$. At last, since $P = 0.1003 > 0.05$ we cannot reject the null hypothesis H_0 and therefore, we accept the null hypothesis

H_0. In other words, there is a weak/little evidence that to support the claim that the proportion of HL+ users with a rhinovirus infection is less than the proportion of non-HL+ users with a rhinovirus infection.