Science Methods & Practice BES 301

February 3 and 8, 2011

Describing & Examining Scientific Data

Describing Scientific Data

32.6 cm	23.2
23.2	31.6
14.1	35.6
35.2	26.2
36.8	36.7
45.1	32.4
33.5	42.6
33.9	27.8
16.6	42.8
38.2	47.6

Length of coho salmon returning to North Creek October 8, 2010 *

These data need to be included in a report to Pacific Salmon Commission on salmon return in streams of the Lake Washington watershed

So, what now?

Do we just leave these data as they appear here?

*** pretend data**

Describing Scientific Data

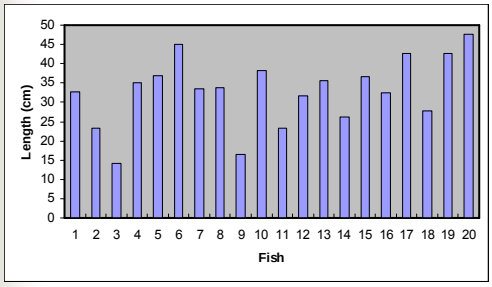
We can create a _____

Length of coho salmon returning to North Creek October 8, 2003

Describing Scientific Data

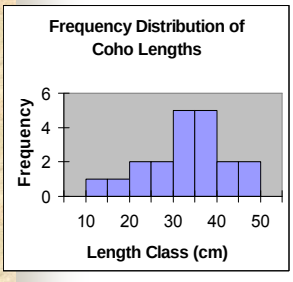
Is this a frequency distribution ?

Length of coho salmon returning to North Creek October 8, 2010



Describing Scientific Data

This is also known as a "Histogram"



Length of coho salmon returning to North Creek October 8, 2010 *

These data need to be included in a report to Pacific Salmon Commission on salmon return in streams of the Lake Washington watershed

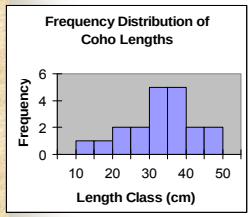
What information does a frequency distribution reveal?

Next steps?

*** pretend data**

Describing Scientific Data

Length of coho salmon returning to North Creek October 8, 2010



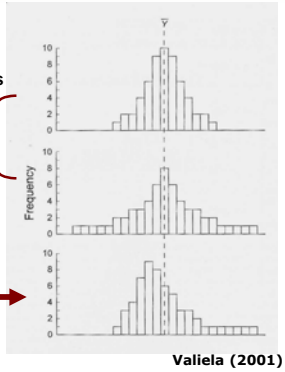
Describing Scientific Data

Mean values can hide important information

Populations with
- the same mean value but
- different frequency distributions

Normal distribution of data

Non-normal distribution



Valiela (2001)

Describing Scientific Data

"Problems" with non-normal distributions

Normal distributions are desirable and required for many statistical methods.

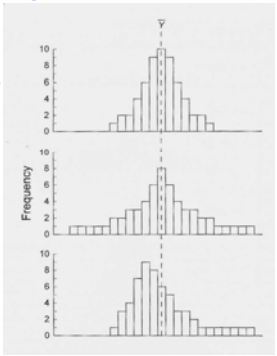
Even simple comparisons of mean values is NOT good if distributions underlying those means are not "normal".

Thus, data with non-normal distributions are usually mathematically transformed to create a normal distribution.

Describing Scientific Data

Mean values can hide important information

Populations with
- the same mean value but
- different frequency distributions



Describing Scientific Data

Expressing Variation in a Set of Numbers

Range: difference between largest and smallest sample
(or sometimes expressed as both smallest and largest values)

32.6 cm	23.2
23.2	31.6
14.1	35.6
35.2	26.2
36.8	36.7
45.1	32.4
33.5	42.6
33.9	27.8
16.6	42.8
38.2	47.6

Range = 33.5

Range = 14.1 – 47.6

Describing Scientific Data

Expressing Variation in a Set of Numbers

Variance & Standard Deviation: single-number expressions of the degree of spread in the data

32.6 cm	23.2
23.2	31.6
14.1	35.6
35.2	26.2
36.8	36.7
45.1	32.4
33.5	42.6
33.9	27.8
16.6	42.8
38.2	47.6

