CSC 315, Extra Credit (towards Exam II)

Perform a two-sample hypothesis test comparing two proportions, based on a Gallup Poll (http://news.gallup.com/home.aspx) or another poll used with my permission. In order to do this, you must determine the appropriate frequencies based on the information provided by Gallup (e.g., the sample size and relevant percentages). If comparing across different years, you may assume that the same number of individuals were surveyed each year. Your assignment must include the following information, submitted as a single Word document.

1) A word document containing the following:
    a. a link to the survey data
    b. an explanation of how you calculated the appropriate sample sizes (see below for an example).
    c. The null and alternative hypotheses, in the format given in the Formula Sheet, with the proportions appropriately defined

2) An R Notebook (created as a Word Document) that
    a. carries out the appropriate hypothesis test and outputs the results
    b. that extracts the chi-squared test statistic and calculates the corresponding Z test statistic
    c. that extracts the p-value
    d. that calculates the p-value from the Z test statistic in (b).

3) A word document that includes
    a. The conclusion of your hypothesis test in the context of this problem
    b. An explanation of what it would mean, in the context of your problem, if a Type I error occurred
    c. An explanation of what it would mean, in the context of your problem, if a Type II error occurred

When you are finished, combine each of the above into a single Word Document.

Example for 1 (a) and (b) (which is what I did for question 10 on Exam II)

The data for this question was based on the following poll:
http://news.gallup.com/poll/196394/americans-divided-priorities-criminal-justice-system.aspx

The survey methods (at the bottom of the link) state that "Results for this Gallup poll are based on telephone interviews conducted Oct. 5-9, 2016, on the Gallup U.S. Daily survey, with a random sample of 1,017 adults, aged 18 and older, living in all 50 U.S. states and the District of Columbia."

The survey results are reproduced in **Table 1**.

Table 1. Survey Results from October 2016 Gallup Poll

	Law and Order (%)	Reducing Bias against minorities (%)	No opinion (%)
National Adults	**49**	**43**	**8**
Republicans	77	17	6
Independents	32	60	8

From the survey results, and based on the fact that 1,017 individuals are surveyed, we can calculate the total number of individuals preferring Law and Order (n = 498) and Reducing Bias against minorities (437), as shown in **Table 2**.

Table 2. Number of individuals with each preference

	Law and Order (%)	Reducing Bias against minorities (%)
National Adults	**.49(1017) = 498**	**.43(1017 = 437**
Republicans	77	17
Democrats	32	60

If we let *R* = the total number of Republicans surveyed, and *D* = the total number of Democrats, then the following system of equations must hold:

$$\begin{cases} 0.77R + 0.32D = 498 \\ 0.17R + 0.60D = 437 \end{cases}$$

Solving the systems of equations we find *R* = 390 and *D* = 617. This gives us the following contingency table for Republicans and Democrats (**Table 3**).

Table 3. Contingency table for political affiliation and criminal justice preference

	Law and Order (%)	Reducing Bias against minorities (%)
National Adults	**.49(1017) = 498**	**.43(1017 = 437**
Republicans	.77(390) = 300	.17(390) = 66
Democrats	.32(617) = 197	.60(617) = 370

(The segment below would appear in your R Notebook for question (2)).

The *R* code provided on the exam generated this frequency table and sample sizes for Republicans and Democrats, and is reproduced below:

```
counts <- cbind(LawAndOrder = c(300, 197), ReducingBias = c(66, 370))
rownames(counts) <- c("Republicans", "Democrats")
counts

# total number of Republicans and Democrats
rowSums(counts)
```