1. (5 points) The following question relies on this file: minarets.Rdata (see blackboard or "`http://www.unige.ch/ses/sococ/cl/edat/minarets.Rdata`"). The data examine whether people voted to prohibit minaret (a tower, typically part of a mosque) construction in their neighborhood. The variable *vote1* is a cleaned up version of the data, where a Yes vote signifies support for prohibition. Using *gender* as an explanatory variable:
   - Present a cross-tabulation of gender and vote choice.
   - Use a $\chi^2$ test to determine whether these variables are independent.
   - Report your results. Include discussion on how to interpret the p-value for your test.

2. (2 points) Why do we usually rely on the t-distribution rather than a normal distribution for hypothesis testing?

3. (4 points) Discuss and give (political science) examples at least four different types of t-tests. Your discussion should include when they are appropriate to use. For the political science examples, attempt to identify political relevant variables that could be used in these different tests. Use different variable in each example.

4. (a) (2 points) What is statistical power? Be precise. Why is this important?

   (b) (2 points) Suppose you are planning a survey and want to eventually conduct an one sample, two-tailed t-test. You expect a difference of $\bar{X} - \mu$ of 5, and a standard deviation of 10. What number of survey respondents do you need to recruit if you want 95% power.

5. The following questions rely on this file: cereal.csv.
   (a) (2 points) Explore the distributions of sugar in cereals labelled "Children" and cereals labelled "Adult". Do they follow a normal distribution? Use a qq-plot to support your argument.

   (b) (2 points) I'm interested in the sugar differences in cereals labelled "Children" and cereals labelled "Adult". What is the difference in means of these two types of cereals' sugar level?

   (c) (2 points) Given that I'm interested in the difference of sugar levels of the two types of cereals, what would we call the dependent variable and what would we call the independent variable in this analysis?

   (d) (2 points) Now I'm interested in a statistical difference. Assuming the cereals have unequal variance, report an appropriate test statistic? Is there a statistical difference in the means of the two types of cereals?

   (e) (2 points) Does the exploration of the data in the first part of the question impact the validity of this test at all? In other words, is a t-test the appropriate hypothesis test given the distribution of the variables?

   (f) (2 points) For the moment, assume variances of the two types of cereals are actually equal. Report an appropriate test statistic? Is there a statistical difference in the means of the two types of cereals.

(g) (2 points) What the difference in our results based on our assumptions of the variance.

(h) (3 points) Is the assumption of equal variance defensible?

6. Earlier in the semester I asked you to identify a dataset that you want work with this semester. This dataset might be part of an existing published research or be a dataset that you downloaded from an organization (i.e. the World Bank or Pew).

   (a) (2 points) Describe the dataset and where you found it. Your description must identify the unit of observation, the temporal scope of the data, cases that the dataset covers. Assume that the reader knows nothing about the data.

   (b) (2 points) Identify a continuous and binary variable in your data. (You may need to transform one of the variables in your dataset to create a binary variable). Describe these variables in words, descriptive statistics, and graphs.

   (c) (2 points) Let's prepare to run a difference of means t-test using the aforementioned variables. First, what is the dependent variable and what is the independent variable?

   (d) (2 points) Will you assume equal variance? Why or why not?

   (e) (2 points) What is the research/alternative hypothesis? What is the null? What level of significance will your test your null hypothesis at?

   (f) (2 points) Derive a test statistic using your data.

   (g) (2 points) Interpret the results as if you were reporting your results in an academic article.