Lesson 3 Homework - Ellen Chancey

Ellen Chancey

September 9, 2018

Question 2.2a

According to FBI in 2008, 92% of blacks killed where killed by blacks. In 2005, 85% of whites killed were killed by whites. X is race of killer, Y is race of victim. Which conditional distribution do these statistics refer to: Y given X or X given Y?
Killer's race given the victims, so X given Y.

Question 2.4

Using the seatbelt data from table 2.10

```
# set up data, do prob calcs manually
n11 <- 1085 # fatal, no sb
n12 <- 55623 # nonfatal, no sb
n21 <- 703 # fatal, yes sb
n22 <- 441239 # nonfatal, yes sb
n <- n11+n12+n21+n22

n1j <- n11+n12 # no sb
n2j <- n21+n22 # yes sb
ni1 <- n11+n21 # fatal
ni2 <- n12+n22 # not fatal

p11 <- n11/n # fatal no sb
p12 <- n12/n # nonfatal no sb
p21 <- n21/n # fatal yes sb
p22 <- n22/n # nonfatal yes sb
p <- p11+p12+p21+p22

p1j <- p11+p12 # prob across no sb
p2j <- p21+p22 # prob across yes sb
pi1 <- p11+p21 # prob across fatality
pi2 <- p12+p22 # prob across no fatality
```

Part A: Estimate probability of fatality conditional on no seatbelt use and yes seatbelt use.

```
py1x2 <- p21/p2j  # prob fatal given yes sb
py1x2
```

```
## [1] 0.001590706
```

```
py1x1 <- p11/p1j   # prob fatal given no sb
py1x1

## [1] 0.0191331
```

Part B : Estimate probability of wearing seatbelt conditional on fatality or no fatality

```
px2y1 <- p21/pi1 # prob sb yes given fatality
px2y1

## [1] 0.3931767

px2y2 <- p22/pi2 # prob sb yes given no fatality
px2y2

## [1] 0.8880514
```

Part C: Difference of proportions, relative risk and odds ratio. Why are RR and OR approximately equal?

```
# difference of proportions
# prob of fatality when no sb - prob fatality when yes sb
p1 <- n11/n1j
p2 <- n21/n2j
dp <- p1 - p2
dp

## [1] 0.0175424

# relative risk
rr <- p1/p2
rr

## [1] 12.02805

# oddsratio
o <- (p1/(1-p1))/(p2/(1-p2))
o

## [1] 12.24317
```

The relative risk and the oddsratio are similar because fatalities in both seatbelt wearing and non seatbelt wearing accidents occurs in small numbers. Because both probability are small, the RR and OR are similiar.

Question 2.9

Occurence of lung cancer in non smokers (sp1) compared with occurence of lung cancer in smokers (sp2). Find and interpret the difference of proportion and relative risk. Which measure is more informative?

```
sp1 <- 0.00023
sp2 <- 0.01284
sdp <- sp2-sp1 # diff proportions
sdp

## [1] 0.01261

srr <- sp2/sp1 # relative risk
srr

## [1] 55.82609
```

The difference of proportions is 0.01261 while the relative risk is 55.826087. Because both probabilities are quite small, the relative risk is more informative at identifying the magnitude of difference between the two values.

Question 2.10

Odds ratio between gender and survival for Titanic passengers is 11.4.
Part A: what is wrong with the interpretation "the probability of surivival for females qas 11.4 times that for males"?
This interpretation is incorrect because the interpretation used is for that of the odds, not the odds ratio. The correct interpretation is the women are far more likely to survive than men.

Part B: The odds of survival for females is 2.9. For each gender, find the proportion that survived.

```
# pi = o/(1+o)
pi <- 2.9/(1+2.9)
pi

## [1] 0.7435897
```

Session Info

```
sessionInfo()

## R version 3.4.1 (2017-06-30)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 16299)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
```

```
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## loaded via a namespace (and not attached):
##  [1] compiler_3.4.1  backports_1.1.0 magrittr_1.5    rprojroot_1.2
##  [5] tools_3.4.1     htmltools_0.3.6 yaml_2.1.14     Rcpp_0.12.13
##  [9] stringi_1.1.5   rmarkdown_1.6   knitr_1.16      stringr_1.2.0
## [13] digest_0.6.12   evaluate_0.10.1
```