Chi-Square test by hand

SDS 291

4/15/2020

The Intuition of a χ^2 test and doing it by hand

There's more reading on this in Chapter 11.4. I'd read that first and then come back to this.

From Section 9.2 in the book (p.465, 1st Ed; p.425-426, 2nd Ed).

Marital Status	Boy Child	Girl Child	Total
Married	176	148	324
Did Not Marry	134	142	276
Total	310	290	600

H_0 : Single Mothers Marrying and Child Gender are independent

H_A : Single Mothers Marrying and Child Gender are related/dependent

Calculating the Expected Values for Each Cell

$E_{ij} = \frac{RowTotal \cdot ColumnTotal}{n}$

E: Expected (under the null) O: Observed

Marital Status	Boy Child	Girl Child	Total
Married	O: 176, E: $\frac{324 \cdot 310}{600} = 167.4$	O: 148, E: $\frac{324 \cdot 290}{600} = 156.6$	324
Did Not Marry	O: 134, E: $\frac{276 \cdot 310}{600} = 142.6$	O: 142, E: $\frac{276 \cdot 290}{600} = 133.4$	276
Total	310	290	600

Calculating the χ^2 test statistic

$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected}$

$\chi^2 = \frac{176-167.4}{167.4} + \frac{148-156.6}{156.6} + \frac{134-142.6}{142.6} + \frac{143-133.4}{133.4}$

$\chi^2 = 0.441816 + 0.5186536 + 0.4722861 + 0.554428 = 1.987178$

If we compare this to an F distribution with 1 df (http://gallery.shinyapps.io/dist_calc), we see that the p-value is approximately 0.158. We fail to reject the null hypothesis that the a single mother marrying is independent from her child's gender.

Testing against R Output

```
M <- as.table(rbind(c(176,148), c(134,142)))
dimnames(M) <- list(Marital = c("Married","Did Not Marry"),Child = c("Boy", "Girl"))
M
```

```
##                Child
## Marital         Boy Girl
##   Married       176  148
##   Did Not Marry 134  142
```

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

```
##
##  Pearson's Chi-squared test
##
## data:  M
## X-squared = 1.9872, df = 1, p-value = 0.1586
```

Working with Data from a DataFrame

There is also a function in the `gmodels` package called CrossTable that is helpful for this purpose when both variables are in a dataset already.

```
library(Stat2Data)
data("ICU")
gmodels::CrossTable(ICU$AgeGroup,ICU$Survive, prop.t = FALSE, prop.c = FALSE, prop.chisq = FALSE, chisq
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |-------------------------|
##
##
## Total Observations in Table:  200
##
##
##              | ICU$Survive
## ICU$AgeGroup |         0 |         1 | Row Total |
## -------------|-----------|-----------|-----------|
##            1 |         5 |        54 |        59 |
##              |     0.085 |     0.915 |     0.295 |
## -------------|-----------|-----------|-----------|
##            2 |        17 |        60 |        77 |
##              |     0.221 |     0.779 |     0.385 |
## -------------|-----------|-----------|-----------|
##            3 |        18 |        46 |        64 |
##              |     0.281 |     0.719 |     0.320 |
## -------------|-----------|-----------|-----------|
## Column Total |        40 |       160 |       200 |
## -------------|-----------|-----------|-----------|
```

```
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## ------------------------------------------------------------
## Chi^2 =  7.746722     d.f. =  2     p =  0.02078838
##
##
##
```

Proof this approach gets you the same thing as the approach we used before for the married example.

```
IC <- as.table(rbind(c(5,54), c(17,60), c(18,46)))
dimnames(IC) <- list(AgeGroup = c("Young", "Middle","Old"),Survied = c("No", "Yes"))
IC
```

```
##         Survied
## AgeGroup No Yes
##   Young   5  54
##   Middle 17  60
##   Old    18  46
```

```
chisqIC<-chisq.test(IC)
chisqIC
```

```
##
##  Pearson's Chi-squared test
##
## data:  IC
## X-squared = 7.7467, df = 2, p-value = 0.02079
```

Now You Try!

Try for yourself using the approach from the married example to see whether you can get the same χ^2 statistic by hand!