STA138: Project 1

Graham Smith (912355584), Chad Pickering (913328497)

Part One: Successful Treatment

The Trial dataset has three nominal variables: group, treatment success, and year treated. The data describes patients with a particular condition, followed over two years.

Column descriptions:

Column 1: Success: Whether the patient successfully treated their condition - Yes, No

Column 2: Group: What group the patient was in - A, B

Column 3: Year: What year of treatment the patient was in - One, Two

First, to prepare for analysis, we set up the required tables.

1. Estimate the probability of a successful treatment overall.

```
## Overall risk for successful treatment, 95% CI bounds: (0.6590979, 0.7523462)
```

While our sample proportion/risk is 0.7057, we are 95% confident that the true probability of a successful treatment regardless of year or group is between 0.6591 and 0.7523.

2. Estimate the probability of a successful treatment, comparing only groups.

```
## Risk for Group A successful treatment, 95% CI bounds: (0.6465238, 0.7773892)
```

```
## Risk for Group B successful treatment, 95% CI bounds: (0.6330245, 0.7658826)
```

```
## Difference in group risk proportions, 95% CI bounds: (-0.08074001, 0.105746)
```

The sample risks of success for group A and group B are 0.7120 and 0.6995, respectively. We are 95% confident that the true probability of successful treatment for group A is between 0.6465 and 0.7774. Similarly, we are 95% confident that the true probability of successful treatment for group B is between 0.6330 and 0.7659. When explicitly comparing the risks, we are 95% confident that the difference in risk proportions between groups A and B is between -0.0807 and 0.1057. Because the interval contains 0, we can say that the two groups do not have a significantly different risk of successful treatment.

3. Estimate the probability of a successful treatment, comparing only years.

```
## Risk for Year 1 successful treatment, 95% CI bounds: (0.7051823, 0.8173402)
```

```
## Risk for Year 2 successful treatment, 95% CI bounds: (0.541713, 0.6996663)
```

```
## Difference in risk proportions for year, 95% CI bounds: (0.04371006, 0.2374332)
```

The sample risk proportions for year 1 and year 2 are 0.7613 and 0.6207, respectively. We are 95% confident that the true probability of successful treatment for year 1 is between 0.7052 and 0.8173. In contrast, we are 95% confident that the true probability of successful treatment for year 2 is between 0.5417 and 0.6997. At this point, we do recognize that these 95% confidence intervals do not overlap, and thus, we expect

the difference in risk between years to be significant. This is true - upon calculating the difference of risk proportion between years, we are 95% confident that the difference in proportions here is between 0.0437 and 0.2374. Because 0 is not in the interval, we can say that the two years have a significantly different true risk proportion. In fact, because the bounds of the confidence intervals are both positive, it seems that year 1 has a significantly higher success proportion in comparison with year 2 (this can be quantitatively assessed with a separate one-sided hypothesis test with alternative hypothesis $\pi_1 - \pi_2 > 0$, but our current analysis is sufficient).

4. Assess the relationship between Group and Success, with and without information from year.

```
## Odds ratio for successful treatment unconditional on year, 95% CI bounds: (0.6778305, 1.664084)
```

We are 95% confident that, unconditional on year, the odds of successfully treating group A is between 0.6778 and 1.6641 times higher than the odds of successfully treating group B. This interval contains 1, so the odds of treating group A is not significantly higher than the odds of treating group B; in this way, we can say that, unconditional on year, group and success are independent.

```
## Odds ratio for successful treatment, conditional on year 1, 95% CI bounds: (0.5804306, 1.994671)
```

```
## Odds ratio for successful treatment, conditional on year 2, 95% CI bounds: (0.5301455, 2.028586)
```

We are 95% confident that, conditional on year 1, the odds of successfully treating group A is between 0.5804 and 1.9947 times higher than the odds of successfully treating group B. This interval contains 1, so the odds of treating group A is not significantly higher than the odds of treating group B; in this way, we can say that, conditional on year 1, group and success are independent. We find similar results for year 2, in that we are 95% confident that, conditional on year 2, the odds of successfully treating group A is between 0.5301 and 2.0286 times higher than the odds of successfully treating group B. In part 3 we assessed the difference in risks between years and concluded that 0 was not in the 95% confidence interval, so we will refrain from redundant analysis about the dependence of risk of success on year, which was confirmed based on that earlier result.

