Homework - Week 10

Solution

Questions marked with "(OS3: X.X)" are from the textbook with "X.X" as the exercise number. The answers to the odd questions (odd by book numbering that is) will be in the back of the book.

1. (OS3: 6.15) A 2012 survey of 2,254 American adults indicates that 17% of cell phone owners do their browsing on their phone rather than a computer or other device.

    a. According to an online article, a report from a mobile research company indicates that 38 percent of Chinese mobile web users only access the internet through their cell phones. Conduct a hypothesis test to determine if these data provide strong evidence that the proportion of Americans who only use their cell phones to access the internet is different than the Chinese proportion of 38%.

    ```
    # We need the number of "successes"
    n.american <- round(2254 * 0.17)
    n.american
    ```

    ```
    ## [1] 383
    ```

    ```
    cell.test <- prop.test(n.american, 2254, p=0.38)
    cell.test
    ```

    ```
    ##
    ##  1-sample proportions test with continuity correction
    ##
    ## data:  n.american out of 2254, null probability 0.38
    ## X-squared = 421.34, df = 1, p-value < 2.2e-16
    ## alternative hypothesis: true p is not equal to 0.38
    ## 95 percent confidence interval:
    ##  0.1547670 0.1862125
    ## sample estimates:
    ##         p
    ## 0.1699201
    ```

    b. Interpret the p-value in this context.

    $p < 0.0001$. If the true proportion of Americans who only access internet through cell phones is 0.38, he probabability of obtaining a sample like we did, or one more extreme, is very small (≈ 0).

    c. Calculate a 95% confidence interval for the proportion of Americans who access the internet on their cell phones, and interpret the interval in this context.

    95% confidence interval: (0.155, 0.186)

    We are 95% confident that the true proportion of Americans who only access the internet through cell phones is between 15.5% and 18.6%.

2. (OS3: 6.29) A 2010 survey asked 827 randomly sampled registered voters in California "Do you support? Or do you oppose? Drilling for oil and natural gas off the Coast of California? Or do you not know enough to say?" Below is the distribution of responses, separated based on whether or not the respondent graduated from college.

	College	*Grad*
	Yes	No
Support	154	132
Oppose	180	126
Do not know	104	131
Total	438	389

a. What percent of college graduates and what percent of the non-college graduates in this sample do not know enough to have an opinion on drilling for oil and natural gas off the Coast of California?

College grad: (104/438)*100 = 23.7%

Non-college grad: (131/389)*100 = 33.7%

b. Conduct a hypothesis test to determine if the data provide strong evidence that the proportion of college graduates who do not have an opinion on this issue is different than that of non-college graduates.

```
prop.test(c(104, 131), c(438, 389))
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  c(104, 131) out of c(438, 389)
## X-squared = 9.5084, df = 1, p-value = 0.002045
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.16333768 -0.03529832
## sample estimates:
##    prop 1    prop 2
## 0.2374429 0.3367609
```

p = 0.02 $<$ α = 0.05. Reject H_0. There is evidence that the rates of people who do not know their opinions on iol drilling off the California coast is different for college grads and non-college grads.

3. (OS3: 6.30) Exercise 6.28 provides data on sleep deprivation rates of Californians and Oregonians. The proportion of California residents who reported insufficient rest or sleep during each of the preceding 30 days is 8.0%, while this proportion is 8.8% for Oregon residents. These data are based on simple random samples of 11,545 California and 4,691 Oregon residents.

a. Conduct a hypothesis test to determine if these data provide strong evidence the rate of sleep deprivation is different for the two states. (Reminder: Check conditions.)

```
n.ca <- round(0.08 * 11545)
n.or <- round(0.088 * 4691)

prop.test(c(n.ca, n.or), c(11545, 4691))
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  c(n.ca, n.or) out of c(11545, 4691)
## X-squared = 2.7246, df = 1, p-value = 0.09882
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.017656096  0.001643531
```

```
## sample estimates:
##     prop 1     prop 2
## 0.08003465 0.08804093
```

p = 0.099 > α = 0.05. Do not reject H_0. There is not evidence that the rates of people who report insufficient rest is different for California residents and Oregon residents.

b. It is possible the conclusion of the test in part (a) is incorrect. If this is the case, what type of error was made?

Since we failed to reject the null hypothesis, it is possible we have a false negative. That is, it is possible that there is a true difference in proportions that we failed to detect. This is a type II error.

4. (OS3: 6.43) Rock-paper-scissors is a hand game played by two or more people where players choose to sign either rock, paper, or scissors with their hands. For your statistics class project, you want to evaluate whether players choose between these three options randomly, or if certain options are favored above others. You ask two friends to play rock-paper-scissors and count the times each option is played. The following table summarizes the data:

Rock	Paper	Scissors
43	21	35

Use these data to evaluate whether players choose between these three options randomly, or if certain options are favored above others. Make sure to clearly outline each step of your analysis, and interpret your results in context of the data and the research question.

If players choose the positions randomly, we would expect an equal probability for each option. We will conduct a goodness-of-fit test against a uniform distribuion.

