

More on Sampling

Standard errors, Central Limit
Theorem, Bootstrapping

Standard Error Revisited

- Property of the sampling distribution
 - Distribution of means (derived from estimates)
 - Not distribution of data
- Can be estimated from data
 - Variance
 - Sample size

Central Limit Theorem

- As sample size increases, the sampling distribution (distribution of means) approaches a normal distribution, no matter what the underlying distribution of data is
- http://onlinestatbook.com/stat_sim/sampling_dist/index.html
- Why is this important
- Many of the statistics we estimate assume a normal distribution

For instance

- Estimating required sample size
- How many samples must I measure to have a xx% chance of coming within E units of the true mean?
- Clutch size in grouse nests
 - Want to be within ± 2 eggs of true mean 95% of the time
- $n = (t^2 s^2) / E^2$
 - E is specified by researcher
 - s^2 can be obtained by a preliminary sample
 - $t?$
 - # of standard errors on either side of mean enclosed with confidence
 - a function of a normal distribution

Bootstrapping

- Simulation of many sampling efforts from the data of one sampling effort
- Assumes only that the data are representative of the population
- Allows examination of the properties of the samples
 - bias
- Allows estimation of variance when it can't be reliably estimated with one sample

How to bootstrap

- Take sample of size n
- Select a random sample (w/ replacement) of those data
- Repeat that resampling many times to generate sampling distribution
- Jackknife is a similar procedure but samples without replacement