Simpson's Paradox:

```
##      Year
## Group   No Year    Year 1    Year 2
##     A 0.7119565 0.7678571 0.6250000
##     B 0.6994536 0.7545455 0.6164384
```

To check whether the confounding variable, year, changes the direction of association, we assessed the risk of successful treatment conditional and unconditional on year. Upon comparison, we determined that the direction of association remained the same regardless of if year is considered. Thus, Simpson's Paradox is not present.

Part Two: Horror

A sociologist is interested in how gender and death type of people in horror films are related.

Column descriptions:

Column 1: Gender: The gender of the subject - Female, Male

Column 2: Death/Method: The type of death of the subject - Shot, Stabbed, BFT (Blunt force trauma), Other

As customary, our null hypothesis is that gender and method are independent, whereas the alternative hypothesis is that they are dependent in some fashion. We will determine the dependence of these variables using multiple pairwise odds ratio confidence intervals, and then use partitioning to confirm our results. First, we construct a table with the count data and compute risks.

```
##        Method
## Gender BFT Other Shot Stabbed
##      F  38    28   23      61
##      M  14     6   39      42
```

The following are the risks of the various methods for males and females:

```
##        Method
## Gender       BFT     Other      Shot   Stabbed
##      F 0.7307692 0.8235294 0.3709677 0.5922330
##      M 0.2692308 0.1764706 0.6290323 0.4077670
```

We notice that the risks for females are all higher than that for males except in the group "Shot". That could lend to a higher chi-squared statistic, and a rejection of independence for the relationship between gender and method of death. We can find out by conducting further tests.

Pearson's Chi-Squared Test for Independence

We would like to see if gender and method are dependent in general using a Pearson's Chi-Squared Test for Independence.

```
##
##  Pearson's Chi-squared test
##
## data:  horror_tbl
## X-squared = 24.307, df = 3, p-value = 2.155e-05
```

With a very small P-value and test statistic value of 24.307, we can conclude that gender and method are NOT independent, rejecting our null hypothesis (this would be rejected with significance levels of even as low as 0.001). Now, we would like to see which method contributes the most to the deviation from expected counts.

The following are the expected counts, to perhaps see how far off the observed counts were without scaling the data in any way:

```
##        Method
## Gender     BFT    Other     Shot  Stabbed
##      F 31.0757 20.31873 37.05179 61.55378
##      M 20.9243 13.68127 24.94821 41.44622
```

We also include standardized residuals, measuring how far from the expected counts the observed counts were, but in a way such that each residual is standardized (normally distributed with mean 0 and variance 1); one can easily see which response category deviated the most:

```
##        Method
## Gender        BFT      Other       Shot    Stabbed
##      F  2.1991361  2.8891377 -4.1938149 -0.1449092
##      M -2.1991361 -2.8891377  4.1938149  0.1449092
```

We see that the "Shot" category contributed the most to the dependence of these variables, and this is a significant deviance from independence as this specific r_{ij} is greater than 3. We see that unlike the other three categories, men tend to be shot more often than women. This relationship will be confirmed with different strategies as we continue.

Inference with Odds Ratio CIs

Now we calculate the sample odds ratios for each pair.

```
## [1] 0.5816327 4.6024845 1.8688525 7.9130435 3.2131148 0.4060530
```

Now that we have calculated the sample odds ratios for each combination of response variable, we will now construct confidence intervals for each of the pairs. There are $\binom{4}{2}$, or 6 odds ratios, and thus our Bonferroni correction will be with g=6, such that the type 1 error will be controlled to about $\alpha = 0.05$.

```
## Odds ratio for methods BFT and Other, 99.167% CI bounds: (0.1566993, 2.15889)

## Odds ratio for methods BFT and Shot, 99.167% CI bounds: (1.730968, 12.23758)

## Odds ratio for methods BFT and Stabbed, 99.167% CI bounds: (0.7681043, 4.547051)

## Odds ratio for methods Other and Shot, 99.167% CI bounds: (2.273028, 27.54751)

## Odds ratio for methods Other and Stabbed, 99.167% CI bounds: (0.9882086, 10.44729)

## Odds ratio for methods Shot and Stabbed, 99.167% CI bounds: (0.1840008, 0.8960776)
```