$H_0 : p_r = p_p = p_s$

H_a : At least one probability differs from the others.

```r
chisq.test(c(43, 21, 35))
```

```
##
##  Chi-squared test for given probabilities
##
## data:  c(43, 21, 35)
## X-squared = 7.5152, df = 2, p-value = 0.02334
```

p = 0.023 < 0.05. Reject the null hypothesis. There is evidence that the option choices do not follow a uniform distribution.

5. (OS3: 6.44) Microhabitat factors associated with forage and bed sites of barking deer in Hainan Island, China were examined from 2001 to 2002. In this region woods make up 4.8% of the land, cultivated grass plot makes up 14.7%, and deciduous forests makes up 39.6%. Of the 426 sites where the deer forage, 4 were categorized as woods, 16 as cultivated grassplot, and 61 as deciduous forests. The table below summarizes these data.

Woods	Cultivated grassplot	Deciduous forests	Other	Total
4	16	61	345	426

a. Write the hypotheses for testing if barking deer prefer to forage in certain habitats over others.

$H_0 : p_w = 0.048,\ p_g = 0.147,\ p_f = 0.396,\ p_o = 0.409(1 - p_w - p_g - p_f)$

H_a : At least one probability differs from the expected probability.

b. What type of test can we use to answer this research question?

Chi-sqaure goodness-of-fit test.

c. Check if the assumptions and conditions required for this test are satisfied.

The smallest expected value will be woods because it has the smallest proportion of land. E(woods) = 0.048 * 426 = 20.448. Since this is greater than 5, we can use this test.

d. Do these data provide convincing evidence that barking deer prefer to forage in certain habitats over others? Conduct an appropriate hypothesis test to answer this research question.

```
chisq.test(c(4, 16, 61, 345), p=c(0.048, 0.147, 0.396, 0.409))
```

```
##
##  Chi-squared test for given probabilities
##
## data:  c(4, 16, 61, 345)
## X-squared = 284.06, df = 3, p-value < 2.2e-16
```

p $<$ 0.0001 $<$ 0.05. Reject the null hypothesis. There is evidence that barking deer prefer to forage in certain habitats over others.

6. (OS3: 6.46) The table below summarizes a data set we first encountered in Exercise 6.32 regarding views on full-body scans and political affiliation. The differences in each political group may be due to chance. Complete the following computations under the null hypothesis of independence between an individual's party affiliation and his support of full-body scans. It may be useful to first add on an extra column for row totals before proceeding with the computations.

		Party Affiliation		
		Republican	Democrat	Independent
Answer	Should	264	299	351
	Should not	38	55	77
	Don't know/No answer	16	15	22
	Total	318	369	450

```
# Create matrix of data and add marginal totals
bs.mat <- matrix(c(264, 38, 16, 299, 55, 15, 351, 77, 22), nrow=3,
                 dimnames = list(ans=c("Should", "Should not", "Don't know"),
                                 party=c("Rep", "Dem", "Ind")))
bs.mat <- addmargins(bs.mat)
bs.mat
```

```
##               party
## ans            Rep Dem Ind  Sum
##   Should       264 299 351  914
##   Should not    38  55  77  170
##   Don't know    16  15  22   53
##   Sum          318 369 450 1137
```

a. How many Republicans would you expect to not support the use of full-body scans?

```
# Expected value = # rep (row 4 col 1) * # should not (row 2 col 4) / # total (row 4 col 4)
exp.rep.sn <- bs.mat[4,1] * bs.mat[2,4] / bs.mat[4,4]
exp.rep.sn
```

```
## [1] 47.54617
```

b. How many Democrats would you expect to support the use of full-body scans?

```r
# Expected value = # dem (row 4 col 2) * # should (row 1 col 4) / # total (row 4 col 4)
exp.dem.s <- bs.mat[4,2] * bs.mat[1,4] / bs.mat[4,4]
exp.dem.s
```

```
## [1] 296.628
```

c. How many Independents would you expect to not know or not answer?

```r
# Expected value = # ind (row 4 col 3) * # don't know (row 3 col 4) / # total (row 4 col 4)
exp.ind.dk <- bs.mat[4,3] * bs.mat[3,4] / bs.mat[4,4]
exp.ind.dk
```

```
## [1] 20.97625
```

7. (OS3: 6.47) The table below summarizes a data set we first encountered in Exercise 6.29 that examines the responses of a random sample of college graduates and non-graduates on the topic of oil drilling. Complete a chi-square test for these data to check whether there is a statistically significant difference in responses from college graduates and non-graduates.

	College Grad	
	Yes	No
Support	154	132
Oppose	180	126
Do not know	104	131
Total	438	389

```r
oil.mat <- matrix(c(154, 180, 104, 132, 126, 131), nrow=3)
oil.mat
```

```
##      [,1] [,2]
## [1,]  154  132
## [2,]  180  126
## [3,]  104  131
```

```r
chisq.test(oil.mat)
```

```
##
##  Pearson's Chi-squared test
##
## data:  oil.mat
## X-squared = 11.461, df = 2, p-value = 0.003246
```

$p = 0.0032 < 0.05$. Reject the null hypothesis. There is evidence that there is a difference in responses from college graduates and non-graduates.