Verzani Problem Set

Next are considered the problems from Verzani's book on page 76.

Problem 12.1

In an effort to increase student retention, many colleges have tried block programs. Suppose 100 students are broken into two groups of 50 at random. One half are in a block program, the other half not. The number of years in attendance is then measured. We wish to test if the block program makes a difference in retention. The data is:

Program	1 yr	2 yr	3 yr	4 yr	5+ yr
Non-Block	18	15	5	8	4
Block	10	5	7	18	10

Do a test of hypothesis to decide if there is a difference between the two types of programs in terms of retention.

```
> nonBlock <- c(18, 15, 5, 8 ,4)
> block <- c(10, 5, 7, 18, 10)
> chisq.test(rbind(nonBlock, block))

    Pearson's Chi-squared test

data:  rbind(nonBlock, block)
X-squared = 14.037, df = 4, p-value = 0.007179
```

The $p-value = 0.007179 < 0.05 = \alpha$, so we reject H_0. The block programs makes a difference in a retention.

Problem 12.2

A survey of drivers was taken to see if they had been in an accident during the previous year, and if so was it a minor or major accident. The results are tabulated by age group:

Age \ Accident Type	None	Minor	Major
under 18	67	10	5
18 - 25	42	6	5
26 - 40	75	8	4
41 - 65	56	4	6
over 65	57	15	1

Do a chi-squared hypothesis test of homogeneity to see if there is difference in distributions based on age.

```
> under18 <- c(67, 10, 5)
> between18and25 <- c(42, 6, 5)
> between26and40 <- c(75, 8, 4)
> between40and65 <- c(56, 4, 6)
> over65 <- c(57, 15, 1)
> chisq.test(rbind(under18, between18and25,
between26and40, between40and65, over65))
Warning in chisq.test(rbind(under18, between18and25,
between26and40,
between40and65, : Chi-squared approximation may be
incorrect

    Pearson's Chi-squared test

data:  rbind(under18, between18and25, between26and40,
between40and65,      over65)
X-squared = 12.586, df = 8, p-value = 0.1269
```

The $p-value = 0.1269 > 0/05 = \alpha$, so we have no evidence to reject H_0. The age does not influence the accident type.

Problem 12.3

A fish survey is done to see if the proportion of fish types is consistent with previous years. Suppose, the 3 types of fish recorded: parrotfish, grouper, tang are historically in a $5:3:4$ proportion and in a survey the following counts are found

Parrotfish	Grouper	Tang
53	22	49

Do a test of hypothesis to see if this survey of fish has the same proportions as historically.

We perform goodness of fit test

```
> freq <- c(53, 22, 49)
> prob <- c(5, 3, 4) / 12
> chisq.test(freq, p = prob)

    Chi-squared test for given probabilities

data:  freq
X-squared = 4.0694, df = 2, p-value = 0.1307
```

The $p-value = 0.1307 > 0.05 = \alpha$, so we have no evidence to reject H_0. This survey of fish have the same proportion as historically observed.

Problem 12.4

The R data set `UCBAdmissions` contains data on admission to UC Berkeley by gender. We wish to investigate if the distribution of males admitted is similar to that of females. To do so, we need to first do some spade work as the data set is presented in a complex contingency table. The `ftable` (flatten table) command is needed. To use it try

```
> library(UsingR)
Warning: package 'UsingR' was built under R version 4.0.3
Loading required package: MASS
Loading required package: HistData
Loading required package: Hmisc
Loading required package: lattice
Loading required package: survival
Loading required package: Formula
Loading required package: ggplot2

Attaching package: 'Hmisc'
```

```
The following objects are masked from 'package:base':

    format.pval, units

Attaching package: 'UsingR'
The following object is masked from 'package:survival':

    cancer
> x = ftable(UCBAdmissions)
> x
                Dept   A   B   C   D   E   F
Admit    Gender
Admitted Male        512 353 120 138  53  22
         Female       89  17 202 131  94  24
Rejected Male        313 207 205 279 138 351
         Female       19   8 391 244 299 317
```

We want to compare rows 1 and 2. Treating x as a matrix, we can access these with x[1:2,]. Do a test for homogeneity between the two rows. What do you conclude? Repeat for the rejected group.

```
> chisq.test(x[1:2,])

    Pearson's Chi-squared test

data:  x[1:2, ]
X-squared = 463.09, df = 5, p-value < 2.2e-16
```

The $p-value = 2.2e-16 < 0.05 = \alpha$, so we reject H_0. The difference in the admitted men and women is statistically significant.

```
> chisq.test(x[3:4,])

    Pearson's Chi-squared test

data:  x[3:4, ]
X-squared = 552.62, df = 5, p-value < 2.2e-16
```

The $p-value < 2.2e-16 < 0.05 = \alpha$, so we reject H_0. The difference in the rejected men and women is statistically significant.