STAT 1293 - Assignment 5

Gordon Lu

7/26/2020

Problem 1: Gastric bypass surgery (15 points)

1a) Are the conditions for the use of the large sample procedures met? Explain. (2 points)

Solution:

Yes. First observing that our data comes from a SRS, we can assume the outcomes are unbiased and random. Additionally, notice how each individual's response can be summarized with a binary outcome. That being, either "Receiving gastic bypass surgery and maintaining at least a 20% loss six years after surgery" or "Receiving gastic bypass surgery and not maintaining at least a 20% loss six years after surgery". Then, note how the response of one individual does not impact the response of another, thus individuals are independent. Lastly, note that `np` $\geq 10 = 418(0.76) \approx 318$ and $n(1-p) \geq 10 = 418(0.24) \approx 100$. Thus, we can conclude that the sample size if sufficiently large, and importantly that the sampling distribution of p is normally distributed.

1b) Give a 90% confidence interval for the proportion of those receiving gastric bypass surgery that maintained at least a 20% weight loss six years after surgery. (4 points)

Solution:

```r
np <- 0.76*418
prop.test(np, 418, conf.level = 0.90, correct = F)$conf.int
```

```
## [1] 0.7240379 0.7926179
## attr(,"conf.level")
## [1] 0.9
```

1c) Does the study provide sufficient evidence to claim that more than 70% of those receiving gastric bypass surgery maintained at least a 20% weight loss six years after surgery? Use $\alpha = 0.05$ and choose correct=F. (9 points)

Solution:

```r
prop.test(np, 418, 0.70, conf.level = 0.95, correct = F, alternative = "greater")
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  np out of 418, null probability 0.7
## X-squared = 7.1657, df = 1, p-value = 0.003716
## alternative hypothesis: true p is greater than 0.7
## 95 percent confidence interval:
##  0.7240379 1.0000000
## sample estimates:
##    p
## 0.76
```

Conclusion: Yes, since $p < 0.05$, there is sufficient evidence to claim that more than 70% of those receiving gastic bypass surgery maintained at least a 20% weight loss six years after surgery.

Problem 2: I refuse! (15 points)

2a) What are the sample proportions of offers being rejected for the two groups? Are they similar? (6 points)

Solution:

```
human_reject <- 18/38
computer_reject <- 6/38
human_reject; computer_reject
```

```
## [1] 0.4736842
```

```
## [1] 0.1578947
```

The sample propertions of offers being rejected for the human group is: $\frac{18}{38} \approx 0.47$ and for the computer group is: $\frac{6}{38} \approx 0.16$. So, it appears that the sample proportions are not similar.

2b) We suspect that emotion will lead to offers from another person being rejected more often than offers from an impersonal computer. Do a test to assess the evidence for this conjecture. Use \$alpha = 0.01\$ and correct=T. (9 points)

Solution:

```
prop.test(c(18, 6), c(38,38), conf.level = 0.99, correct = T, alternative = "greater")
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(18, 6) out of c(38, 38)
## X-squared = 7.3686, df = 1, p-value = 0.003319
## alternative hypothesis: greater
## 99 percent confidence interval:
```

```
## 0.05614464 1.00000000
## sample estimates:
##    prop 1    prop 2
## 0.4736842 0.1578947
```

Decision: Since the p-value is 0.003319, we reject the null at the significance level `0.01`, and conclude that there is sufficient evidence to conclude that offers from another person is significantly larger than from an impersonal computer.

Conclusion: In other words, we have sufficient evidence to conclude that emotion will lead to offers from another person being rejected more than offers from an impersonal computer.

Problem 3: Where do young people live? (30 points)

3a) Based on the data, is there evidence that the proportions of living at home among the four age groups are different? Conduct a chi-squared test at $\alpha = 0.05$. (12 points)

Solution:

```r
matA <- matrix(c(324, 216, 378, 388, 337, 464, 318, 559), 2, 4)
rownames(matA) <- c("Parents' home", "Other places")
colnames(matA) <- c(19, 20, 21, 22)
qchisq(0.95, 3)
```

```
## [1] 7.814728
```

```r
chisq.test(matA, correct = F)
```

```
##
##  Pearson's Chi-squared test
##
## data:  matA
## X-squared = 84.354, df = 3, p-value < 2.2e-16
```

Decision: Since the $\chi^2 > \chi^2_{3,0.95} \to 84.354 > 7.815$ we have sufficient evidence to reject the null hypothesis.

