Math 3080 § 1. **Telecommuting Example:** Name: Example
Treibergs **Chi-Squared Test for Independence** April 8, 2014

In this **R**© program, we run a χ^2 tests for independence. The data is from the article "No Shows," *American Demographics,* 2003 as quoted by Mendenhall, Beaver & Beaver, *Introduction to Probability and Statistics* 14th ed., Brooks Cole, Boston, 2013. The article concluded that "people who work exclusively at home tend to be ... better educated than those who have to leave home to report to work." The authors polled a random sample of 300 workers. The category "Mixed" are those who reported working at home for at least one full day in a typical week.

Education	Non-home	Mixed	Home
Less than H.S. Diploma	23	3	5
H.S. Graduate	54	12	11
Some College/Assoc. Degree	53	24	14
B.A. or more	41	42	18

We run chi-squared tests of independence. Let us assume that $p(x,y)$ is the joint pmf and that there are I levels of the education factor X and J levels of the work factor Y. The null and alternative hypotheses are

$$\mathcal{H}_0 : \text{The factors are independent: } p(x,y) = p_X(x)p_Y(y) \text{ for all } x = 1,\dots,I \text{ and all } y = 1,\dots,J$$
$$\mathcal{H}_a : \text{The factors are dependent: } \mathcal{H}_0 \text{ is not true}$$

where $p_X(x)$ and $p_Y(y)$ are marginal probabilities. The test statistic

$$\chi^2 = \sum_{i=1}^{3}\sum_{j=1}^{3} \frac{(n_{ij} - \hat{e}_{ij})^2}{\hat{e}_{ij}} = 10.8286$$

where the estimated expected frequency in each cell is given by $\hat{e}_{ij} = \dfrac{n_{i\cdot}\, n_{\cdot j}}{n_{\cdot\cdot}}$. The null hypothesis is rejected if $\chi^2 \ge \chi^2_{\alpha,(I-1)(J-1)}$.

Here $I = 4$ and $J = 3$. At the $\alpha = .05$ significance level $\chi^2_{\alpha,(I-1)(J-1)} = \chi^2_{.05,6} = 12.592$ from Table A7. The statistic works out to be $\chi^2 = 24.23$ so we reject the null hypothesis: there is highly significant evidence to indicate the two variables are not independent. The p-value was 0.0004728. There was a warning that some expected cell counts were below five. Looking at the table of expected counts, we see that only one cell was under, and that was 4.96, which is very close to five and the others were much more, so we are willing to believe in the validity of the large sample assumption.

To see if we agree with the researchers about their conclusion that the telecommuters were more highly educated, we computed the marginal proportions in each work category. The proportion of B.A. or more in the Mixed category is higher than in the others, but otherwise the proportions are not so different. We tried to see if lumping the Mixed and Home together. The proportion of B.A. or more in telecommuting group is 46.5% whereas it is 24.0% in the telecommuting group. Thus we agree with the summary assessment of the researchers.

R Session:

```
R version 2.13.1 (2011-07-08)
Copyright (C) 2011 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: i386-apple-darwin9.8.0/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.41 (5874) i386-apple-darwin9.8.0]

[History restored from /Users/andrejstreibergs/.Rapp.history]

> ################# ENTER TABLE  ######################################
> M=matrix(c(23,3,5,54,12,11,53,24,14,41,42,18),ncol=3,byrow=T)
> colnames(M)=c("Non-home","Mixed","Home")
> rownames(M)=c("Less than H.S. Diploma","H.S. Graduate",
                "Some College/Assoc. Degree","B.A. or More")
> names(dimnames(M))=c("Age","Workers")
> M
                           Workers
Age                         Non-home Mixed Home
  Less than H.S. Diploma           23     3    5
  H.S. Graduate                    54    12   11
  Some College/Assoc. Degree       53    24   14
  B.A. or More                     41    42   18

> ############  RUN CHI-SQ TEST OF INDEPENDENCE  ########################

> t1 = chisq.test(M); t1

Pearson's Chi-squared test

data:  M
X-squared = 24.2346, df = 6, p-value = 0.0004728

Warning message:
In chisq.test(M) : Chi-squared approximation may be incorrect
```

```
> ################  CHECK EXPECTED CELL COUNTS  ########################
> t1$expected
                           Workers
Age                         Non-home Mixed  Home
  Less than H.S. Diploma        17.67  8.37  4.96
  H.S. Graduate                 43.89 20.79 12.32
  Some College/Assoc. Degree    51.87 24.57 14.56
  B.A. or More                  57.57 27.27 16.16

> ####  COMPUTE COLUMN TOTALS TO GET ED. PROP IN EACH WORK CATEGORY  ##
> colsum=margin.table(M,2)
> colsum
Workers
Non-home    Mixed     Home
     171       81       48
> nidot=matrix(rep(colsum,times=4),ncol=3,byrow=T); nidot
     [,1] [,2] [,3]
[1,]  171   81   48
[2,]  171   81   48
[3,]  171   81   48
[4,]  171   81   48

> ############## PROPORTION EDucation IN EACH WORK CAT.  ###############
> M/nidot
                           Workers
Age                          Non-home      Mixed      Home
  Less than H.S. Diploma     0.1345029 0.03703704 0.1041667
  H.S. Graduate              0.3157895 0.14814815 0.2291667
  Some College/Assoc. Degree 0.3099415 0.29629630 0.2916667
  B.A. or More               0.2397661 0.51851852 0.3750000
> margin.table(M/nidot,2)
Workers
Non-home    Mixed     Home
       1         1        1
```

```
> ###  LUMP HOME AND MIXED TOGETHER AND RECOMPUTE PROPS  ############
> M2=cbind(M[,1],M[,2]+M[,3]); M2
                           [,1] [,2]
Less than H.S. Diploma       23    8
H.S. Graduate                54   23
Some College/Assoc. Degree   53   38
B.A. or More                 41   60

> colsum2=margin.table(M2,2); colsum2
[1] 171 129

> #######  PROP. EDUCATION IN NONHOME & TELECOMMUTING GROPS  ####
> M2/matrix(rep(colsum2,times=4),ncol=2,byrow=T)
                                [,1]      [,2]
Less than H.S. Diploma     0.1345029 0.0620155
H.S. Graduate              0.3157895 0.1782946
Some College/Assoc. Degree 0.3099415 0.2945736
B.A. or More               0.2397661 0.4651163
>
```