Announcements
Lab for this week will be on Web page
Lab materials due on Thursday
Exploratory data analysis covered in lab

Last Lecture
Requirements for inferential statistics
Probabilities and distributions
Normal distributions
Estimation of center of a continuous variable

Estimation of the center of a distribution for a continuous variable (Table 2.1)
- Median-
  - middle measurement of a set of data, which estimates the center of a distribution
  - Is equal to the mean if data are normally distributed
  - Is less sensitive than the mean to outliers if data are skewed
- Mean
  - Closest constant variable to a random variable
  - If equal weighting used, calculated by multiplying the sum by 1/N, the sample size

Estimating the spread of a distribution (Table 2.1)
- Variance is the average of the sum of squared deviations
  - If we are looking at population variance, SS/n
  - Usually we look at sample variance, SS/(n-1) to remove bias
- With a greater dispersion of observations from the mean, the variance increases
- Standard deviation is the square root of the variance and occurs in the same units as the original variable (e.g. mm instead of mm²)
- C.V. is the standard deviation divided by the mean, and allows comparison of variables with different means.

Standard error (fig 2.2)
- For sample, we also want to know how much confidence we can place onto the sample mean
- If we collect many samples, the distribution of sample means should conform to a normal distribution (Central Limit Theorem)
- The mean of the distribution of sample means = population mean
- The standard deviation of the distribution of sample means = standard error
- We often calculate standard error by dividing the standard deviation by the sample size
- Standard errors, but not standard deviations, depend on the sample size

Confidence intervals(fig 2.1)
- Determined by estimating the area under the probability density function and backtracking to the distance from the mean for that given area
- Mean and standard deviation determine the shape of the curve, so the confidence intervals are calculated based on the distance from the mean multiplied by the standard deviation
- With higher confidence intervals, we move further from the population mean

Basing CI on sample statistics
- In practice, population mean and standard deviation are unknown
- We use the t distribution to calculate confidence intervals (t is based on difference between sample mean and population mean, divided by SE)
- t distribution is similar to normal, but changes with degrees of freedom (n – 1)
- CI means that if we sampled repeatedly, that proportion of the samples would contain the population mean (page 20 of text)