Calculating Means, Variances and Standard Deviations:

1. Absenteeism in our plant varies considerably from day to day. Over the last ten days the number of employees absent has been

   10, 15, 5, 6, 7, 22, 3, 14, 8, 21

   a. Calculate the mean, variance and standard deviation of absences. The plant has 80 employees. Repeat these calculations for the percent of employees absent.

   b. Calculate the median and mode for this data set for the values. How do the median, mode and mean compare?

2. The midterm test for a statistics course has a time limit of 1 hour. However, like most statistics exams, this one was quite easy. To assess how easy, the professor recorded the amount of time taken by a sample of nine students to hand in their test papers. The times, rounded to the nearest minute, are

   33   29   45   60   42   19   52   38   36

   a. Compute the mean, median and mode. What can you learn about the exam from these three statistics?

3. A plant which is under our supervision is in the midst of an organizing drive by Local 9 of the United Clerical Employees. There are ten employees in the clerical unit at this plant. Prior subtle and illegal research suggests that 45% of our clerical employees at all of our locations, sympathetic to union representation. Using the binomial distribution, what is the likelihood that six or more of the employees in this ten employee unit will vote for a union? To calculate this outcome, you will need to use the binomial formula and calculate the likelihood of six, seven, eight, nine or ten employees will vote for a union (and then add these probabilities). Before trying to use the formula you will need to establish:

   a. What is the probability of a success (a vote for the union)?
   b. What is the probability of a failure (a vote against the union)?
   c. What is the total number of experiments?
   d. What is the number of successes (this will vary with the outcome you are calculating).

Remember, n! is n*(n-1)*(n-2)*(n-3)*.....*5*4*3*2*1 and 0! = 1

so, for example, 6! = 6*5*4*3*2*1.
4. We are developing a business plan for a psychic hot line. Based on our work with other similar services and the type of advertising we intend to do, we expect that the line will average 14 calls per hour. We also know that people who use these hot lines hate to wait to get answers to life’s most troubling questions. Calculate that probability that we will get at least 14 15, ..... 24 calls per hour. If a typical psychic can handle four calls per hour, how many psychics do we need to assure that 75% of calls are covered in any one answer. That virtually all (99%) of calls are taken (at $4.00 per minute, you want to get all the calls). Can we be absolutely certain that with the manning you suggest for total coverage we will never miss a call? Explain your answer.

5. Six employees have taken an examination to assess their skills. Their scores on the exam were 500, 520, 540, 560, 580 and 600.

   A. What was the mean test score, variance and standard deviation for this population?

   B. How many unique samples of four can be drawn from this population?

   C. List each of these samples and calculate their mean

   D. Create a probability distribution for the means of these samples.

   E. Demonstrate that the mean of the sample means is the population mean.

   F. What is the standard deviation of the sample means? How does this compare to the standard deviation of the population?

6. If random variable z is distributed N(0,1), calculate the likelihood (probability) that a randomly chosen z will:

   a. $Z \geq 1.96$

   b. $Z \leq -1.96$

   c. $0 \leq Z \leq 1.96$

   d. $Z \leq -1.96$ or $Z \geq 1.96$

   e. $1.28 \leq z \leq 1.96$

   f. $-1.96 \leq Z \leq 1.96$

   g. $Z = 1.96$
7. Returning to the employment exam, data on the exam indicates that the mean score for the examination is 520 with a standard deviation of 20. Exam scores are normally distributed. What is the likelihood (probability) that an individual will score over 550 on the exam? What is the likelihood that the mean score for a group of four employees will be over 550? Explain why the two likelihoods differ.

8. We have 25 employees in an establishment owned by our firm. As part of our annual meet the President day, we choose five employees to have lunch with the President of the firm. How many unique samples of five employees (without regard to the order in which employees are chosen) are there?

9. We are interested in the effect of extended overtime and travel on the blood pressure of managerial employees in our firm. As we have a mandatory annual health checkup for all managerial employees, we know that our average manager has a diastolic blood pressure of 88 with standard deviation of 8 psi.