So, with the simultaneous CIs, we see that gender and method are dependent when the methods considered are BFT/Shot, Other/Shot, and Shot/Stabbed because the confidence intervals do not include 1. We can even interpret a particular confidence interval as follows if needed. Choosing the OR between BFT and Shot here, we are about 99.167% confident that the true odds of being shot for men is between 1.730968 and 12.23758 times higher than for women relative to that of BFT.

Again, we see that "Shot" contributes the most to the large chi-square value, and is the main category that contributes toward the dependence of gender and method of death.

Partitioning Methodology

Finally, we wish to confirm these results with partitioning methodology. We will keep the columns in the same order and carry out the strategies contained in the class notes, Lesson 10, and in the text, Ch 3.3.

```
##   X-squared   X-squared   X-squared
##  0.99112848 23.62324327  0.02099867

## 24.63537 is the sum of the partitioned chi-square values.

## 24.30671 is the full chi-squared test statistic.
```

In conclusion, we again see that the "Shot" category is contributing to the chi-square value the most because, as said previously, men are more likely to get shot than women whereas women are more likely to encounter any other method framed by the question. Interestingly, the sum of the partitioned chi-square values does not sum to the full test statistic, but instead goes over slightly. We attribute this to the sample size (Piazza, #86). Otherwise, the origin of any other deviation from the theory introduced in class is unknown.

The sociologist should also look into confounding variables such as age of victim (e.g. young/old), release period of film (e.g. pre-1990, post-1990), gender of perpetrator, etc. With the conclusions reached thus far and these additional examinations, they would perhaps be able to make even more interesting developments into the modernization of horror films, and perhaps how society has changed over time.

