BIOS 4120 Lab 2

January 28 - 29, 2020

Objectives

In today's lab we will:

1. Learn to make tables and proportion tables in R
2. Learn bracketing techniques for indexing tables
3. Work on calculating weighted averages of data, adjusting for confounders

Constructing Tables

First, let's read in the 'titanic' dataset and compute some summary statistics.

```
titanic <- read.delim("http://myweb.uiowa.edu/pbreheny/data/titanic.txt")
summary(titanic)
```

```
##   Class         Sex          Age           Survived
##  1st :325   Female: 470   Adult:2092   Died    :1490
##  2nd :285   Male  :1731   Child: 109   Survived: 711
##  3rd :706
##  Crew:885
```

By default, when the summary() function encounters categorical data, it produces a table for that column, as evidenced above, when it created 4 separate tables. We can replicate that using the table() function.

```
table(titanic$Class)
```

```
##
##  1st  2nd  3rd Crew
##  325  285  706  885
```

But the table function is more versatile than that. For example, we can create 2x2 tables:

```
table(titanic$Class,titanic$Survived)
```

```
##
##        Died Survived
##   1st   122      203
##   2nd   167      118
##   3rd   528      178
##   Crew  673      212
```

If we give the function more than two variables, it creates multiple tables (one for each level):

```
table(titanic$Class,titanic$Survived,titanic$Sex)
```

```
## , ,  = Female
##
##
##        Died Survived
##   1st     4      141
##   2nd    13       93
##   3rd   106       90
##   Crew    3       20
```

```
##
## , ,  = Male
##
##
##        Died Survived
##   1st   118       62
##   2nd   154       25
##   3rd   422       88
##   Crew  670      192
```

I'd recommend keeping the number of variables down to 2 or 3, as more than that begins to get a bit cluttered and confusing.

If we save a table, we can use brackets to access individual numbers [row,column]. When the row or column entry is left blank using this type of indexing, R will take all the rows or columns (whichever one was left blank).:

```
table1 <- table(titanic$Class,titanic$Survived)

table1
```

```
##
##        Died Survived
##   1st   122      203
##   2nd   167      118
##   3rd   528      178
##   Crew  673      212
```

```
table1[3,2]
```

```
## [1] 178
```

```
# The 3 indicates the third row and the 2 indicates the second column,
# so this is the number of 3rd class passengers who survived.

table1[4, 1]
```

```
## [1] 673
```

```
table1[,2]
```

```
##  1st  2nd  3rd Crew
##  203  118  178  212
```

```
table1[3,]
```

```
##     Died Survived
##      528      178
```

Additionally, we can use the "==" to access data with certain specifications. For instance, if we wanted to access only the data in which people survived, we could use the following code:

```
justSurv <- titanic[titanic$Survived == "Survived",]

# Or if we only want 2nd class travelers

second <- titanic[titanic$Class == "2nd",]
```

We can also use prop.table() to get the proportions for each cell of a table:

```r
prop.table(table1, 1) # Gives proportions for each class (by row)
```

```
##
##             Died  Survived
##   1st  0.3753846 0.6246154
##   2nd  0.5859649 0.4140351
##   3rd  0.7478754 0.2521246
##   Crew 0.7604520 0.2395480
```

```r
prop.table(table1, 2) # Gives proportions for the survival groups (by column)
```

```
##
##              Died   Survived
##   1st  0.08187919 0.28551336
##   2nd  0.11208054 0.16596343
##   3rd  0.35436242 0.25035162
##   Crew 0.45167785 0.29817159
```

Weighted Averages

Let's investigate Titanic survival rates based on class.
(For the sake of practice, we will do this by hand and then R.)

Recall the steps for computing a weighted average:

1) Identify the outcome, the group you are interested in, and the confounder that you are adjusting for
2) Calculate $w_1, w_2, ..., w_n$, the overall proportion of observations that belong to each level of the confounder
3) Calculate $\bar{x}_1, \bar{x}_2, ..., \bar{x}_n$, the average (or percentage) for that group at each level of the confounder

4)Calculate the weight average $\bar{x} = \Sigma_i w_i \bar{x}_i$

```
, , Sex = Female

      Survived
Class  Survived Total
  1st       141   145
  2nd        93   106
  3rd        90   196
  Crew       20    23

, , Sex = Male

      Survived
Class  Survived Total
  1st        62   180
  2nd        25   179
  3rd        88   510
  Crew      192   862
```

Part a

From "table1" above, calculate the overall percentages of survival for each class.

Part b

Now, create a table listing the percentage of passengers in each class who survived, broken down by sex. ##### Part c Create a table listing the total proportion of male and female passengers on the ship.

Part d

Finally, construct a weighted average of the percentage of passengers in each class who survived, controlling for the effect of sex (i.e., report one number for each class).

Do any of these results surprise you? What changed in Part a when compared to Part d? What conclusions can we draw from this?

Answers

```
## Part a
```

```
##       1st       2nd       3rd      Crew
## 0.6246154 0.4140351 0.2521246 0.2395480
```

```
## Part b
```

```
## Part c
```

```
##
##    Female      Male
## 0.2135393 0.7864607
```

```
## Part d
```

```
## 1st :  0.479
```

```
## 2nd :  0.297
```

```
## 3rd :  0.234
```

```
## Crew:  0.361
```

Weighted Averages in R

We can also use R, to solve these problems. There is no simple function that allows you to calculate the weighted average. Below are a few of ways to do this. Note that the first method could introduce mistakes since you are inputting values individually and also could be lengthy process depending on the structure of the dataset. Typically, we would prefer to use the second method which are more efficient and have less opportunity for error.

```
overallSex <- prop.table(table(titanic$Sex))

#First Method
firstclass <- c(141, 62)/ c(145, 180) # From table provided

(first <- weighted.mean(firstclass, overallSex))
```

```
## [1] 0.4785406
```

```
#Second Method
classtable <- table(titanic$Sex,titanic$Class,titanic$Survived)
classes <- prop.table(classtable, 1:2)[,,2]

(class1 <- weighted.mean(classes[,1], overallSex))
```

```
## [1] 0.4785406
```

Practice

Now let's say that we want to investigate the difference in survival by sex for the Titanic data set. Construct a weighted average of the percentage of passengers for each sex who survived, controlling for the effect of class. For the sake of learning the process do not use the weighted.mean() function except to check your answer

Practice Answers

```
##    Female      Male
## 0.7541256 0.2138535
```

```
## [1] 0.7541256
```

```
## [1] 0.2138535
```