Project 10 - Solutions

1. From the *Bozeman Daily Chronicle Article* from October 27, 2006, "Attorneys argue over unsealing homicide documents" available at the STAT401 web site. Table 1 gives the observed counts of respondents in category as given by the defendants' attorney's:

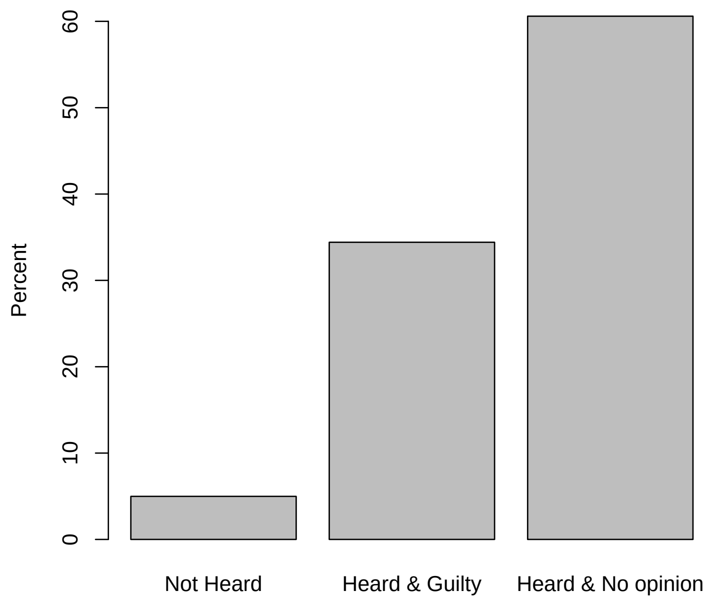
   Table 1: The observed counts of 401 respondents about Jason Wright homicide

Had not Heard of Case	Heard of Case and Guilty	Heard of Case and No Opinion	Total
20	148	233	401

   A Goodness of Fit Test to determine whether the percentage of people who have heard of the case is 90%, and that these people are evenly split on whether or not the defendants are guilty, is:

   (a) Figure 1 shows a bar chart of the results of the 401 respondents about Jason Wright's homicide.

   Figure 1: The observed distribution of respondents to a survey about Jason Wright homicide

   (b) The hypotheses are:
   $H_0 : \pi_1 = .1,\ \pi_2 = .45,$ and $\pi_3 = .45$
   H_a : At least one of the following are true: $\pi_1 \neq .1$ or $\pi_2 \neq .45$ or $\pi_3 \neq .45$

   (c) The attorneys claim that we have a SRS of Gallatin county residents. As seen in Table 2, the expected counts, computed via $n\pi_{i_0}$, are all greater than 5.

   (d) **Table 2: The expected counts of 401 respondents about Jason Wright's homicide if H_0 were true**

Had not Heard of Case	Heard of Case and Guilty	Heard of Case and No Opinion	Total
40.10	180.45	180.45	401

(e) Table 3 gives the χ^2 contributions. The i^{th} contribution is computed via $X_i^2 = \frac{((\text{observed count})_i - (\text{expected count})_i)^2}{(\text{expected count})_i}$:

Table 3: The χ^2 contributions for the three categories of respondents about Jason Wright's homicide

Had not Heard of Case	Heard of Case and Guilty	Heard of Case and No Opinion
10.0751	5.835	15.3034

(f) The test statistic is $X_i^2 = \sum_i \frac{((\text{observed count})_i - (\text{expected count})_i)^2}{(\text{expected count})_i} = 31.2139$.

(g) The distribution of the test statistic assuming H_0 is true is $X^2 \dot{\sim} \chi^2(2)$.

(h) The p-value is $P(X^2 > x^2) = 1.667 \times 10^{-7}$.

(i) Since the p-value is tiny, then we reject H_0.

