# Contingency Tables with R

```
> # Pauling's Skier Data
> rm(list=ls())
> ski = rbind(c(31,109),
+             c(17,122))
> rownames(ski) = c('Placebo','Ascorbic Acid')
> colnames(ski) = c('Cold','No Cold')
> ski # Take a look

              Cold No Cold
Placebo         31     109
Ascorbic Acid   17     122

> addmargins(ski)

              Cold No Cold Sum
Placebo         31     109 140
Ascorbic Acid   17     122 139
Sum             48     231 279

> # help(prop.table)
> prop.table(ski,margin=1) # 1=row, 2=col

                  Cold   No Cold
Placebo       0.2214286 0.7785714
Ascorbic Acid 0.1223022 0.8776978
```

$$\widehat{\mu}_{ij} = \frac{n_{i+}n_{+j}}{n} \qquad 2\sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\log\left(\frac{n_{ij}}{\widehat{\mu}_{ij}}\right) \qquad X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(n_{ij}-\widehat{\mu}_{ij})^2}{\widehat{\mu}_{ij}}$$

```
> # help(chisq.test)
> # Option correct=FALSE means no correction for continuity
> skitest = chisq.test(ski, correct=FALSE); skitest

      Pearson's Chi-squared test

data:  ski
X-squared = 4.8114, df = 1, p-value = 0.02827

> ls(skitest) # What's in the object skitest?
[1] "data.name" "expected"  "method"    "observed"  "p.value"   "parameter"
"residuals" "statistic"  "stdres"
> skitest$method
[1] "Pearson's Chi-squared test"

> muhat = skitest$expected; muhat
                 Cold No Cold
Placebo       24.08602 115.914
Ascorbic Acid 23.91398 115.086

> # Check Pearson X^2 test
> sum( (ski-muhat)^2/muhat )
[1] 4.811413

> # Likelihood ratio test
> G2 = 2*sum( ski*log(ski/muhat) ); G2
[1] 4.871697
>
> # Residuals are observed minus expected
> skitest$residuals
                  Cold    No Cold
Placebo       1.408787 -0.6421849
Ascorbic Acid -1.413846  0.6444908
```

```
> # Berkeley graduate admissions data (1973)
> # UCBAdmissions is a built-in R data set
> UCBAdmissions
, , Dept = A

          Gender
Admit      Male Female
  Admitted  512     89
  Rejected  313     19

, , Dept = B

          Gender
Admit      Male Female
  Admitted  353     17
  Rejected  207      8

, , Dept = C

          Gender
Admit      Male Female
  Admitted  120    202
  Rejected  205    391

, , Dept = D

          Gender
Admit      Male Female
  Admitted  138    131
  Rejected  279    244

, , Dept = E

          Gender
Admit      Male Female
  Admitted   53     94
  Rejected  138    299

, , Dept = F

          Gender
Admit      Male Female
  Admitted   22     24
  Rejected  351    317

> # Look at dimensions 1 and 2, summing over 3 (Department)
> AG = apply(UCBAdmissions,MARGIN=c(1,2),FUN=sum); AG
          Gender
Admit      Male Female
  Admitted 1198    557
  Rejected 1493   1278
> GenderAdmit = t(AG); GenderAdmit
        Admit
Gender   Admitted Rejected
  Male       1198     1493
  Female      557     1278
> round( prop.table(GenderAdmit,margin=1), 3) # Round to 3 decimal places
        Admit
Gender   Admitted Rejected
  Male      0.445    0.555
  Female    0.304    0.696
> chisq.test(GenderAdmit,correct=FALSE)

      Pearson's Chi-squared test

data:  GenderAdmit
X-squared = 92.205, df = 1, p-value < 2.2e-16
```

```
>
> AD = apply(UCBAdmissions,MARGIN=c(1,3),FUN=sum); AD
          Dept
Admit        A   B   C   D   E   F
  Admitted 601 370 322 269 147  46
  Rejected 332 215 596 523 437 668
> DeptAdmit = t(AD); DeptAdmit
     Admit
Dept Admitted Rejected
   A      601      332
   B      370      215
   C      322      596
   D      269      523
   E      147      437
   F       46      668
> round(prop.table(DeptAdmit,margin=1),3) # Round to 3 decimal places
     Admit
Dept Admitted Rejected
   A    0.644    0.356
   B    0.632    0.368
   C    0.351    0.649
   D    0.340    0.660
   E    0.252    0.748
   F    0.064    0.936
> chisq.test(DeptAdmit,correct=FALSE) # Necessary?

      Pearson's Chi-squared test

data:  DeptAdmit
X-squared = 778.91, df = 5, p-value < 2.2e-16

>
> # Test on sub-table, comparing Dept. D to E
> DvsE = DeptAdmit[c(4,5),] # Rows 4 and 5, all columns
> chisq.test(DvsE,correct=FALSE)

      Pearson's Chi-squared test

data:  DvsE
X-squared = 12.323, df = 1, p-value = 0.0004475

>
> choose(7,2) # Number of pairwise comparisons of departments
[1] 21
>
> apply(UCBAdmissions,MARGIN=c(2,3),FUN=sum)
        Dept
Gender     A   B   C   D   E   F
  Male   825 560 325 417 191 373
  Female 108  25 593 375 393 341
> apply(UCBAdmissions,MARGIN=c(3,2),FUN=sum)
     Gender
Dept Male Female
   A  825    108
   B  560     25
   C  325    593
   D  417    375
   E  191    393
   F  373    341
> DeptGender = apply(UCBAdmissions,MARGIN=c(3,2),FUN=sum)
>
```

```
> round(prop.table(DeptGender,margin=1),3)
    Gender
Dept  Male Female
   A 0.884  0.116
   B 0.957  0.043
   C 0.354  0.646
   D 0.527  0.473
   E 0.327  0.673
   F 0.522  0.478
> round(prop.table(DeptAdmit,margin=1),3) # Repeat for comparison
    Admit
Dept Admitted Rejected
