1. An unnamed political science professor recently served on a federal grand jury. He was told that he was selected at random from adults in the Boston metropolitan area. The grand jury had 23 members on it, all of them White. What is the chance this could have happened randomly?

2. True or false: The average of 10 random draws from a distribution is 10 times more variable than the average of 1,000 random draws from the same distribution.

3. If we gave everyone in America an IQ test, we would get an average of 100 with a standard deviation of 16. It would be normally distributed. If we draw a sample of 100 people,
   2.a what is the probability that the average of the sample will be more than 105?
   2.b what fraction of the sample should we expect to be between 90 and 110?
   2.c what fraction of the sample should we expect to be above 140?

4. You are conducting research into how much money MIT undergraduates spent on their spring break trip. You send out a survey that asks people to tell you, in ranges, how much they spent. Here are the results:

<table>
<thead>
<tr>
<th>Range</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0</td>
<td>150</td>
</tr>
<tr>
<td>$0-$100</td>
<td>50</td>
</tr>
<tr>
<td>$100-$200</td>
<td>40</td>
</tr>
<tr>
<td>$200-$500</td>
<td>30</td>
</tr>
<tr>
<td>$500-$1000</td>
<td>20</td>
</tr>
<tr>
<td>More than $1000</td>
<td>5</td>
</tr>
</tbody>
</table>

Provide your estimate of the average cost of a spring break trip and the standard error associated with it.
5. In May 1975 *Time Magazine* reported a survey conducted by Yankelovich, Skelly and White concerning views that a random sample of adults had about antiwar protesters. The sample size was 1,014. Here is the question and the results:

Question: Looking back, do you give a lot of credit to the college youth in the 1960's who demonstrated against the (Vietnam) war, or do you feel that they went about things the wrong way?

- 32% Give credit
- 61% Went about things wrong way
- 7% Not sure

What is the 95% confidence interval around the “give credit” answer? Write a sentence that explains what this means.

6. The following two histograms show the distribution of grade point averages of two universities. Both have a mean GPA of 3.0 and standard deviation of 0.7. If you sampled 100 students from University A and 100 students from University B, how would the averages and standard errors of the two samples differ?

7. The 2000 Current Population Survey Special Report on Voting and Registration interviewed 74,174 adults of voting age in November 2000, after the general election. 49,269 of them reported they had voted. Is this number “too high” or “too low,” given the number of eligible voters we know participated in the 2000 presidential election (roughly 52% turnout)?

8. In the survey mentioned above, 22,898 of the 34,580 men and 27,112 of the 39,594 women said they voted. Run the t-test to see if the fraction of men voting equals the fraction of women voting.
9. Find the probability that a normally distributed random variable with mean of 0 and standard deviation of 1 takes on the following values:
   a. greater than 0
   b. greater than 1.96
   c. greater than 2.54
   d. greater than 1.96 or below -1.96
   e. between -1.96 and 1.96

10. Imagine a variable with mean of 0 and standard deviation of 1. What is the value of \( z \) such that the variable takes on ....
    a. a value less than \( z \) 99% of the time.
    b. a value less than \( z \) 64% of the time.
    c. a value greater than \( z \) 35% of the time.

11. (The following is based on some real research I was asked to review for a journal this week.) Political scientists are interested in knowing whether candidates for office are responsive to the voters in their districts. Specifically, we want to know whether liberal candidate run in liberal districts and conservatives run in conservative districts.

    In the research here, there’s even greater specificity. In the U.S., candidate for office must first be nominated through primaries. Primary voters consist of only party registrants from the candidate’s party. Therefore, the specific question is this: are congressional candidates more conservative whenever the primary electorate within their own party is more conservative, and vice versa?

    To study this question, a political scientist uses the 1988–1992 Senate Election Study to study the relationship between the ideological disposition of a state’s primary voters and the ideological position that the Republican and Democratic candidates for Senate took in the election. He reports that the correlation between a candidate’s ideological position and the average ideological position of a voter in his/her party within the state is .77. Therefore, he concludes that Senate candidates (at least in these elections) were “responsive.”

    The data to replicate this analysis are located in /mit/17.871/Examples/senate.dta. The variables are as follows:

    state: postal abbreviation for the state
    year: year of the Senate election
    party: party of the Senate candidate
**cand_ideo**: candidate ideological position on a 0-6 scale (0 = most liberal; 6 = least liberal). The scale is constructed by taking the average answer that survey respondents gave when they were asked to place the candidate on this scale.

**voter_ideo**: the average ideological placement of the respondents in the state who were from the candidate’s party on the same 0–6 scale.

Note that the data set is constructed so that there are two observations for each state; an observation is a state-candidate-year.

12. You are interested in reporting to Dean Robert Redwine about what makes MIT students like the classes they take. You use the data from the Spring 1995 Course Evaluation Guide to do your research. (Don’t worry, the basic story hasn’t changed.) You will find a codebook for this data located on the 17.871 web site. The survey instrument has also basically stayed the same, so if you want to refresh yourself, you can go to the current Course Evaluation Guide web site to see what the questions look like.

Construct a regression equation that explains student satisfaction with their classes, based on at least four independent variables.

Hand in a “do file” that will reproduce the regression that you report.

Type up the table. Write a page that describes the variables you chose, why you chose them, and what you concluded from this exercise.