STAT 301 HW14 due Wed Oct 28

You may work with one partner on this assignment, submitting one report with both names, provided that both students contribute substantially to the work. Word-processed reports are preferred to hand-written ones. Please copy/paste relevant computer output into your report as appropriate.

Blood Donations?

Blood donations from volunteers serve a crucial role in the nation's health care system. Have Americans become any more or less generous about donating blood in recent years? To investigate this question we can analyze data from the General Social Survey, which is a national survey conducted every two years on a random sample of adult Americans. Data from the 2002 and 2004 surveys are summarized in the following table:

	2002	2004
Donated blood in previous 12 months	210	230
Did not donate blood in previous 12 months	1152	1106
Total	1362	1336

a) For each year, calculate the proportion of adult Americans who donated blood in the previous 12 months. Use appropriate symbols to represent them. Also calculate the (absolute) difference between these proportions.

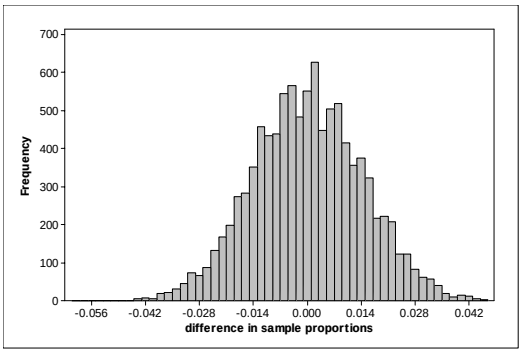
The sample proportion who donated blood in 2002 is $\hat{p}_{02} = \frac{210}{1362} \approx .154$. The sample proportion who donated blood in 2004 is $\hat{p}_{04} = \frac{230}{1336} \approx .172$. The (absolute) difference between them is .172 - .154 = .018.

b) State the appropriate null and alternative hypotheses, in words and in symbols, to address the research question in the first paragraph.

The null hypothesis is that the population proportions who gave blood are the same in 2002 and 2004 (H_0: $\pi_{02} = \pi_{04}$). The alternative hypothesis is that the population proportions who gave blood are different in 2002 and 2004 (H_a: $\pi_{02} \neq \pi_{04}$).

c) Conduct a Minitab simulation to approximate a p-value for this significance test. Be sure to report the appropriate parameter values (n and π) for the binomial distribution that you simulate from. Also submit a well-labeled histogram of your simulation results. Finally explain how you calculate the approximate p-value and report its value.

The simulation analysis uses sample sizes of $n_{02} = 1362$ and $n_{04} = 1336$. But since the simulation is conducted assuming the null model to be true, we use the same success probability π for both years. We approximate π with the combined sample proportion of blood donators: (210+230)/(1362+1336) = 440/2698 ≈ .163. The following histogram displays the difference in sample proportions ($\hat{p}_{02} - \hat{p}_{04}$)for 10,000 repetitions of the simulation:



To determine the approximate p-value, we count how many of these repetitions produced a difference of .018 or higher (in absolute value) and then divide by 10,000. As shown in the following Minitab output, this gives an approximate p-value of .2049.

```
MTB > let c6=(abs(c5)>=.018)
MTB > tally c6
```

Tally for Discrete Variables: C6

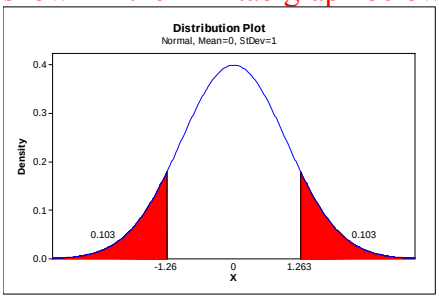
```
C6  Count
 0   7951
 1   2049
N=  10000
```

d) Check the conditions for whether the normal approximation is appropriate for this significance test.

Both years involved a random sample of adult Americans, and both samples have far more than 5 people who donated blood and far more than 5 people who did not give blood, so the technical conditions are easily met.

e) Calculate the *z*-test statistic and p-value based on the normal distribution.

The *z*-statistic is $z = \dfrac{.172 - .154}{\sqrt{.163(.837)\left(\dfrac{1}{1336} + \dfrac{1}{1362}\right)}} \approx 1.263$. The (two-sided) p-value is .206, as shown in the Minitab graph below:



f) Summarize your conclusion from these analyses, with regard to the research question in the first paragraph. Also explain the reasoning process that leads to your conclusion.

The sample data provide little or no evidence of a difference in the population proportions who donated blood between 2002 and 2004. Our simulation and normal-based analyses both reveal that the observed difference in sample proportions who gave blood (.018) would not be surprising if the null model (that the population proportions who gave blood were the same in both years) were true.