(j) The preponderance of evidence suggests that the distribution of responses to the survey about Jason Wright's homicide is something different than P(Not Heard)=10%, P(Heard & Guilty)=45% and P(Heard and No Opinion)=45%.

(k) A follow-up is necessary since we rejected H_0 in the Goodness of Fit Test. To perform a follow-up, observe the χ^2 contributions given in Table 3. Since the largest contribution is for the category "Heard and No Opinion" and the observed count is greater than the expected count for this category, it appears that this proportion is significantly larger than 45%.

2. In November, 1990, the *Student Right To Know and Campus Security Act* was signed into law. This law has been expanded in subsequent years and is widely referred to as the *Clery Act*. The Act mandates that institutions of higher education report and make available to students and employees the reported occurrences of specific types of crime which occur on campus. Table 4, excerpted from the report *Montana State University - Bozeman Calender Year 2002 Annual Crime Report*, gives the observed counts of reported Burglaries and Forcible Sex Offences for three different years:

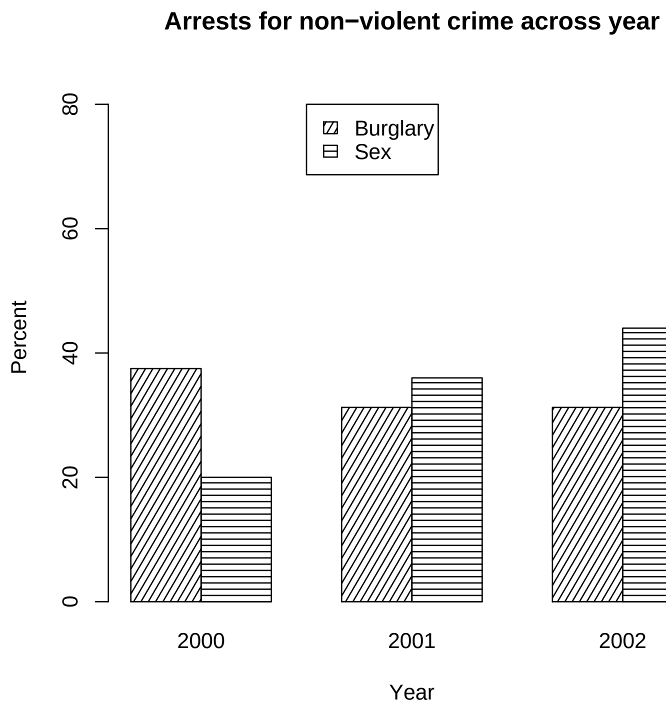
Table 4: Observed Counts of Burglaries and Sex Offenses at MSU-Bozeman

OFFENSE	2000	2001	2002
Burglary	12	10	10
Sex-Offense-Forcible	5	9	11

A Test of Independence will determine whether there is an association between the type of offense and the year:

(a) Figure 2 shows a bar chart of the proportion of crimes versus year.

Figure 2:



(b) The hypotheses are:
H_0 : Type of Crime and Year are independent
H_a : There is an association between Type of Crime and Year

(c) The data consists of ALL reported burglaries and forcible sex-offenses for 2000-2002, and so it is not a SRS. That is, the assumption that $.05N > n$ is violated. To check the second assumption, Table 5 shows that the expected counts, computed via (row total)(column total)/57 are all larger than 5.

(d) **Table 5: Expected Counts of Burglaries and Sex Offenses at MSU-Bozeman**

OFFENSE	2000	2001	2002
Burglary	9.54	10.67	11.79
Sex-Offense-Forcible	7.46	8.33	9.21

(e) Table 6 gives the χ^2 contributions:

Table 6: χ^2 contributions for Burglaries and Sex Offenses at MSU-Bozeman

OFFENSE	2000	2001	2002
Burglary	0.6321	0.0417	0.2716
Sex-Offense-Forcible	0.8091	0.0533	0.3477

(f) The test statistic value is the sum of the contributions from Table 6, $X^2 = 2.1555$.