Appendix

```
# Set up initial tables.
library(readr)
Trial <- read_csv("C:/Users/cpickering/Syncplicity Folders/ChadSync/STATISTICS/STA138/138 Project1/Trial
```

```r
table1 <- table(Trial$Group, Trial$Success)
split.data <- split(Trial, Trial$Year)
year_1_table <- table(split.data$One$Group, split.data$One$Success)
year_2_table <- table(split.data$Two$Group, split.data$Two$Success)
n_T <- 367
total_prob <- (131+128)/n_T
lower_bound_T <- total_prob-qnorm(1-(.05/2))*sqrt((total_prob*(1-total_prob))/n_T)
upper_bound_T <- total_prob+qnorm(1-(.05/2))*sqrt((total_prob*(1-total_prob))/n_T)
cat("Overall risk for successful treatment, 95% CI bounds: (",
    lower_bound_T, ", ", upper_bound_T, ")", sep = "")
n_A <- 53+131
grp_A_prob <- (131)/n_A
lower_bound_A <- grp_A_prob-qnorm(1-(.05/2))*sqrt((grp_A_prob*(1-grp_A_prob))/n_A)
upper_bound_A <- grp_A_prob+qnorm(1-(.05/2))*sqrt((grp_A_prob*(1-grp_A_prob))/n_A)

cat("Risk for Group A successful treatment, 95% CI bounds: (",
    lower_bound_A, ", ", upper_bound_A, ")", sep = "")
n_B <- 55+128
grp_B_prob <- (128)/n_B
lower_bound_B <- grp_B_prob-qnorm(1-(.05/2))*sqrt((grp_B_prob*(1-grp_B_prob))/n_B)
upper_bound_B <- grp_B_prob+qnorm(1-(.05/2))*sqrt((grp_B_prob*(1-grp_B_prob))/n_B)

cat("Risk for Group B successful treatment, 95% CI bounds: (",
    lower_bound_B, ", ", upper_bound_B, ")", sep = "")
lower_bound_diff <- (grp_A_prob-grp_B_prob)-qnorm(1-(.05/2))*sqrt(((grp_A_prob*(1-grp_A_prob))/n_A)+((gr
upper_bound_diff <- (grp_A_prob-grp_B_prob)+qnorm(1-(.05/2))*sqrt(((grp_A_prob*(1-grp_A_prob))/n_A)+((gr

cat("Difference in group risk proportions, 95% CI bounds: (",
    lower_bound_diff, ", ", upper_bound_diff, ")", sep = "")
n_y1 <- 86+83+26+27
year_1_prob <- (86+83)/n_y1
lower_bound_1 <- year_1_prob-qnorm(1-(.05/2))*sqrt((year_1_prob*(1-year_1_prob))/n_y1)
upper_bound_1 <- year_1_prob+qnorm(1-(.05/2))*sqrt((year_1_prob*(1-year_1_prob))/n_y1)

cat("Risk for Year 1 successful treatment, 95% CI bounds: (",
    lower_bound_1, ", ", upper_bound_1, ")", sep = "")
n_y2 <- 45+45+27+28
year_2_prob <- (45+45)/n_y2
lower_bound_2 <- year_2_prob-qnorm(1-(.05/2))*sqrt((year_2_prob*(1-year_2_prob))/n_y2)
upper_bound_2 <- year_2_prob+qnorm(1-(.05/2))*sqrt((year_2_prob*(1-year_2_prob))/n_y2)

cat("Risk for Year 2 successful treatment, 95% CI bounds: (",
    lower_bound_2, ", ", upper_bound_2, ")", sep = "")
lower_bound_diffyr <- (year_1_prob-year_2_prob)-qnorm(1-(.05/2))*sqrt(((year_1_prob*(1-year_1_prob))/n_y
upper_bound_diffyr <- (year_1_prob-year_2_prob)+qnorm(1-(.05/2))*sqrt(((year_1_prob*(1-year_1_prob))/n_y

cat("Difference in risk proportions for year, 95% CI bounds: (",
    lower_bound_diffyr, ", ", upper_bound_diffyr, ")", sep = "")
# Without year:
theta_hat_y <- (55*131)/(53*128)
lower_bound_wo <- exp(log(theta_hat_y) - qnorm(1-(.05/2))*sqrt((1/53)+(1/131)+(1/55)+(1/128)))
upper_bound_wo <- exp(log(theta_hat_y) + qnorm(1-(.05/2))*sqrt((1/53)+(1/131)+(1/55)+(1/128)))
```

```r
