In this **R**© program, we run a χ^2 test for homogeneity. The data is from the article by Rosenman and Hertzberg *et al.*, "Chronic Beryllium Disease and Sensitization at a Beryllium Processing Facility," in *Environmental Health Perspectives,* 2005; as quoted by Navidi, *Statistics for Engineers and Scientists,* 2nd ed., Mc Graw Hill, New York, 2008. The authors discuss the effects of exposure to beryllium in a cohort of workers. Workers were categorized by their duration of exposure (in years) and by their disease status (chronic beryllium disese, sensitization to beryllium, or no disease). Can we conclude that the proportion of workers in the various disease categories differ among exposure levels? The results were as follows

	Duration	of	Exposure
Disease Status	< 1Yr.	$1 \le$ Yr. < 5	$5 \le$ Yrs.
Diseased	10	8	23
Sensitized	9	19	11
Normal	70	136	206

We run our chi-squared test of Homogeneity. The null hypothesis is that the proportion of durations is the same for each level of disease. 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

$$\hat{e}_{ij} = \frac{n_{i.}\, n_{.j}}{n_{..}}$$

Since the $\alpha = .05$ critical value is $\chi^2_{\alpha,(I-1)(J-1)} = \chi^2_{.05,4} = 9.487729$, we reject the null hypothesis: there is significant evidence to indicate that workers having different disease levels received different levels of exposure. The p-value was 0.02856.

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 THE CONTINGENCY TABLE AS MATRIX  ######

> exposure.disease=matrix(c(10,8,23,9,19,11,70,136,206),nrow=3,byrow=T)
> rownames(exposure.disease)=c("Diseased","Sensitized","Normal")
> colnames(exposure.disease)=c("<1 Yr","1  Yrs < 5","5  Yrs")
> names(dimnames(exposure.disease))=c("Disease Status","Duration of Exposure")
> exposure.disease
               Duration of Exposure
Disease Status <1 Yr 1  Yrs < 5 5  Yrs
    Diseased       10          8      23
    Sensitized      9         19      11
    Normal         70        136     206
>
> ###########  AS A DATA FRAME OF COUNTS  ####################
> as.data.frame(as.table(exposure.disease))
  Disease.Status Duration.of.Exposure Freq
1       Diseased                <1 Yr   10
2     Sensitized                <1 Yr    9
3         Normal                <1 Yr   70
4       Diseased            1  Yrs < 5    8
5     Sensitized            1  Yrs < 5   19
6         Normal            1  Yrs < 5  136
7       Diseased               5  Yrs   23
8     Sensitized               5  Yrs   11
9         Normal               5  Yrs  206
>
> #############  RUN THE CHI SQUARED TEST OF HOMOGENEITY  #####
> X2<-chisq.test(exposure.disease)
> X2

	Pearson's Chi-squared test

data:  exposure.disease
X-squared = 10.8286, df = 4, p-value = 0.02856
```

```
> ############   CHI-SQUARED TEST "BY HAND" ####################

> ########  ROW SUMS, COLUMN SUMS, GRAND TOTAL   #############
> rowsum = margin.table(exposure.disease,1); rowsum
Disease Status
  Diseased Sensitized     Normal
        41          39           412
> colsum = margin.table(exposure.disease,2); colsum
Duration of Exposure
     <1 Yr 1  Yrs < 5     5  Yrs
         89           163            240
> n=sum(exposure.disease); n
[1] 492

> ############    TABLE OF EXPECTED FREQUENCIES  ###############

> outer(rowsum, colsum)/n
              Duration of Exposure
Disease Status     <1 Yr 1  Yrs < 5    5  Yrs
    Diseased    7.416667    13.58333  20.00000
    Sensitized  7.054878    12.92073  19.02439
    Normal     74.528455   136.49593 200.97561
> X2$expected
              Duration of Exposure
Disease Status     <1 Yr 1  Yrs < 5    5  Yrs
    Diseased    7.416667    13.58333  20.00000
    Sensitized  7.054878    12.92073  19.02439
    Normal     74.528455   136.49593 200.97561

> X2$observed
              Duration of Exposure
Disease Status <1 Yr 1  Yrs < 5 5  Yrs
    Diseased       10           8        23
    Sensitized      9          19        11
    Normal         70         136       206


> ################  CONTRIBUTION TO CHI-SQ FROM EACH CELL  ###

> E <- X2$expected;  O <- X2$observed;
> (E-O)^2/E
              Duration of Exposure
Disease Status      <1 Yr 1  Yrs < 5    5  Yrs
    Diseased    0.8998127 2.294989775 0.4500000
    Sensitized 0.5362955 2.860325856 3.3846467
    Normal      0.2751554 0.001801896 0.1256098
```

```
> ################  CHI-SQ  SUM  ##########################
> ch2=sum((O-E)^2/E); ch2
[1] 10.82864

> ######  CRITICAL VALUE FOR ALPHA=.05, (I-1)(J-1)=4  ########
> qchisq(.05,4,lower.tail=F)
[1] 9.487729
> ################  SO REJECT H0  #########################

> #################  P-VALUE  ############################
> pchisq(ch2,4,lower.tail=F)
[1] 0.02855892

>
```