   A    0.644    0.356
   B    0.632    0.368
   C    0.351    0.649
   D    0.340    0.660
   E    0.252    0.748
   F    0.064    0.936
> addmargins(DeptGender)
     Gender
Dept  Male Female  Sum
  A    825    108  933
  B    560     25  585
  C    325    593  918
  D    417    375  792
  E    191    393  584
  F    373    341  714
  Sum 2691   1835 4526
> # Departments A and C are REALLY big.
>
>
> UCBAdmissions # Look at the 3-d table again
, , Dept = A

          Gender
Admit      Male Female
  Admitted  512     89
  Rejected  313     19

, , Dept = B

          Gender
Admit      Male Female
  Admitted  353     17
  Rejected  207      8

, , Dept = C

          Gender
Admit      Male Female
  Admitted  120    202
  Rejected  205    391

, , Dept = D

          Gender
Admit      Male Female
  Admitted  138    131
  Rejected  279    244

, , Dept = E

          Gender
Admit      Male Female
  Admitted   53     94
  Rejected  138    299
```

```
, , Dept = F

          Gender
Admit      Male Female
  Admitted   22     24
  Rejected  351    317

> # I want rows of sub-tables to be M vs F
>
> UCB = as.data.frame(UCBAdmissions); UCB
      Admit Gender Dept Freq
1  Admitted   Male    A  512
2  Rejected   Male    A  313
3  Admitted Female    A   89
4  Rejected Female    A   19
5  Admitted   Male    B  353
6  Rejected   Male    B  207
7  Admitted Female    B   17
8  Rejected Female    B    8
9  Admitted   Male    C  120
10 Rejected   Male    C  205
11 Admitted Female    C  202
12 Rejected Female    C  391
13 Admitted   Male    D  138
14 Rejected   Male    D  279
15 Admitted Female    D  131
16 Rejected Female    D  244
17 Admitted   Male    E   53
18 Rejected   Male    E  138
19 Admitted Female    E   94
20 Rejected Female    E  299
21 Admitted   Male    F   22
22 Rejected   Male    F  351
23 Admitted Female    F   24
24 Rejected Female    F  317

> GenderAdmitDept = xtabs(Freq ~ Gender + Admit + Dept, data=UCB)
> GenderAdmitDept
, , Dept = A

        Admit
Gender   Admitted Rejected
  Male        512      313
  Female       89       19

, , Dept = B

        Admit
Gender   Admitted Rejected
  Male        353      207
  Female       17        8

, , Dept = C

        Admit
Gender   Admitted Rejected
  Male        120      205
  Female      202      391

, , Dept = D

        Admit
Gender   Admitted Rejected
  Male        138      279
  Female      131      244
```

```
, , Dept = E

        Admit
Gender   Admitted Rejected
  Male         53      138
  Female       94      299

, , Dept = F

        Admit
Gender   Admitted Rejected
  Male         22      351
  Female       24      317
```

```
> round(prop.table(GenderAdmitDept, margin=c(1,3)), 3) # Trial and error
, , Dept = A

        Admit
Gender   Admitted Rejected
  Male      0.621    0.379
  Female    0.824    0.176

, , Dept = B

        Admit
Gender   Admitted Rejected
  Male      0.630    0.370
  Female    0.680    0.320

, , Dept = C

        Admit
Gender   Admitted Rejected
  Male      0.369    0.631
  Female    0.341    0.659

, , Dept = D

        Admit
Gender   Admitted Rejected
  Male      0.331    0.669
  Female    0.349    0.651

, , Dept = E

        Admit
Gender   Admitted Rejected
  Male      0.277    0.723
  Female    0.239    0.761

, , Dept = F

        Admit
Gender   Admitted Rejected
  Male      0.059    0.941
  Female    0.070    0.930
```

```
>
> # Tests on sub-tables: Gender by Admission
# Remember
#   A: Higher proportion of women admitted
#   B: Slightly higher proportion of women admitted
#   C: Slightly higher proportion of men admitted
#   D: Slightly higher proportion of women admitted
#   E: Slightly higher proportion of men admitted
#   F: Slightly higher proportion of women admitted

> dept = as.character(sort(unique(UCB$Dept))); dept
[1] "A" "B" "C" "D" "E" "F"

> results = matrix(NA,6,2) # 6x2 matrix of NAs
> rownames(results) = dept; colnames(results) = c('X-squared','p-value')
>
> for(k in 1:6)
+     {
+      x2test = chisq.test(GenderAdmitDept[,,k],correct=FALSE)
+      results[k,1] = x2test$statistic; results[k,2] = x2test$p.value
+     }
> options(scipen=999) # To suppress scientific notation

> results
   X-squared       p-value
A 17.2480134 0.00003280404
B  0.2537215 0.61446676567
C  0.7535389 0.38535809298
D  0.2979776 0.58515307222
E  1.0010686 0.31705206682
F  0.3840933 0.53542068131
>
> # Test of conditional independence
> X2pooled = sum(results[,1]) # Sum of chi-squared values
> pval = 1-pchisq(X2pooled,6)
> c(X2pooled,pval)
[1] 19.938413378  0.002840164
```

---

This document was prepared by Jerry Brunner, University of Toronto. It is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License: http://creativecommons.org/licenses/by-sa/3.0/deed.en_US. Use any part of it as you like and share the result freely. It is available in OpenOffice.org format from the course website: http://www.utstat.toronto.edu/brunner/oldclass/312f22