We test 16 employees on a Friday after a week of overtime which involved travel. We find that the mean blood pressure for this group of employees is 92 psi. Answer the following questions.

a. What is the likelihood that, given the mean and standard deviation for blood pressure for our managerial employees, a single employee would be found to have blood pressure of 92?

b. What is the likelihood, given the mean and standard deviation in blood pressure for managerial employees, that a group of four employees would have a mean blood pressure of 92?

c. What is the likelihood, given the mean and standard deviation in blood pressure for managerial employees, that the full group of sixteen employees would have a mean blood pressure of 92?

d. Given the result to part c, is it reasonable to conclude that overtime and travel contribute to higher blood pressure?
It can be wise to prepare for a worst case, but it is difficult to know what a reasonable worst case is. We can use statistical information to help decide what type of conditions are sufficiently likely to occur that it is worthwhile preparing for those conditions. Here is an extreme case.

In 1910 the last bit “undiscovered” place on earth was the South Pole. Two explorers, Robert Falcon Scott (a British navel captain) and R. Amundson (of Norway), put together expeditions to try to reach the pole. Amundson was an experienced Arctic explorer who drew on his experience to Inuit to equipt and train his expedition. Scott was much less knowledgeable about the type of conditions his team would face. Their eventual fate was due, in part, to bad decisions by Scott (which have been due, in part, to British explorers reluctance to learn from indigenous peoples. When you have developed a doctrine of White and British superiority as a part of running a colonial system, its difficult to accept that one’s subjects - - or any other lower race - - are more knowledgeable than you). But that is beyond our issue for this problem.

As with many of his day, Scott believed in scientific knowledge and was particularly keen on using the developing science of weather forecasting to understand how his team of five would need to be equipt. For example, the type and amount of clothes, of food and of fuel needed to take five men on foot to the South Pole and back would be vitally determined by the weather conditions. The more extreme the temperature the fewer days on which the team could travel (lengthening the trip). Lower temperatures also require more clothes to remain warm and more fuel to melt water, cook food and heat the tents.

Fortunately, a British meteorologist had been observing temperatures in Antarctic for fifteen years and had calculated that the average temperature during the Antarctic summer was -10°F. Trying to build in a margin of error, Scott planned for an average temperature of -15°F. The actual average temperature in 1910 was -25°F. For this and other reasons, all members of the expedition perished on the return trip.

Although we still tend to focus on the mean value, we have a somewhat better understanding of dispersion (variance) than did scientists did in Scott’s day. Reworking the meteorological data Scott used, we find that (1) the mean temperature over the prior 15 years was -10°F, the standard deviation of temperatures was 10°F and temperatures are normally distributed around the mean (this result is actually due to observations beyond those available in Scott’s day). Given this information, answer the following questions:

a. What was the likelihood that the average temperature would be -10°F or colder?

b. What was the likelihood that the average temperature would be -15°F or colder?

c. What was the likelihood that the average temperature would be -20°F or colder?

d. What was the likelihood that the average temperature would be -25°F or colder?

e. This one is a little harder. Sometimes it is a good idea to prepare for worst cases,
say conditions which will occur 5% of the time or 1% of the time. If Scott had wanted to prepare for a low temperature which would occur only 5% of the time (5 times in 100 years), what temperature should he have prepared for. How about a low temperature which would have occurred only once in 100 years?

11. A variant on the last problem. We are concerned that some of our 10,000 employees are taking excessive days away from work without a legitimate cause. We wish to review extreme cases, to be able to distinguish between legitimate absences (say for cancer treatment or other long term illnesses and non-legitimate absences. The latter may be supported by doctor’s reviews, but we are suspicious of some area medical personnel). Reviews are expensive and we currently lack criteria for excessive time away from work. The latter may be a problem as any disciplinary action we take is subject to review by a third party and we need a neutral criteria in deciding which cases are subject to review.

We look at our data and find that the average employee takes six days of sick leave annually with a standard deviation of 3 days. Calculate the following:

A. How many days of sick leave are taken by the 5% employees with the greatest use of sick leave? How many employees are in this group?

B. How many days of sick leave are taken by the 1% employees with the greatest use of sick leave? How many employees are in this group?

C. How many days of sick leave are taken by the 0.5% employees with the greatest use of sick leave? How many employees are in this group.

D. How could we use this information to develop a criteria for deciding which employees illnesses will be reviewed?