Stat511 Hw9

Amy Fox

11/7/2019

Question 1

This is question 10.21 from Ott & Longnecker 7 th Edition. The toxicity of two sludge fertilizer treatments are compared. One hundred tomato plants were randomly assigned to pots containing sludge processed by either the "new" or "old" treatment. The tomatoes harvested from the plants were evaluated to determine if the nickel was at a toxic level. Let πNew be the true population proportion of all plants treated with the "New" treatment that will be toxic and let πOld be the true population proportion of all plants treated with the "Old" treatment that will be toxic. For the tests considered in parts B,C and E below we will test H0: πNew - πOld = 0.

Toxic Non-Toxic Total New 5 45 50 Old 9 41 50 Total 14 86 100

A. Calculate the estimated proportion of toxic plants for EACH treatment.

```
toxic_matrix <- matrix(c(5, 45, 9, 41), byrow = TRUE, nrow = 2)
colnames(toxic_matrix) <- c("Toxic", "Non-toxic")
rownames(toxic_matrix) <- c("New", "Old")

prop.table(toxic_matrix, 1)
```

```
##     Toxic Non-toxic
## New  0.10      0.90
## Old  0.18      0.82
```

New treatment toxic: 0.1

Old treatment toxic: 0.18

B. Use the two-sample Z test (prop.test function in R) to compare the proportion of Toxic plants for the two treatments. Use correct = TRUE (default). Give your (chi-square) test statistic and p-value.

```
prop.test(toxic_matrix, correct = TRUE)
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  toxic_matrix
## X-squared = 0.74751, df = 1, p-value = 0.3873
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.23510963  0.07510963
## sample estimates:
## prop 1 prop 2
##   0.10   0.18
```

Chi-square test statistic: 0.74751

p-value: 0.3873

C. Use the chi-square test for contingency tables (chisq.test function in R) to compare the proportion of Toxic plants for the two treatments. Use correct = TRUE (default). Give your test statistic and p-value. Note: You should find that these results exactly match the results above.

```
chisq.test(toxic_matrix, correct = TRUE)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  toxic_matrix
## X-squared = 0.74751, df = 1, p-value = 0.3873
```

Test statitic: 0.74751

p-value = 0.3873

D. Calculate the expected cell counts from the chi-square test.

```
chisq.test(toxic_matrix)$expected
```

```
##     Toxic Non-toxic
## New     7        43
## Old     7        43
```

E. Use Fisher's exact test (fisher.test function in R) to compare the proportion of Toxic plants for the two treatments. Give the resulting p-value.

```
fisher.test(toxic_matrix)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  toxic_matrix
## p-value = 0.3881
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.1235385 1.8596694
## sample estimates:
## odds ratio
##  0.5095856
```

p-value: 0.3881

F. Considering the expected cell counts from part D above, is the chi-square test or fisher's exact test preferred for this data. Justify your choice.

Fisher's test is used when the cell size is small, and chi-squared is used when the cell sizes are larger. The smallest cell size is still greater than 5, meaning that a chi-square test can be used.

G. The "New" treatment would be preferred if it can be shown to reduce toxicity. Hence a one-side alternative could be justified here. A benefit of the two-sample Z test is that it lends itself more easily to a one-sided alternative. Use prop.test to test H0: πNew - πOld $\geq$ 0 versus HA: πNew - πOld $<$ 0. Give the resulting one-sided p-value.

```
prop.test(toxic_matrix, alternative = "less", correct = TRUE)
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
```

```
## data:  toxic_matrix
## X-squared = 0.74751, df = 1, p-value = 0.1936
## alternative hypothesis: less
## 95 percent confidence interval:
##  -1.00000000  0.05338758
## sample estimates:
## prop 1 prop 2
##   0.10   0.18
```

p-value: 0.1936

Question 2

This is question 10.69 from Ott and Longnecker 7th Edition. A study was conducted to compare two anesthetic drugs for use in minor surgery using n = 45 men who were similar in age and physical condition. The two drugs were applied on the right and left ankles of each patient, and after a fixed period of time, the doctor recorded whether or not the ankle remained anesthetized.

