Lab 9 χ^2

Senie et al. (1981) investigated the relationship between age and frequency of breast self-examination in a sample of women (Senie, R. T., Rosen, P. P., Lesser, M. L., and Kinne, D. W. Breast self-examinations and medical examination relating to breast cancer stage. *American Journal of Public Health*, **71**, 583-590.)

A summary of the results is presented in the following table:

Age	Monthly	Occassionally	Never
under 45	91	90	51
45 – 59	150	200	155
60 and over	109	198	172

The data have already been tabled for us in most textbook problems. We just have to get the data into an R data object. There are several ways to do this...

```
> row1 = c(91,90,51)                    # or col1 = c(91,150,109)
> row2 = c(150,200,155)                 # and col2 = c(90,200,198)
> row3 = c(109,198,172)                 # and col3 = c(51,155,172)
> data.table = rbind(row1,row2,row3)    # and data.table = cbind(col1,col2,col3)
data.table
     [,1] [,2] [,3]
row1   91   90   51
row2  150  200  155
row3  109  198  172
> chisq.test(data.table)

  Pearson's Chi-squared test

data:  data.table
X-squared = 25.086, df = 4, p-value = 4.835e-05
```

Capt Jim takes a random sample of students enrolled in Statistics 2228 at UNH. He finds the following: there are 25 freshman in the sample, 32 sophomores, 18 juniors, and 20 seniors. Test the null hypothesis that freshman, sophomores, juniors, and seniors are equally represented among students signed up for Stat 2228. This is a goodness of fit test with equal expected frequencies. The "p" vector does not need to be specified, since equal frequencies is the default...

```
> chisq.test(c(25,32,18,20))

Chi-squared test for given probabilities

data:  c(25, 32, 18, 20)
X-squared = 4.9158, df = 3, p-value = 0.1781
```

You could also have begun by assigning the observed frequencies to a vector, and then have used the vector name as "x"...

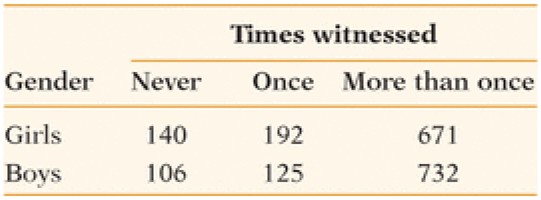
```
> ofs <- c(25,32,18,20)
> chisq.test(ofs)

Chi-squared test for given probabilities

data:  ofs
X-squared = 4.9158, df = 3, p-value = 0.1781
```

Either way, the null hypothesis cannot be rejected at alpha = 0.05.

Lab 9 On Your Own

Name ___________________ Score ________

1. **Sexual harassment in middle and high schools.** A nationally representative survey of students in grades 7 to 12 asked about the experience of these students with respect to sexual harassment.[11] One question asked how many times the student had witnessed sexual harassment in school. Here are the data categorized by gender:



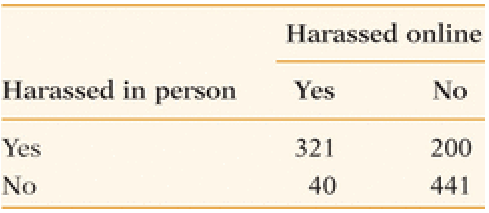
	Times witnessed		
Gender	Never	Once	More than once
Girls	140	192	671
Boys	106	125	732

a) State Ho and Ha

b) Use R to find the test statistic, χ^2 . What is the p-value of the test statistic?

c) Is there a relationship between gender and harassment?

2. **Sexual harassment online or in person.** In the study described above, the students were also asked whether or not they were harassed in person and whether or not they were harassed online. Here are the data for the girls:

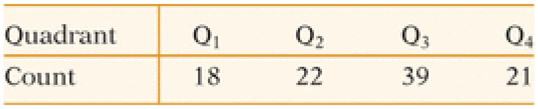


	Harassed online	
Harassed in person	Yes	No
Yes	321	200
No	40	441

a) State Ho and Ha.

b) Using R analyze these data for examining a relationship between two categorical variables in a 2 × 2 table. State your conclusion.

3. **Is there a random distribution of trees?** In Example 6.1 (page 352) we examined data concerning the longleaf pine trees in the Wade Tract and concluded that the distribution of trees in the tract was not random. Here is another way to examine the same question. First, we divide the tract into four equal parts, or quadrants, in the east–west direction. Call the four parts Q_1 to Q_4. Then we take a random sample of 100 trees and count the number of trees in each quadrant. Here are the data:



Quadrant	Q_1	Q_2	Q_3	Q_4
Count	18	22	39	21

a) If the trees are randomly distributed, we expect to find 25 trees in each quadrant. Why? Explain your answer.

b) We do not really expect to get *exactly* 25 trees in each quadrant. Why? Explain your answer.

c) Using R, perform the goodness-of-fit test for these data to determine if these trees are randomly scattered. Write a short report giving the details of your analysis and your conclusion.