Problem 1: Assume that 50% of the US population votes. Of those who vote, 50% vote Democrat and 50% vote Republican. From our survey, 32 report that they do not (or cannot) vote. Of the students who do vote, 71 students vote Democrat and 5 students vote Republican. Are these observed frequencies significantly different from what you’d expect from a random sample from the US population? We'll run a $\chi^2$ test on frequencies using $\alpha = 0.01$.

a) Make bar graph showing these frequencies:

![Bar Graph](image.png)

b) Calculate the expected frequencies for those who vote "Democrat", "Republican" and "I never (or can't) vote".

There are a total of $71 + 5 + 32 = 108$ students.
We'd expect $0.25 \times 108 = 27$ to vote Democrat
$0.25 \times 108 = 27$ to vote Republican
and $0.5 \times 108 = 54$ not to vote.
c) Calculate the $\chi^2$ statistic using:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

$$\chi^2 = \frac{(71-27)^2}{27} + \frac{(5-27)^2}{27} + \frac{(32-54)^2}{54} =$$

$$71.7037 + 17.9259 + 8.963 = 98.5926$$

d) Use table I to find the critical value of $\chi^2$

$$df = (3-1) = 2$$

$$\chi^2_{crit} = 9.21$$

e) Make your decision and state your conclusion using APA format

Our observed value of $\chi^2 = 98.59$ is greater than the critical value of 9.21. We reject $H_0$. The distribution of voting preference in our class is significantly different from the U.S. population, $\chi^2(2, N=108) = 98.59$, $p < 0.01$. 
Problem 2 Of the 75 students who vote, voter preference in our class breaks down into the following frequencies:

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democrat</td>
<td>57</td>
<td>13</td>
</tr>
<tr>
<td>Republican</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

We’ll use a $\chi^2$ test for independence with $\alpha = 0.05$ to determine if voting preference depends on gender for our class.

a) Make a bar graph of the observed frequencies
b) Calculate the rows and column sums and the expected frequencies for the null hypothesis

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Male</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democrat</td>
<td>(70)(61)/75  = 56.9333</td>
<td>(70)(14)/75 = 13.0667</td>
<td>70</td>
</tr>
<tr>
<td>Republican</td>
<td>(5)(61)/75 = 4.0667</td>
<td>(5)(14)/75 = 0.9333</td>
<td>5</td>
</tr>
<tr>
<td>sum</td>
<td>61</td>
<td>14</td>
<td>75</td>
</tr>
</tbody>
</table>

\[
\chi^2 = \frac{(57-56.9333)^2}{56.9333} + \frac{(4-4.0667)^2}{4.0667} + \frac{(13-13.0667)^2}{13.0667} + \frac{(1-0.9333)^2}{0.9333} = \\
0.0001 + 0.0011 + 0.0003 + 0.0048 = 0.0063
\]

c) Calculate the \( \chi^2 \) statistic.

d) Use table I to find the critical value of \( \chi^2 \)

\[ \text{df} = (2-1)(2-1) = 1 \]

for \( \alpha = 0.05 \), \( \chi^2_{\text{crit}} = 3.84 \)

e) Use the \( \chi^2 \) calculator in the excel spreadsheet to find the p-value for this test. Make your decision and state your conclusion using APA format.

Our observed value of \( \chi^2 \) is 0.0063 which is not greater than the critical value of 3.84. We fail to reject \( H_0 \). The distribution of voting preference in our class does not vary with gender, \( \chi^2(1, N=75) = 0.0063, p = 0.9367. \)
**Problem 3** Use R’s `chisq.test` function on the survey data to conduct the hypothesis test from problem 1. Voter preferences can be found in the field ‘vote’, and you’ll need to pull out just the three categories of ‘Democrat’ (#1), ‘Republican’ (#5) and ‘I never (or can’t) vote’ (#2).

Hint: for an example on running chi-squared tests for frequencies see [Chi2TestFrequencies.R](#).

```r
# HW9_problem3.R
#
# Load in the survey data
survey <- read.csv("http://www.courses.washington.edu/psy315/datasets/Psych315W20survey.csv")
# Make table of voter preferences
fo <- table(survey$vote)  # observed frequencies
# Restrict the table (fo) to Democrats, Republicans, and those who can’t vote
fo <- fo[c(1,5,2)]
fe = c(.25,.25,.5)  # expected frequencies
# run the chi-squared test:
out <- chisq.test(fo,p=fe)
# The chi-squared statistic is:
out$statistic
X-squared
98.59259

# The degrees of freedom is:
out$parameter
df
2

# And the p-value is:
out$p.value
[1] 3.898437e-22

# Writing in APA format can be done like this:
sprintf('Chi-Squared(%d,N=%d) = %5.2f, p = %0.20f',out$parameter,sum(fo),out$statistic,out$p.value)
[1] "Chi-Squared(2,N=108) = 98.59, p = 0.00000000000000000000"
```