Statistics 305/605: Introduction to Biostatistical Methods for Health Sciences

R Demo for Chapter 15, part 2: Chi-Square Tests

Jinko Graham

Association and the WHI data

```r
uu <- url("http://people.stat.sfu.ca/~jgraham/Teaching/S305_17/Data/whi.csv")
WHI <- read.csv(uu)
wtab <- table(WHI)
```

- The table of proportions below gives the conditional distributions of BC status given EP status.

```r
prop.table(wtab,margin=1)
```

```
##      BC
## EP           BC-        BC+
##   EP- 0.98494199 0.01505801
##   EP+ 0.98048436 0.01951564
```

Categorical Variables in R

- R calls a categorical variable a `factor` and refers to its categories as `levels`.
- For example, the categorical variables EP and BC, for hormone replacement therapy and breast cancer status, respectively, are called `factors` by R.
- When we cross-tabulate factors, R chooses the order of the columns and rows in our table.
  - The order is set by the order of the categories in the EP and BC factors.
  - Generally the categories, or levels, of a factor are ordered alphabetically

```
wtab
```

```
##      BC
## EP     BC-  BC+
##   EP- 7980  122
##   EP+ 8340  166
```

Chi-square test for WHI example

```
cc <- chisq.test(wtab,correct=FALSE)
cc
```

```
##
##  Pearson's Chi-squared test
##
## data:  wtab
## X-squared = 4.8387, df = 1, p-value = 0.02783
```

- The argument `correct=FALSE` specifies that we do **not** want to do a continuity correction.

Another way: use a dataframe rather than a table.

- Say that we are not given the data; all we have to work with are the counts:

EP+, BC+	166	EP+, BC−	8340
EP−, BC+	122	EP−, BC−	7980

- Then we can do the chi-square test as follows:

```r
mydf <- data.frame(BCpos=c(166,122),BCneg=c(8340,7980)) #WHI data
rownames(mydf)=c("EP+", "EP-")
mydf
```

```
##     BCpos BCneg
## EP+   166  8340
## EP-   122  7980
```

```r
chisq.test(mydf, correct=FALSE)
```

```
##
##  Pearson's Chi-squared test
##
## data:  mydf
## X-squared = 4.8387, df = 1, p-value = 0.02783
```

Chi-square test with continuity correction for WHI example

```r
cc <- chisq.test(wtab) #apply the default continuity correction
cc
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  wtab
## X-squared = 4.5807, df = 1, p-value = 0.03233
```

Expected Cell Counts

- The expected cell counts under the null hypothesis of no association can be extracted from the output of `chisq.test()`.
- In general, you can find the names of an R object with `names()` and extract components with `with()`.

```
names(cc)
```

```
## [1] "statistic" "parameter" "p.value"   "method"    "data.name" "observed" 
## [7] "expected"  "residuals" "stdres"
```

```
with(cc,expected)
```

```
##      BC
## EP         BC-      BC+
##   EP- 7961.503 140.4971
##   EP+ 8358.497 147.5029
```