#1.
The following table lists length of stay in hospital (days) for a sample of 25 patients.

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>10</td>
<td>6</td>
<td>11</td>
<td>5</td>
<td>14</td>
<td>30</td>
<td>11</td>
<td>17</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>8</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
<td>11</td>
<td>9</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

By any means you like (e.g. – StatKey, R or Stata, construct a frequency/relative frequency table for these data using 5-day class intervals. Include columns for the frequency counts, relative frequencies, and cumulative frequencies.

#2.
The following table lists fasting cholesterol levels (mg/dl) for two groups of men.

<table>
<thead>
<tr>
<th>Group 1:</th>
<th>Group 2:</th>
</tr>
</thead>
<tbody>
<tr>
<td>233 291 312 250 246</td>
<td>344 185 263 246 224</td>
</tr>
<tr>
<td>254 276 234 181 248</td>
<td>212 188 250 148 169</td>
</tr>
<tr>
<td></td>
<td>226 175 242 252 153</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>183 137 202 194 213</td>
</tr>
</tbody>
</table>

By any means you like (e.g. Statkey, R or Stata, construct the following graphical comparisons of the two groups:

2a. Side-by-side box plot

2b. Side-by-side histograms with same definitions (starting value, ending value, tick marks, etc) of the horizontal and vertical axes.

In 1-2 sentences, compare the two distributions. What conclusions do you draw?
#3.

Consider the following setting. Seventy-nine firefighters were exposed to burning polyvinyl chloride (PVC) in a warehouse fire in Plainfield, New Jersey on March 20, 1985. A study was conducted in an attempt to determine whether or not there were short- and long-term respiratory effects of the PVC. At the long term follow-up visit at 22 months after the exposure, 64 firefighters who had been exposed during the fire and 22 firefighters who where not exposed reported on the presence of various respiratory conditions. Eleven of the PVC exposed firefighters had moderate to severe shortness of breath compared to only 1 of the non-exposed firefighters.

Calculate the probability of finding 11 or more of the 64 exposed firefighters reporting moderate to severe shortness of breath if the rate of moderate to severe shortness of breath is 1 case per 22 persons. **Show your work.**

#4.

The Air Force uses ACES-II ejection seats that are designed for men who weigh between 140 lb and 211 lb. Suppose it is known that women’s weights are distributed Normal with mean 143 lb and standard deviation 29 lb.

4a. What proportion of women have weights that are outside the ACES-II ejection seat acceptable range?

4b. In a sample of 1000 women, how many are expected to have weights below the 140 lb threshold?

#5.

Consider the setting of a single sample of n=16 data values that are a random sample from a normal distribution. Suppose it is of interest to perform a type I error $\alpha = 0.01$ statistical hypothesis test of $H_0: \mu \geq 100$ versus $H_A: \mu < 100$, $\alpha = 0.01$. Suppose further that $\sigma$ is unknown.

5a. State the appropriate test statistic

5b. Determine the critical region for values of the sample mean $\bar{X}$. 
#6.

An investigator is interested in the mean cholesterol level $\mu$ of patients with myocardial infarction. S/he drew a simple random sample of $n=50$ patients and from these data constructed a 95% confidence interval for the mean $\mu$. In these calculations, it was assumed that the data are a simple random sample from a normal distribution with known variance. The resulting width of the confidence interval was 10 mg/dl.

How large a sample size would have been required if the investigator wished to obtain a confidence interval width equal to 5 mg/dl?

#7.

In (a) – (e) below, you may assume that the data are a simple random sample (or samples) from a normal distribution (or distributions). Each setting is a different setting of confidence interval estimation. In each, state the values of the confidence coefficients (recall – these will be the values of specific percentiles from the appropriate probability distribution).

7a. For a single sample size of $n=15$ and the estimation of the population mean $\mu$ when the variance is unknown using a 90% confidence interval, what are the values of the confidence coefficients?

7b. For a single sample size $n=35$ and the estimation of a variance parameter $\sigma^2$ using a 95% confidence interval, what are the values of the confidence coefficients?

7c. For a single sample size of $n=25$ and the estimation of the population mean $\mu$ when the variance is known using an 80% confidence interval, what are the values of the confidence coefficients?

7d. For the setting of two independent samples, one with sample size $n_1 = 13$ and the other with sample size $n_2 = 22$, it is of interest to construct a 90% confidence interval estimate of the ratio of the two population variances, $[\sigma_1^2/\sigma_2^2]$. What are the values of the confidence coefficients?
#8.

A study was investigated of length of hospital stay associated with seat belt use among children hospitalized following motor vehicle crashes. The following are the observed sample mean and sample standard deviations for two groups of children: 290 children who were not wearing a seat belt at the time of the accident plus 123 children who were wearing a seat belt at the time of the accident.

<table>
<thead>
<tr>
<th>Group</th>
<th>Sample size, $n$</th>
<th>Sample mean $\bar{X}$</th>
<th>Sample standard deviation $S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seat belt = no</td>
<td>$n_{NO} = 290$</td>
<td>$\bar{X}_{NO} = 1.39$ days</td>
<td>$S_{NO} = 3.06$ days</td>
</tr>
<tr>
<td>Seat belt = yes</td>
<td>$n_{YES} = 123$</td>
<td>$\bar{X}_{YES} = 0.83$ days</td>
<td>$S_{YES} = 2.77$ days</td>
</tr>
</tbody>
</table>

You may assume normality. You may also assume that the unknown variances are equal. Construct a 95% confidence interval estimate of the difference between the two population means. In developing your answer, you may assume that the population variances are unknown but EQUAL.

8a. What is the value of the point estimate?

8b. What is the value of the estimated standard error of the point estimate?

8c. What is the value of the confidence coefficient?

8d. What are values of the lower and upper limits of the confidence interval?

8e. Write a clear interpretation of the confidence interval.
#9.
A test consists of multiple choice questions, each having four possible answers, one of which is correct. What is the probability of getting exactly four correct answers when six guesses are made?

#10.
After being rejected for employment, woman “A” learns that company “X” has hired only 2 women among the last 20 new employees. She also learns that the pool of applicants is very large, with an approximately equal number of qualified men and women. Help her address the charge of gender discrimination by finding the probability of getting 2 or fewer women when 20 people hired under the assumption that there is no discrimination based on gender. Does the resulting probability really support such a charge?

#11.
Suppose the length of newborn infants is distributed normal with mean 52.5 cm and standard deviation 4.5 cm. What is the probability that the mean of a sample of size 15 is greater than 56 cm?

#12.
Suppose that 25 year old males have a remaining life expectancy of an additional 55 years with a standard deviation of 6 years. Suppose further that this distribution of additional years life is normal. What proportion of 25 year-old males will live past 65 years of age?