cat("Odds ratio for successful treatment unconditional on year, 95% CI bounds: (",
    lower_bound_wo, ", ", upper_bound_wo, ")", sep = "")
# With year:
theta_hat_y1 <- (86*27)/(83*26)
lower_bound_1 <- exp(log(theta_hat_y1) - qnorm(1-(.05/2))*sqrt((1/86)+(1/26)+(1/83)+(1/27)))
upper_bound_1 <- exp(log(theta_hat_y1) + qnorm(1-(.05/2))*sqrt((1/86)+(1/26)+(1/83)+(1/27)))

cat("Odds ratio for successful treatment, conditional on year 1, 95% CI bounds: (",
    lower_bound_1, ", ", upper_bound_1, ")", sep = "")
theta_hat_y2 <- (45*28)/(45*27)
lower_bound_2 <- exp(log(theta_hat_y2) - qnorm(1-(.05/2))*sqrt((1/45)+(1/28)+(1/45)+(1/27)))
upper_bound_2 <- exp(log(theta_hat_y2) + qnorm(1-(.05/2))*sqrt((1/45)+(1/28)+(1/45)+(1/27)))

cat("Odds ratio for successful treatment, conditional on year 2, 95% CI bounds: (",
    lower_bound_2, ", ", upper_bound_2, ")", sep = "")
# Without Z:
n_A <- 53+131
grp_A_prob <- (131)/n_A

n_B <- 55+128
grp_B_prob <- (128)/n_B

# Z_1
n_y1A <- 26+86
grp_A_y1 <- (86)/n_y1A

n_y1B <- 27+83
grp_B_y1 <- (83)/n_y1B

# Z_2
n_y2A <- 27+45
grp_A_y2 <- (45)/n_y2A

n_y2B <- 28+45
grp_B_y2 <- (45)/n_y2B

matrix(c(grp_A_prob, grp_B_prob, grp_A_y1, grp_B_y1, grp_A_y2, grp_B_y2),
       nrow = 2, ncol = 3, byrow = FALSE,
                           dimnames = list(Group = c('A', 'B'),
                                           Year = c('No Year', 'Year 1', 'Year 2')))
library(readr)
horror <- read_csv("C:/Users/cpickering/Syncplicity Folders/ChadSync/STATISTICS/STA138/138 Project1/hor
horror_tbl <- as.table(matrix(c(38, 14, 28, 6, 23, 39, 61, 42),
                           nrow = 2, ncol = 4, byrow = FALSE,
                           dimnames = list(Gender = c('F', 'M'),
                                           Method = c('BFT', 'Other', 'Shot', 'Stabbed'))))
horror_tbl
prop.table(horror_tbl, margin = 2)
horror_full <- chisq.test(horror_tbl, correct = FALSE)
horror_full
horror_full$expected
horror_full$stdres
# Sample odds ratios:
```

```r
theta_12 <- (38*6)/(28*14)
theta_13 <- (38*39)/(14*23)
theta_14 <- (38*42)/(14*61)
theta_23 <- (28*39)/(6*23)
theta_24 <- (28*42)/(6*61)
theta_34 <- (23*42)/(39*61)

c(theta_12, theta_13, theta_14, theta_23, theta_24, theta_34)

# CIs:
lb_12 <- exp(log(theta_12) - qnorm(1-(.05/6))*sqrt((1/38)+(1/6)+(1/28)+(1/14)))
ub_12 <- exp(log(theta_12) + qnorm(1-(.05/6))*sqrt((1/38)+(1/6)+(1/28)+(1/14)))

cat("Odds ratio for methods BFT and Other, 99.167% CI bounds: (",
    lb_12, ", ", ub_12, ")", sep = "")
lb_13 <- exp(log(theta_13) - qnorm(1-(.05/6))*sqrt((1/38)+(1/39)+(1/23)+(1/14)))
ub_13 <- exp(log(theta_13) + qnorm(1-(.05/6))*sqrt((1/38)+(1/39)+(1/23)+(1/14)))

cat("Odds ratio for methods BFT and Shot, 99.167% CI bounds: (",
    lb_13, ", ", ub_13, ")", sep = "")
lb_14 <- exp(log(theta_14) - qnorm(1-(.05/6))*sqrt((1/38)+(1/42)+(1/14)+(1/61)))
ub_14 <- exp(log(theta_14) + qnorm(1-(.05/6))*sqrt((1/38)+(1/42)+(1/14)+(1/61)))

cat("Odds ratio for methods BFT and Stabbed, 99.167% CI bounds: (",
    lb_14, ", ", ub_14, ")", sep = "")
