Lab 3: 2x2 tables and related terms

January 31 - February 1, 2017

Setup

```
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
```

Making tables and R review

By default, when the summary() function encounters categorical data, it produces a table for that column, as evidenced above, where 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:
(The with() function lets us just use the column names as variables, instead of writing out titanic$ every time.)

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

```
##       Survived
## Class  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, 1 for each level:

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

```
## , , Sex = Female
##
##       Survived
## Class  Died Survived
##   1st     4      141
##   2nd    13       93
##   3rd   106       90
##   Crew    3       20
##
## , , Sex = Male
##
##       Survived
## Class  Died Survived
##   1st   118       62
##   2nd   154       25
##   3rd   422       88
##   Crew  670      192
```

```
# with(titanic, table(Class,Survived,Sex,Age))
```

It can be done with 4 and so the code is included, but it is not worth taking up space to print the tables in the lab.
(I'd recommend keeping the number of variables down to 2 or 3, as 4 is getting a bit cluttered and confusing.)

If we save a table, we can use brackets to access individual numbers [row,column]:

```
tableDemo<-with(titanic, table(Class,Survived))
print(tableDemo)
```

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

```
tableDemo[3,2] #This is the number of 3rd class passengers who survived.
```

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

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

```
prop.table(tableDemo)
```

```
##       Survived
## Class        Died   Survived
##   1st  0.05542935 0.09223080
##   2nd  0.07587460 0.05361199
##   3rd  0.23989096 0.08087233
##   Crew 0.30577010 0.09631985
```

Vocab Recap

```
##        H0 False H0 True
## Reject        A       B
## FtR           C       D
```

Type I Error

A Type I error is committed when a true null hypothesis is rejected.
In terms of disease detection (where the null hypothesis is no disease), this is a false positive.
In the table above, this is B.

Type I Error Rate (α)

The Type I error rate is the proportion of true hypotheses that were rejected.
In the table above, this is B/(B+D).

Type II Error

A Type II error is committed when a false null hypothesis is not rejected.
In the table above, this is C.

Type II Error Rate (β)

The Type II error rate is the proportion of false null hypotheses that failed to be rejected.
In the table above, this is C/(C+A).

False Discovery Rate

The false discovery rate is the fraction of null hypothesis rejections that were incorrect.
In the table above, this is B/(B+A).

Selection bias

Instead of random sampling, certain subgroups of the population were more likely to be included than others.

Nonresponse bias

Nonresponders can differ from responders in many important ways

Perception bias

The perception of benefit from a treatment (placebo effect)

Confirmation bias (touched on, but not named in notes)

The tendency to interpret new evidence as confirmation of one's existing beliefs or theories
(If doctors think that the polio vaccine causes polio, a patient with a borderline instance of disease is more likely to be diagnosed with polio if the doctor knows that the vaccine was administered.)

Confounding

The two things being studied are both highly correlated to a third thing.
(Think ice cream sales and murder rates both being related to weather.)

Weighted Averages

Weighted averages can be tricky, so here's an example:
Let's investigate sexual bias in Titanic survival.
(For the sake of practice, do this by hand.)

```
, , 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 the tables above, calculate the overall percentages of men and women who survived the Titanic.

Part b

Create a table listing the percentage of men and women who survived, broken down by class

Part c

Construct a weighted average of the percentage of male and female passengers who survived, controlling for the effect of class (i.e., report one number for men and one number for women).

```
## Part a

## Women:  0.7319149 , Men:  0.2120162

## Part b

##          Women       Men
## 1st  0.9724138 0.3444444
## 2nd  0.8773585 0.1396648
## 3rd  0.4591837 0.1725490
## Crew 0.8695652 0.2227378

## Part c

## Women:  0.754

## Men:  0.214
```