(g) The distribution of the test statistic assuming H_0 is true is $X^2 \dot{\sim} \chi^2(2)$.

(h) The p-value is $P(X^2 > x^2) = 0.3404$.

(i) Since the p-value $= 0.3404 . \alpha = .05$, then FAIL TO REJECT H_0.

(j) The evidence fails to suggest that there is an association between crime type and year.

(k) A follow-up is not necessary since we failed to reject H_0 in the χ^2 test

3. In March 25, 2005, *Animal Behavior* published Creel and Winnie's "Responses of elk herd size to fine scale spatial and temporal variation in the risk of predation by wolves." In the paper, Creel and Winnie are studying how far from the timber elk herds aggregate when there are wolves present. Table 7 shows the observed numbers bulls, cows and calves at three different ranges of distance to the timber when there are wolves present:

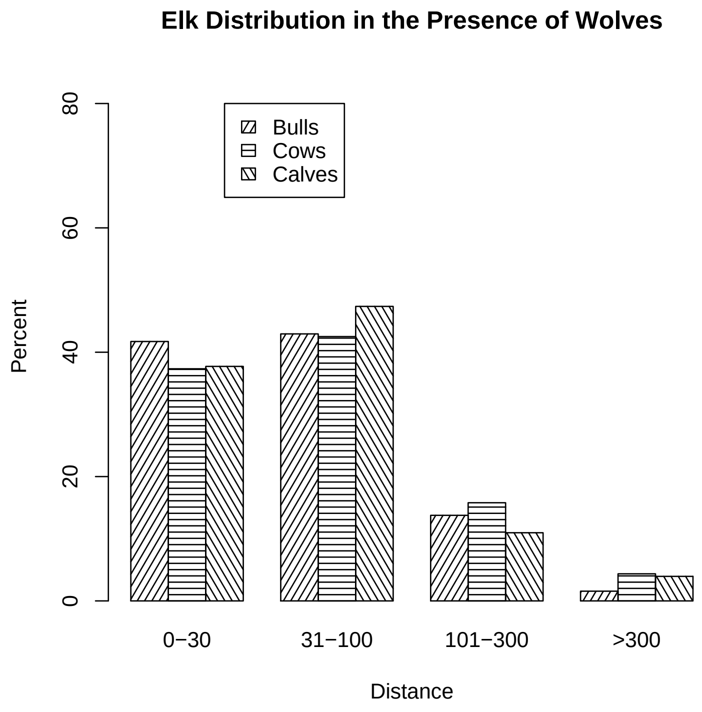
Table 7: Observed Number of Elk at different distances to the timber when wolves are present

	0-30	31-100	101-300	> 300	Totals
Bulls	479	493	158	18	1148
Cows	454	517	192	53	1216
Calves	86	108	25	9	228

A Test of Homogeneity will determine whether the distributions of the distances from the timber are different for all three populations of elk:

(a) Figure 3 shows a bar chart comparing the observed distance distribution for each population of elk.

Figure 3:



(b) The hypotheses are:

H_0 : The distribution of the distance to the timber is the same for all three populations of elk.

H_a : The distribution of the distance to the timber is different for at least one of the three populations of elk.

(c) By assumption, we have SRS's from each elk population. Table 8 shows that the expected counts are all above five.

(d) **Table 8: Expected Number of Elk at different distances to the timber when wolves are present**

	0-30	31-100	101-300	> 300
Bulls	451.32	495.16	166.09	35.43
Cows	478.05	524.49	175.93	37.53
Calves	89.63	98.34	32.99	7.04

(e) Table 9 gives the χ^2 contributions

Table 9: χ^2 contributions for the 3 populations of Elk at different distances to the timber when wolves are present

	0-30	31-100	101-300	> 300
Bulls	1.70	0.0095	0.3939	8.58
Cows	1.21	0.1071	1.47	6.38
Calves	0.1474	0.9484	1.93	0.5476

(f) The sum of the χ^2 contributions is $X^2 = 23.42$.