lb_23 <- exp(log(theta_23) - qnorm(1-(.05/6))*sqrt((1/28)+(1/39)+(1/6)+(1/23)))
ub_23 <- exp(log(theta_23) + qnorm(1-(.05/6))*sqrt((1/28)+(1/39)+(1/6)+(1/23)))

cat("Odds ratio for methods Other and Shot, 99.167% CI bounds: (",
    lb_23, ", ", ub_23, ")", sep = "")
lb_24 <- exp(log(theta_24) - qnorm(1-(.05/6))*sqrt((1/28)+(1/42)+(1/6)+(1/61)))
ub_24 <- exp(log(theta_24) + qnorm(1-(.05/6))*sqrt((1/28)+(1/42)+(1/6)+(1/61)))

cat("Odds ratio for methods Other and Stabbed, 99.167% CI bounds: (",
    lb_24, ", ", ub_24, ")", sep = "")
lb_34 <- exp(log(theta_34) - qnorm(1-(.05/6))*sqrt((1/23)+(1/42)+(1/39)+(1/61)))
ub_34 <- exp(log(theta_34) + qnorm(1-(.05/6))*sqrt((1/23)+(1/42)+(1/39)+(1/61)))

cat("Odds ratio for methods Shot and Stabbed, 99.167% CI bounds: (",
    lb_34, ", ", ub_34, ")", sep = "")
# First partition:
horror_1 <- horror_tbl[,1:2]
part_1 <- chisq.test(horror_1, correct = FALSE)

# Second partition:
BFT_other <- c(66, 20)
shot <- c(23, 39)
horror_2 <- rbind(BFT_other, shot)
colnames(horror_2) <- c("F","M")
part_2 <- chisq.test(horror_2, correct = FALSE)

# Third partition:
BFT_other_shot <- c(89, 59)
```

```
stabbed <- c(61, 42)
horror_3 <- rbind(BFT_other_shot, stabbed)
colnames(horror_3) <- c("F","M")
part_3 <- chisq.test(horror_3, correct = FALSE)

# Check to see if the partitioned chi-squared statistics sum to the full one
c(part_1$statistic, part_2$statistic, part_3$statistic)
cat(part_1$statistic + part_2$statistic + part_3$statistic, " is the sum of the partitioned chi-square v
cat(horror_full$statistic, " is the full chi-squared test statistic.", sep="")

# Set up initial tables.
library(readr)
Trial <- read_csv("C:/Users/cpickering/Syncplicity Folders/ChadSync/STATISTICS/STA138/138 Project1/Trial
#1.1
n_T <- 367
total_prob <- (131+128)/n_T
lower_bound_T <- total_prob-qnorm(1-(.05/2))*sqrt((total_prob*(1-total_prob))/n_T)
upper_bound_T <- total_prob+qnorm(1-(.05/2))*sqrt((total_prob*(1-total_prob))/n_T)
cat("Overall risk for successful treatment, 95% CI bounds: (",
    lower_bound_T, ", ", upper_bound_T, ")", sep = "")
#1.2
n_A <- 53+131
grp_A_prob <- (131)/n_A
lower_bound_A <- grp_A_prob-qnorm(1-(.05/2))*sqrt((grp_A_prob*(1-grp_A_prob))/n_A)
upper_bound_A <- grp_A_prob+qnorm(1-(.05/2))*sqrt((grp_A_prob*(1-grp_A_prob))/n_A)
cat("Risk for Group A successful treatment, 95% CI bounds: (",
    lower_bound_A, ", ", upper_bound_A, ")", sep = "")
n_B <- 55+128
grp_B_prob <- (128)/n_B
lower_bound_B <- grp_B_prob-qnorm(1-(.05/2))*sqrt((grp_B_prob*(1-grp_B_prob))/n_B)
upper_bound_B <- grp_B_prob+qnorm(1-(.05/2))*sqrt((grp_B_prob*(1-grp_B_prob))/n_B)
cat("Risk for Group B successful treatment, 95% CI bounds: (",
    lower_bound_B, ", ", upper_bound_B, ")", sep = "")
lower_bound_diff <- (grp_A_prob-grp_B_prob)-qnorm(1-(.05/2))*sqrt(((grp_A_prob*(1-grp_A_prob))/n_A)+((gr
upper_bound_diff <- (grp_A_prob-grp_B_prob)+qnorm(1-(.05/2))*sqrt(((grp_A_prob*(1-grp_A_prob))/n_A)+((gr
cat("Difference in group risk proportions, 95% CI bounds: (",