Drug B Yes Drug B No Total Drug A Yes 12 10 22 Drug A No 9 14 23 Total 21 24 45

A. Calculate the estimated proportion that remain anesthetized (Yes) for EACH Drug. Note the data formatting.

```r
drug_matrix <- matrix(c(12, 10, 9, 14), byrow = TRUE, nrow = 2)
colnames(drug_matrix) <- c("Drug B Yes", "Drug B No")
rownames(drug_matrix) <- c("Drug A Yes", "Drug A No")

# Drug A Yes
22/45
```

```
## [1] 0.4888889
```

```r
# Drug B Yes
21/45
```

```
## [1] 0.4666667
```

B. Considering the design, run an appropriate test comparing the effectiveness of the two drugs. State the name of the test, test statistic and p-value. (4 pts)

McNemar's test should be used because the data is paired

```r
mcnemar.test(drug_matrix)
```

```
##
##  McNemar's Chi-squared test with continuity correction
##
## data:  drug_matrix
## McNemar's chi-squared = 0, df = 1, p-value = 1
```

Test name: McNemar's

Test statistic: 0

p-value: 1

Question 3

A case-control study in Berlin, reported by Kohlmeier, Arminger, Bartolomeycik, Bellach, Rehm and Thamm (1992) and by Hand et al. (1994) asked 239 lung cancer patients and 429 healthy controls (matched to the cases by age and sex) whether or not they had kept a pet bird during adulthood. The data is summarized below:

Healthy Controls Cancer Patients Total No Bird 328 141 469 Bird 101 98 199 Total 429 239 668

A. Estimate the odds ratio (of "Cancer" versus "Control") for the "Bird" versus "No Bird" groups. Does the Bird or No Bird group have higher odds of lung cancer?

```
bird_cancer <- matrix(c(328,141, 101, 98), byrow = TRUE, nrow = 2)
colnames(bird_cancer) <- c("Healthy", "Cancer")
rownames(bird_cancer) <- c("No Bird", "Bird")
bird_cancer
```

```
##         Healthy Cancer
## No Bird     328    141
## Bird        101     98
```

```
# odds of cancer in Bird group/odds of cancer in no bird group
((98/199)/(101/199)) / ((141/469)/(328/469))
```

```
## [1] 2.257145
```

Odds of getting cancer in the bird group is 2.3x the odds of getting cancer in the no bird group.

B. Give a 95% confidence interval for the odds ratio. Based on this interval, can you conclude that there is a relationship between bird ownership and lung cancer? NOTE: Use method="wald". (4 pts)

```
library(epitools)
oddsratio(bird_cancer, method = "wald")
```

```
## $data
##         Healthy Cancer Total
## No Bird     328    141   469
## Bird        101     98   199
## Total       429    239   668
##
## $measure
##                          NA
## odds ratio with 95% C.I. estimate   lower    upper
##                  No Bird 1.000000      NA       NA
##                  Bird    2.257145 1.60518 3.173915
##
## $p.value
##          NA
## two-sided   midp.exact fisher.exact   chi.square
##   No Bird           NA           NA           NA
##   Bird    3.052348e-06 3.938413e-06 2.243712e-06
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```

Based on the 95% CI for Bird: (1.60, 3.18), we can conclude that there is a difference because the CI does not include 1.

C. Run the chi-square test for this data. Give the p-value and conclusion.

```
chisq.test(bird_cancer)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  bird_cancer
## X-squared = 21.547, df = 1, p-value = 3.452e-06
```

p-value: 0.345×10^{-6}

As the p-value $< \alpha = 0.05$, we reject H0 and conclude that the row and column variables are dependent (associated) aka bird ownership and lung cancer are associated.