

# More on Sampling

Standard errors, Central Limit  
Theorem, Bootstrapping

# Standard Error

- 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://www.ruf.rice.edu/%7Elane/stat\\_sim/sampling\\_dist/index.html](http://www.ruf.rice.edu/%7Elane/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  $\pm 2$  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

Table 1. Body length (cm) of long-toed salamanders (*Ambystoma macrodactylum*) sampled at low, medium, and high elevation sites in Montana.

	Low	Medium	High
	8.3	8.9	8.8
	10.2	8.9	10.1
	6.5	5.9	10.9
	7.6	8.4	9
	12.5	10.5	8.1
	11.3	8.4	9.9
	10.9	11.9	10.2
	11.6	10.5	11.0
	8	10.8	9
	8.4	5.8	11.2
	10.6	14	8.3
	8.2	10.1	6.4
	13.6	11.9	8
	10	4.8	17
	9.5	9.6	7.4
	4.9	11.8	9.2
	9.9	7.8	4.3
	9.8	9.5	11.1
	11	15.3	6.4
	8.4	5.4	12
			13.3
			5.7
			12.1
			10.1
			12.7
			6.8
			7.3
			10.2
			5.8
			11.6
$\bar{X}$	9.56	9.51	9.46
s <sup>2</sup>	4.07	9.19	8.44
s	2.02	3.03	2.90

We can also depict these measurements graphically in a histogram.

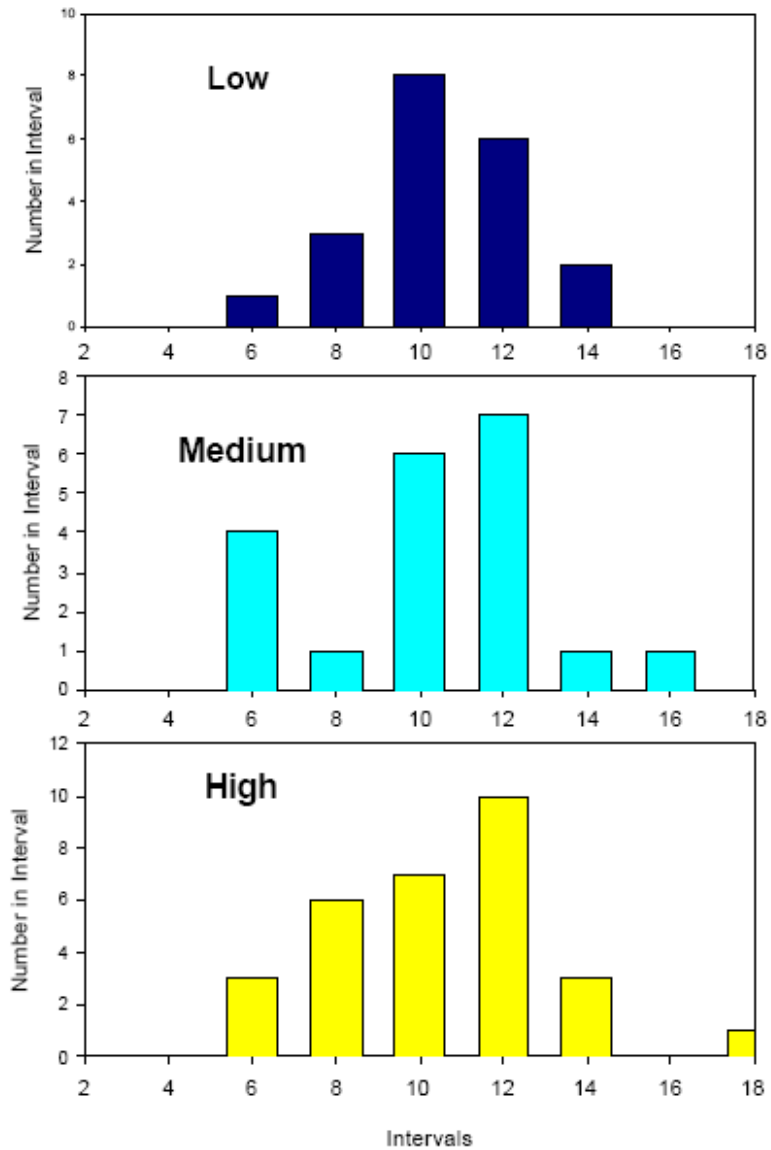


Table 1. Body length (cm) of long-toed salamanders (*Ambystoma macrodactylum*) sampled at low, medium, and high elevation sites in Montana.

	Low	Medium	High
	8.3	8.9	8.8
	10.2	8.9	10.1
	6.5	5.9	10.9
	7.6	8.4	9
	12.5	10.5	8.1
	11.3	8.4	9.9
	10.9	11.9	10.2
	11.6	10.5	11.0
	8	10.8	9
	8.4	5.8	11.2
	10.6	14	8.3
	8.2	10.1	6.4
	13.6	11.9	8
	10	4.8	17
	9.5	9.6	7.4
	4.9	11.8	9.2
	9.9	7.8	4.3
	9.8	9.5	11.1
	11	15.3	6.4
	8.4	5.4	12
			13.3
			5.7
			12.1
			10.1
			12.7
			6.8
			7.3
			10.2
			5.8
			11.6
$\bar{X}$	9.56	9.51	9.46
s <sup>2</sup>	4.07	9.19	8.44
s	2.02	3.03	2.90
Standard error:	0.46	0.62	0.49

# 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