    lower_bound_diff, ", ", upper_bound_diff, ")", sep = "")
#1.3
n_y1 <- 86+83+26+27
year_1_prob <- (86+83)/n_y1
lower_bound_1 <- year_1_prob-qnorm(1-(.05/2))*sqrt((year_1_prob*(1-year_1_prob))/n_y1)
upper_bound_1 <- year_1_prob+qnorm(1-(.05/2))*sqrt((year_1_prob*(1-year_1_prob))/n_y1)

cat("Risk for Year 1 successful treatment, 95% CI bounds: (",
    lower_bound_1, ", ", upper_bound_1, ")", sep = "")
n_y2 <- 45+45+27+28
year_2_prob <- (45+45)/n_y2
lower_bound_2 <- year_2_prob-qnorm(1-(.05/2))*sqrt((year_2_prob*(1-year_2_prob))/n_y2)
upper_bound_2 <- year_2_prob+qnorm(1-(.05/2))*sqrt((year_2_prob*(1-year_2_prob))/n_y2)

cat("Risk for Year 2 successful treatment, 95% CI bounds: (",
    lower_bound_2, ", ", upper_bound_2, ")", sep = "")
lower_bound_diffyr <- (year_1_prob-year_2_prob)-qnorm(1-(.05/2))*sqrt(((year_1_prob*(1-year_1_prob))/n_y
```

```r
upper_bound_diffyr <- (year_1_prob-year_2_prob)+qnorm(1-(.05/2))*sqrt(((year_1_prob*(1-year_1_prob))/n_y

cat("Difference in risk proportions for year, 95% CI bounds: (",
    lower_bound_diffyr, ", ", upper_bound_diffyr, ")", sep = "")
#1.4
# Without year:
theta_hat_y <- (55*131)/(53*128)
lower_bound_wo <- exp(log(theta_hat_y) - qnorm(1-(.05/2))*sqrt((1/53)+(1/131)+(1/55)+(1/128)))
upper_bound_wo <- exp(log(theta_hat_y) + qnorm(1-(.05/2))*sqrt((1/53)+(1/131)+(1/55)+(1/128)))

cat("Odds ratio for successful treatment unconditional on year, 95% CI bounds: (",
    lower_bound_wo, ", ", upper_bound_wo, ")", sep = "")
# With year:
theta_hat_y1 <- (86*27)/(83*26)
lower_bound_1 <- exp(log(theta_hat_y1) - qnorm(1-(.05/2))*sqrt((1/86)+(1/26)+(1/83)+(1/27)))
upper_bound_1 <- exp(log(theta_hat_y1) + qnorm(1-(.05/2))*sqrt((1/86)+(1/26)+(1/83)+(1/27)))

cat("Odds ratio for successful treatment, conditional on year 1, 95% CI bounds: (",
    lower_bound_1, ", ", upper_bound_1, ")", sep = "")
theta_hat_y2 <- (45*28)/(45*27)
lower_bound_2 <- exp(log(theta_hat_y2) - qnorm(1-(.05/2))*sqrt((1/45)+(1/28)+(1/45)+(1/27)))
upper_bound_2 <- exp(log(theta_hat_y2) + qnorm(1-(.05/2))*sqrt((1/45)+(1/28)+(1/45)+(1/27)))

cat("Odds ratio for successful treatment, conditional on year 2, 95% CI bounds: (",
    lower_bound_2, ", ", upper_bound_2, ")", sep = "")
#checking Simpson's Paradox
# Without Z:
n_A <- 53+131
grp_A_prob <- (131)/n_A

n_B <- 55+128
grp_B_prob <- (128)/n_B

# Z_1
n_y1A <- 26+86
grp_A_y1 <- (86)/n_y1A

n_y1B <- 27+83
grp_B_y1 <- (83)/n_y1B

# Z_2
n_y2A <- 27+45
grp_A_y2 <- (45)/n_y2A

n_y2B <- 28+45
grp_B_y2 <- (45)/n_y2B

matrix(c(grp_A_prob, grp_B_prob, grp_A_y1, grp_B_y1, grp_A_y2, grp_B_y2),
       nrow = 2, ncol = 3, byrow = FALSE,
                           dimnames = list(Group = c('A', 'B'),
                                           Year = c('No Year', 'Year 1', 'Year 2')))
#PART 2
#pearson's Chi-Sq
```

```r
library(readr)
horror <- read_csv("C:/Users/cpickering/Syncplicity Folders/ChadSync/STATISTICS/STA138/138 Project1/hori
horror_tbl <- as.table(matrix(c(38, 14, 28, 6, 23, 39, 61, 42),
