From http://www.ma.utexas.edu/users/parker/384/

Prerequisite Activities

Discussion of all the prerequisites required for SSC 384 Topic 2, Mathematical Statistics I is on the website. These are activities – something for you to do – to enhance your reading of the discussion of prerequisites. We will **briefly review some** of these in class. If, now, you are not able to do these problems and other standard types of problems over these topics, it is crucial that you take the prerequisite courses or, if you have had those, then review on your own outside of class. Ask for guidance if needed.

1. (Probability) Let X have the pdf (probability density function) $f(x) = 6x(1-x)$ for $x \in [0,1]$.
   a. Find the mean of the random variable.
   b. Find $P(X < 0.75)$
   c. Find the cdf (cumulative density function) of X and graph the cdf on [-1,3].

2. (Sampling distribution) What is the definition of the sampling distribution of a statistic?

3. (Sampling) The usual methods of inference in an introductory statistics course are valid for data from simple random samples.
   a. Briefly, what should be done differently if the data is from a stratified random sample?
   b. What is different about the analysis if the data is from a judgment sample?

4. (Theory of Estimation) Consider a random variable with uniform distribution on $[0,\theta]$ where $\theta > 0$. You will take a SRS (simple random sample) of size 3. Consider $X_{\max}$ as an estimator of θ. Is it an unbiased estimator?

5. (Confidence interval) Here are the IQ test scores of 21 ninth-grade students in a California school district.
   a. Find a 99% confidence interval for the population mean.
   b. For what population is this a reasonable inference?
   c. What assumptions are necessary for this inference to be valid?

83	113	98	104	72	102	91
74	114	103	105	108	130	120
103	132	111	128	118	119	86

6. (Hypothesis test) Results of a public-opinion poll reported in Time (April 5, 1993) indicated that 54% of the respondents between the ages of 18 and 26 feel that religion is a very important part of their lives. The article also says that 1013 people were interviewed. Is there good evidence that more than half of the people in this age group feel that religion is a very important part of their lives? (Be sure to state hypotheses, give a p-value, interpret the p-value, and state your conclusion.)

7. (Power of a test) The Rockwell hardness index for steel is determined by pressing a diamond point into the steel and measuring the depth of penetration. For a certain type of steel, the manufacturer claims that the mean hardness index is 68 (or more) and the standard deviation of the hardness index is 5. Your client uses this steel in manufacturing machinery and wants to determine whether the product fails to meet this specification by doing the usual hypothesis test on the mean at the 5% significance level. He will assume the manufacturer's claimed standard deviation is correct.
   a. If he uses a sample size of 80, what is the power of this test against the value of 67.6 in the alternative hypothesis?
   b. Do you think this is satisfactory? What are some ways your client could modify this to have a more powerful test? Which of these ways is likely to be most useful to your client?

From http://www.ma.utexas.edu/users/parker/384/

8. (Maximum Likelihood) Consider the random variable with pdf $f(x) = \alpha x^{\alpha-1}$ where $0 < x < 1$ and $\alpha > 0$. Find the maximum likelihood estimator of α.

9. (Regression. Only occasionally referenced in this course.) The following data are mean heights of a group of children in Kalama, an Egyptian village that is the site of a study of nutrition in developing countries. The data were obtained by measuring the heights of all 161 children in the village each month over several years. The variables are age in months and the mean height in centimeters for children at this age. (Moore and McCabe, *Introduction to the Practice of Statistics*. See http://lib.stat.cmu.edu/DASL/ and, under data subjects, choose Health. This dataset is "Age and Height".)
   a. Find the regression equation to predict height from age for this data.
   b. What is the correlation coefficient between these two variables?
   c. If we considered data on individual children, that is, the age and height of each child, what would this scatterplot look like? (Assuming no data were missing, how many points would you have? What would the general shape be? How variable are the data around the line?) How would the correlation coefficient of this data on individual children compare to the correlation coefficient of the given data on groups of children?

age	18	19	20	21	22	23	24	25	26	27	28	29
ht	76.1	77	78.1	78.2	78.8	79.7	79.9	81.1	81.2	81.8	82.8	83.5

10. (Calculation) From any applied statistics text you have, choose an example and carry out the analysis using your statistical software. You might choose a test for the equality of two means or a chi-squared test of independence in a two-way table. We will do very few calculations like these in this class. But, if you need to do some on homework, I assume you could learn fairly quickly, on your own, how to do any of these: tests or confidence intervals for means or proportions, chi-squared tests, regression, ANOVA.

11. (Software skills) Consider an Exp(5) distribution.
    a. (Graph) Using mathematical software or a spreadsheet, graph the density function. Determine an appropriate domain for graphing. Evaluate the function at intervals of 0.01 over that domain. Graph it.
    b. (Solve) Using software or symbolic calculations, Find the x-value(s), correct to two decimal places, at which that density function is equal to 0.15.
    c. (Simulation) Using statistical software, generate 1000 random samples from this distribution and find the mean and variance of that set of data.
    d. Discuss whether the graph and mean and variance in part c. are consistent with what you would expect from your solution to part a.

12. (Sampling distribution of the sample mean) Use statistical software to approximate the sampling distribution of the sample mean for a sample of size 15 from a population with a chi-squared distribution with 6 degrees of freedom.
    a. Find 1000 points from that sampling distribution.
    b. Find the mean and standard deviation and make a histogram of those 1000 values to approximate the full sampling distribution.
    c. Are these results what you would expect? (To know what to expect, use the tables in the back of the book to find the mean and variance of the population distribution, and use the Central Limit Theorem. Would you expect n to be large enough for the CLT to give normality of this sampling distribution?)