Conclusion: Therefore, we have sufficient evidence to conclude that the proportions of living at home among the four given age groups are different.

3b) Conduct a trend test to see if there is a linear trend in the proportions of living at home. (10 points)

Solution:

```r
live_total <- margin.table(matA, 2)
home <- matA["Parents' home", ]
prop.trend.test(home, live_total)
```

```
##
##  Chi-squared Test for Trend in Proportions
##
## data:  home out of live_total ,
##  using scores: 1 2 3 4
## X-squared = 82.63, df = 1, p-value < 2.2e-16
```

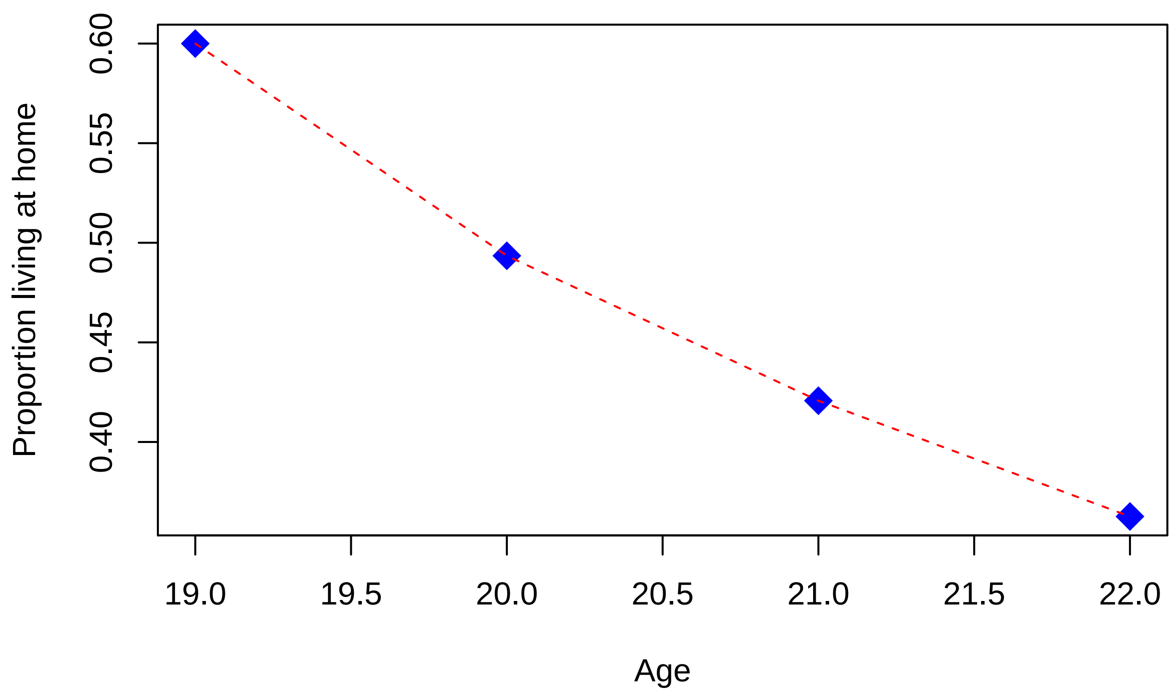
Decision: Since the p-value is very small, p-value < 2.2e-16, we have sufficient evidence to reject the null hypothesis.

Conclusion: Therefore, we have sufficient evidence to conclude that there is a linear trend in the proportions of living at home.

3c) Create a plot the show the relationship between proportion of livng at home and age. (8 points)

Solution:

```
home_p <- home/live_total
plot(colnames(matA), home_p, pch = 18, cex = 2, col = 4, xlab = "Age", ylab = "Proportion living at home
lines(colnames(matA), home_p, lty = 2, col = 2)
```



It appears that the relationship between the proprotion of living at home and age is linear.