                          nrow = 2, ncol = 4, byrow = FALSE,
                          dimnames = list(Gender = c('F', 'M'),
                                          Method = c('BFT', 'Other', 'Shot', 'Stabbed'))))
horror_full <- chisq.test(horror_tbl, correct = FALSE)
horror_full$expected
horror_full$stdres
# Sample odds ratios:
theta_12 <- (38*6)/(28*14)
theta_13 <- (38*39)/(14*23)
theta_14 <- (38*42)/(14*61)
theta_23 <- (28*39)/(6*23)
theta_24 <- (28*42)/(6*61)
theta_34 <- (23*42)/(39*61)
c(theta_12, theta_13, theta_14, theta_23, theta_24, theta_34)
# CIs:
lb_12 <- exp(log(theta_12) - qnorm(1-(.05/6))*sqrt((1/38)+(1/6)+(1/28)+(1/14)))
ub_12 <- exp(log(theta_12) + qnorm(1-(.05/6))*sqrt((1/38)+(1/6)+(1/28)+(1/14)))

cat("Odds ratio for methods BFT and Other, 99.167% CI bounds: (",
    lb_12, ", ", ub_12, ")", sep = "")
lb_13 <- exp(log(theta_13) - qnorm(1-(.05/6))*sqrt((1/38)+(1/39)+(1/23)+(1/14)))
ub_13 <- exp(log(theta_13) + qnorm(1-(.05/6))*sqrt((1/38)+(1/39)+(1/23)+(1/14)))

cat("Odds ratio for methods BFT and Shot, 99.167% CI bounds: (",
    lb_13, ", ", ub_13, ")", sep = "")

lb_14 <- exp(log(theta_14) - qnorm(1-(.05/6))*sqrt((1/38)+(1/42)+(1/14)+(1/61)))
ub_14 <- exp(log(theta_14) + qnorm(1-(.05/6))*sqrt((1/38)+(1/42)+(1/14)+(1/61)))

cat("Odds ratio for methods BFT and Stabbed, 99.167% CI bounds: (",
    lb_14, ", ", ub_14, ")", sep = "")

lb_23 <- exp(log(theta_23) - qnorm(1-(.05/6))*sqrt((1/28)+(1/39)+(1/6)+(1/23)))
ub_23 <- exp(log(theta_23) + qnorm(1-(.05/6))*sqrt((1/28)+(1/39)+(1/6)+(1/23)))

cat("Odds ratio for methods Other and Shot, 99.167% CI bounds: (",
    lb_23, ", ", ub_23, ")", sep = "")

lb_24 <- exp(log(theta_24) - qnorm(1-(.05/6))*sqrt((1/28)+(1/42)+(1/6)+(1/61)))
ub_24 <- exp(log(theta_24) + qnorm(1-(.05/6))*sqrt((1/28)+(1/42)+(1/6)+(1/61)))

cat("Odds ratio for methods Other and Stabbed, 99.167% CI bounds: (",
    lb_24, ", ", ub_24, ")", sep = "")

lb_34 <- exp(log(theta_34) - qnorm(1-(.05/6))*sqrt((1/23)+(1/42)+(1/39)+(1/61)))
ub_34 <- exp(log(theta_34) + qnorm(1-(.05/6))*sqrt((1/23)+(1/42)+(1/39)+(1/61)))

cat("Odds ratio for methods Shot and Stabbed, 99.167% CI bounds: (",
    lb_34, ", ", ub_34, ")", sep = "")
# First partition:
```

```r
horror_1 <- horror_tbl[,1:2]
part_1 <- chisq.test(horror_1, correct = FALSE)

# Second partition:
BFT_other <- c(66, 20)
shot <- c(23, 39)
horror_2 <- rbind(BFT_other, shot)
colnames(horror_2) <- c("F","M")
part_2 <- chisq.test(horror_2, correct = FALSE)

# Third partition:
BFT_other_shot <- c(89, 59)
stabbed <- c(61, 42)
horror_3 <- rbind(BFT_other_shot, stabbed)
colnames(horror_3) <- c("F","M")
part_3 <- chisq.test(horror_3, correct = FALSE)

# Check to see if the partitioned chi-squared statistics sum to the full one
c(part_1$statistic, part_2$statistic, part_3$statistic)
cat(part_1$statistic + part_2$statistic + part_3$statistic, " is the sum of the partitioned chi-square v
cat(horror_full$statistic, " is the full chi-squared test statistic.", sep="")
```