# STAT 511A Homework 9

*Kathleen Wendt*

*11/15/2019*

## Load packages

```
library(broom)
library(epitools)
```

## Question 1

This is question 10.21 from Ott & Longnecker 7th 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 $\pi$new be the true population proportion of all plants treated with the "New" treatment that will be toxic and let $\pi$old 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: $\pi$new - $\pi$old = 0.

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

```
toxic <- matrix(c(5, 45, 9, 41), nrow = 2, byrow = TRUE)
colnames(toxic) <- c("toxic", "nontoxic")
rownames(toxic) <- c("new", "old")
toxic
```

```
##     toxic nontoxic
## new     5       45
## old     9       41
```

```
prop.table(x = toxic, margin = 1)
```

```
##     toxic nontoxic
## new  0.10     0.90
## old  0.18     0.82
```

### Part 1A

Calculate the estimated proportion of toxic plants for EACH treatment.

```
toxic_new <- 5/50
toxic_new
```

```
## [1] 0.1
```

```
toxic_old <- 9/50
toxic_old
```

```
## [1] 0.18
```

## Part 1B

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.

```
toxic_prop_test <- tidy(prop.test(x = c(5, 9),
                                  n = c(50, 50),
                                  alternative = "two.sided",
                                  conf.level = 0.95,
                                  correct = TRUE))
toxic_prop_test
```

```
## # A tibble: 1 x 9
##   estimate1 estimate2 statistic p.value parameter conf.low conf.high method
##       <dbl>     <dbl>     <dbl>   <dbl>     <dbl>    <dbl>     <dbl> <chr>
## 1       0.1      0.18     0.748   0.387         1   -0.235    0.0751 2-sam~
## # ... with 1 more variable: alternative <chr>
```

#### Test Statistic

The test statistic is 0.7475083.

#### P-value

The corresponding p-value is 0.3872663.

## Part 1C

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.

```
toxic_chi <- tidy(chisq.test(x = toxic, correct = TRUE))
toxic_chi
```

```
## # A tibble: 1 x 4
##   statistic p.value parameter method
##       <dbl>   <dbl>     <int> <chr>
## 1     0.748   0.387         1 Pearson's Chi-squared test with Yates' conti~
```

#### Test Statistic

The test statistic is 0.7475083.

#### P-Value

The corresponding p-value is 0.3872663.

### Part 1D

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

```
toxic_exp <- chisq.test(x = toxic)
toxic_exp$expected
```

```
##     toxic nontoxic
## new     7       43
## old     7       43
```

### Part 1E

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.

```
toxic_fisher <- tidy(fisher.test(x = toxic))
toxic_fisher$p.value
```

```
## [1] 0.3880625
```

### Part 1F

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 Exact Test is preferred due to small sample size.

### Part 1G

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: $\pi$new - $\pi$old greater than or equal to 0 versus HA: $\pi$new - $\pi$old less than 0. Give the resulting one-sided p-value.

```
toxic_prop_less <- tidy(prop.test(x = c(5, 9),
                              n = c(50, 50),
                              alternative = "less",
                              conf.level = 0.95,
                              correct = TRUE))
toxic_prop_less$p.value
```

```
## [1] 0.1936331
```

## 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

### Part 2A

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

```
drugs <- matrix(c(22, 23, 21, 24), nrow = 2, byrow = TRUE)
colnames(drugs) <- c("yes", "no")
rownames(drugs) <- c("drugA", "drugB")
drugs
```

```
##       yes no
## drugA  22 23
## drugB  21 24
```

```
drugA_yes <- 22/45
drugA_yes
```

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

```
drugB_yes <- 21/45
drugB_yes
```

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

### Part 2B

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.

#### Name of Test

McNemar's Test for Paired Proportions

```
drugs_mcnemar <- tidy(mcnemar.test(drugs))
```

#### Test Statistic

The test statistic is 0.0227273.

#### P-Value

The p-value is 0.8801685.

## 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

```r
bird <- matrix(c(328, 141, 101, 98), nrow = 2, byrow = TRUE)
colnames(bird) <- c("healthy", "cancer")
rownames(bird) <- c("no", "bird")
bird
```

```
##      healthy cancer
## no       328    141
## bird     101     98
```

```r
prop.table(bird, margin = 1)
```

```
##        healthy    cancer
## no   0.6993603 0.3006397
## bird 0.5075377 0.4924623
```

## Part 3A

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?

```r
oddsratio(bird, method = "wald")
```

```
## $data
##       healthy cancer Total
## no        328    141   469
## bird      101     98   199
## Total     429    239   668
##
## $measure
##                          NA
## odds ratio with 95% C.I. estimate   lower    upper
##                     no   1.000000      NA       NA
##                     bird 2.257145 1.60518 3.173915
##
## $p.value
##          NA
## two-sided   midp.exact fisher.exact   chi.square
##      no             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"
```

The Bird group has higher odds of lung cancer.

## Part 3B

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)

See results in Part 3A. The 95% confidence interval for the odds ratio is 1.61, 3.17. We can conclude there is evidence to suggest a relationship between bird ownership and lung cancer.

## Part 3C

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

```
bird_chi <- tidy(chisq.test(bird))
bird_chi$p.value
```

```
## [1] 3.452389e-06
```

We reject the null hypothesis; there is evidence to suggest a relationship between bird ownership and cancer in adulthood.