(g) The distribution of the test statistic assuming H_0 is true is $X^2 \dot{\sim} \chi^2(6)$.

(h) The p-value is $P(X^2 > x^2) = 0.0006685$.

(i) Since the p-value $< \alpha = .05$, then we REJECT H_0.

(j) The evidence suggests that there is a difference in the distributions distance to the timber for the three different populations of elk.

(k) A follow-up is necessary since H_0 in the chi^2 test was rejected. Looking at the two highest χ^2 contributions, we see that there is a smaller proportion of bull elk at more than 300 yards from the timber and more cow elk at more than 300 yards from the timber.

Appendix

```
> # PROBLEM 1
> freq = c(20,148,233)
> freq.prop = prop.table(freq)*100
> barplot(freq.prop,names=c("Not Heard","Heard and Guilty","Heard and No opinion"),ylab="
> hyp.p=c(0.10,0.45,0.45)
> wright.csq=chisq.test(freq,p=hyp.p)
> wright.csq

        Chi-squared test for given probabilities

data:  freq
X-squared = 31.2139, df = 2, p-value = 1.667e-7


> wright.csq$expected
[1]  40.10 180.45 180.45

> (wright.csq$obs-wright.csq$expected)^2/wright.csq$expected
[1] 10.0751  5.8354 15.3034
```

```
> PROBLEM #2
> freq = matrix(c(12,10,10,5,9,11),ncol=3,byrow=TRUE)
>
> colnames(freq) = c("2000","2001","2002")
> rownames(freq) = c("Burglary", "Sex")
> freq.prop = prop.table(freq,1)*100
> ang=c(60,180)
> barplot(freq.prop,beside=TRUE,angle=ang,density=20,col="black",xlab="Year",ylab="Percen
>
> legend(3.5,80,fill=TRUE,legend=rownames(freq),angle=ang,density=20,merge=TRUE,bg="white
>
>
> crime.csq = chisq.test(freq)
> crime.csq

        Pearson's Chi-squared test

data:  freq
X-squared = 2.1555, df = 2, p-value = 0.3404


> crime.csq$expected
            2000      2001      2002
Burglary 9.54386 10.666667 11.789474
```

```
Sex      7.45614  8.333333  9.210526

> (crime.csq$obs-crime.csq$expected)^2/crime.csq$expected
             2000       2001      2002
Burglary 0.6320949 0.04166667 0.2716165
Sex      0.8090815 0.05333333 0.3476692
>



> PROBLEM #3
> freq = matrix(c(479,493,158,18,454,517,192,53,86,108,25,9),ncol=4,byrow=TRUE)
>
> colnames(freq) = c("0-30","31-100","101-300",">300")
> rownames(freq) = c("Bulls", "Cows", "Calves")
> freq.prop = prop.table(freq,1)*100
> ang=c(60,180,120)
> barplot(freq.prop,beside=TRUE,angle=ang,density=20,col="black",xlab="Distance",ylab="Pe
>
> legend(3.5,80,fill=TRUE,legend=rownames(freq),angle=ang,density=20,merge=TRUE,bg="white
>
>
> elk.csq = chisq.test(freq)
> elk.csq

        Pearson's Chi-squared test

data:  freq
X-squared = 23.4161, df = 6, p-value = 0.0006685


> elk.csq$expected
            0-30   31-100   101-300      >300
Bulls  451.31636 495.1636 166.08796 35.432099
Cows   478.04938 524.4938 175.92593 37.530864
Calves  89.63426  98.3426  32.98611  7.037037


> (elk.csq$obs-elk.csq$expected)^2/elk.csq$expected
            0-30      31-100   101-300      >300
Bulls  1.6981083 0.009453602 0.3938584 8.5763496
Cows   1.2098600 0.107069793 1.4686628 6.3759300
Calves 0.1473526 0.948373592 1.9334795 0.5475634
```