

6. Let confidence interval be denoted as CI, and let contingency table be denoted as CT.

a. The difference in proportion between unintentional gunshot count and terrorism gunshot count, each relative to the total count across all categories, yields a 95% CI of (0.222324, 0.225345). The relative risk of the same two categories yields a 95% CI of (3558.001, 12058.929). The following code in R can confirm these results.

```
# Original count data
gunshot_counts <- list(
  assault = 179793,
  unintentional = 65502,
  suicide = 24624,
  undetermined = 17401,
  law_enforcement = 5266,
  terrorism = 10
)

# Variables
n <- sum(unlist(x = gunshot_counts, use.names = FALSE))
unintentional <- as.integer(x = gunshot_counts['unintentional'])
terrorism <- as.integer(x = gunshot_counts['terrorism'])

# Contingency table
gunshot_ct <- matrix(
  data = c(unintentional, n - unintentional,
           terrorism, n - terrorism),
  nrow = 2,
  byrow = TRUE
)

print(gunshot_ct)

# Difference of proportions tests
cat("\nDifference of proportions using prop.test:\n")
prop.test(
  x = gunshot_ct,
  conf.level = 0.95,
  alternative = 'two.sided',
  correct = FALSE
)

cat("\nDifference of proportions using diffscoreci:\n")
diffscoreci(
```

```
x1 = unintentional,  
n1 = n,  
x2 = terrorism,  
n2 = n,  
conf.level = 0.95  
)  
  
# Risk score tests  
cat("\nRisk score:\n")  
riskscoreci(  
  x1 = unintentional,  
  n1 = n,  
  x2 = terrorism,  
  n2 = n,  
  conf.level = 0.95  
)
```

From averaging the endpoints in CI, the sample difference in proportions seems very small being 0.2238345. However, the odds of each category are both very close to zero (0), with odds being 0.2884356 and 3.417799×10^{-5} for unintentional and terrorism, respectively. This may be misleading. Low odds just means it is unlikely for a gunshot to be resulting from unintentional acts (or similarly terrorism) relative to other categories of reasons. The small odds for both unintentional and terrorism are small partly because many other categories of gunshots in the data not conveyed in the CT. Recall that the first column represents a success (gunshot from unintentional or terrorism), but the second column represents all other categories together. Therefore, relative risk may be a better comparative metric for unintentional and terrorism proportions. Computing the relative risk shows that gunshot victims are more than 3.5k times as likely to be the result of unintentional acts rather than terrorism.

b. The two variables, corresponding to the rows and columns respectively, are ideological group (Democrat and liberal, or Republican and conservative) and support of legal gay marriage (favor or not favor). The odds ratio is 28.68595. Note that favoring legal gay marriage aligns with “success” in the binary outcome (first column in the CT). The R code is below.

```
# Odds ratio manually  
pi_democrat_liberal <- 0.89  
odds_democrat_liberal <- pi_democrat_liberal / (1 - pi_democrat_liberal)  
pi_rep_con <- 0.22  
odds_rep_con <- pi_rep_con / (1 - pi_rep_con)  
odds_ratio <- odds_democrat_liberal / odds_rep_con  
cat("\nOdds ratio manually:\n", odds_ratio, "\n")  
  
# Contingency table
```

Caijun Qin

09/17/2021

```
political_ct <- matrix(
  data = c(pi_democrat_liberal, 1 - pi_democrat_liberal,
    pi_rep_con, 1 - pi_rep_con) * 100,
  nrow = 2,
  byrow = TRUE
)
print(political_ct)

cat("\nOdds ratio manually:\n", odds_ratio, '\n')

# Odds ratio with oddsratio function
cat("\nOdds ratio with epitools oddsratio function:\n")
oddsratio(
  x = c(pi_democrat_liberal, 1 - pi_democrat_liberal,
    pi_rep_con, 1 - pi_rep_con) * 100,
  method = 'wald',
  conf=0.95,
  correct=FALSE
)
```

12. The conditional distribution in the CT refers to Y given X. Although the ordering does not technically matter to the CT, the given answer makes more realistic sense. The rows would be one variable, race of murderer (white or black), and the columns would be the outcome variable, race of slain victim (white or black). The odds ratio is 152.7857, computed by the R code below.

```
# Odds ratio manually
pi_white_white <- 0.93
odds_white_victim <- pi_white_white / (1 - pi_white_white)
pi_black_black <- 0.92
odds_black_victim <- (1 - pi_black_black) / pi_black_black
odds_ratio <- odds_white_victim / odds_black_victim

# Contingency table
crime_ct <- matrix(
  data = c(pi_white_white, 1 - pi_black_black,
    1 - pi_white_white, pi_black_black) * 100,
  nrow = 2,
  byrow = TRUE
)
print(crime_ct)

cat("\nOdds ratio manually:\n", odds_ratio, '\n')
```

```
# Odds ratio with oddsratio function
cat("\nOdds ratio with epitools oddsration function:\n")
oddsratio(
  x = crime_ct,
  method = 'wald',
  conf=0.95,
  correct=FALSE
)
```

16.

a.

The degrees of freedom is $df = (r - 1)(c - 1) = (5 - 1)(3 - 1) = 4 * 2 = 8$.

The p-value is 1.030854e-12 from the given R code below.

```
pchisq(q = 73.4, df = 8, lower.tail = FALSE)
```

Because the χ^2 statistic is so large, and especially the upper-tail p-value is less than the critical level of 0.05 and even 0.001 for testing, we reject the null hypothesis for independence. This means that the absolute deviation of true cell counts from their corresponding expected cell counts sum up to a large enough number to confidently constitute evidence against proportions of each cell being independent from one another.

b.

i. The cells containing counts 21 and 83 have standardized residuals of -2.973 and -5.907 , respectively. For the first cell, the true number of people with above average income that are not too happy does not meet the expected number of people in that income category who theoretically should be unhappy. The second cell also has a negative standardized residual, indicating that the real number of very happy people with below average income is less than the expected number for that income category. Combining interpretations from both cells, high-income level tends to skew more people to be happy while low-income level tend to skew more people towards being unhappy, more so than theoretically expected.

ii. The cells containing counts 110 and 94 have standardized residuals of 3.144 and 7.368 , respectively. The first cell having a positive standardized residual points out that having above average income produced more people that are very happy than expected at that income level. Something in reverse happened to the second cell when interpreted. For the second cell, much more below average income people are not too happy than expected for that income level. These two cells with their standardized residuals further reinforce the evidence from part i. that higher income levels make people happier by a more extreme amount than expected, while lower income levels skew the count of people in the opposite direction in terms of happiness.

23.

a.

i. With a p-value of 0.3808, we cannot reject the plausibility that no treatment is more effective at controlling cancer over another treatment. Specifically, the odds ratio of using surgery versus the odds ratio of using radiation therapy being greater than 1 is not conclusive from any evidence.

ii. With a p-value of 0.6384, the conclusion about no treatment over the other being more effective at controlling cancer remains. Only this time, we are trying to see whether the odds between either treatment are statistically significantly different regardless of direction.