Variance

$x - \bar{x}$

For each value, calculate its deviation from the mean

Describing Scientific Data

Expressing Variation in a Set of Numbers

Variance & Standard Deviation: single-number expressions of the degree of spread in the data

32.6 cm	23.2
23.2	31.6
14.1	35.6
35.2	26.2
36.8	36.7
45.1	32.4
33.5	42.6
33.9	27.8
16.6	42.8
38.2	47.6

Variance

$(x - \bar{x})^2$

Square each deviation to get absolute value of each deviation

Describing Scientific Data

Expressing Variation in a Set of Numbers

Variance & Standard Deviation: single-number expressions of the degree of spread in the data

32.6 cm	23.2
23.2	31.6
14.1	35.6
35.2	26.2
36.8	36.7
45.1	32.4
33.5	42.6
33.9	27.8
16.6	42.8
38.2	47.6

Variance

$$\frac{\sum (x - \bar{x})^2}{n - 1}$$

Add up all the deviations and divide by the number of values (to get average deviation from the mean)

Describing Scientific Data

Expressing Variation in a Set of Numbers

Variance : An expression of the mean amount of deviation of the sample points from the mean value

Example of 2 data points: 20.0 & 28.6

1. Calculate the mean: (20.0 + 28.6)/2 = 24.3
2. Calculate the deviations from the mean: 20.0 – 24.3 = - 4.3
   28.6 – 24.3 = 4.3
3. Square the deviations to remove the signs (negative): $(-4.3)^2 = 18.5$
   $(4.3)^2 = 18.5$
4. Sum the squared deviations: 18.5 + 18.5 = 37.0
5. Divide the sum of squares by the number of samples (minus one*) to standardize the variation per sample: 37.0 / (2-1) = 37.0

Variance = 37.0

* Differs from textbook

Describing Scientific Data

Expressing Variation in a Set of Numbers

Standard Deviation (SD): Also an expression of the mean amount of deviation of the sample points from the mean value

SD = $\sqrt{\text{variance}}$

Using the square root of the variance places the expression of variation back into the same units as the original measured values

$(37.0)^{0.5} = 6.1$

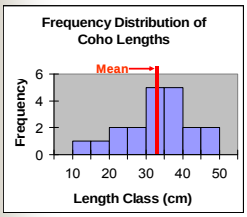
SD = 6.1

Describing Scientific Data

SD captures the degree of spread in the data

32.8 ± 8.9
Mean Standard Deviation

Conventional expression
Mean ± 1 SD



Length of coho salmon returning to North Creek October 8, 2003

Describing Scientific Data

The Bottom Line on expressing variation

Expression	What it is	When to use it
Range	Spread of data	When extreme absolute values are of importance
Variance	Average deviation of samples from the mean	Often an intermediate calculation – not usually presented
Standard Deviation	Average deviation of samples from the mean on scale of original values	Simple & standard presentation of variation; okay when mean values for comparison are similar in magnitude

Bottom Line: use the most appropriate expression; not just what everyone else uses!

The importance of knowing the degree of variation

Why are measures of variation important?

The importance of knowing the degree of variation

Density of sockeye salmon spawning sites along two area creeks (mean ± SD)

	Spawning site density (# / meter)
North Creek	0.40 ± 5.2 a
Bear Creek	0.42 ± 0.2 a

What can you conclude from the means?

What can you conclude from the SDs?

The importance of knowing the degree of variation

Mean annual air temperature at sites in eastern and western WA at 1,000 feet elevation (mean ± SD)

	Annual temperature (°F)
Darrington (western WA)	49.0 ± 11.2 a
Leavenworth (eastern WA)	48.4 ± 24.7 a

What can you conclude from the means?

What can you conclude from the SDs?

The value of examining raw data

Example Problem: Do fluctuating water levels affect the ecology of urban wetlands?

Types of variables?

Wetland	# Plant Species	Mean WLF (cm)
1	11	2
2	10	5
3	11	6
4	2	19
5	2	12
6	5	7
7	10	12
8	10	13
9	10	46
10	2	50

How might we describe these data?

The value of examining raw data

Example Problem: Do fluctuating water levels affect the ecology of urban wetlands?

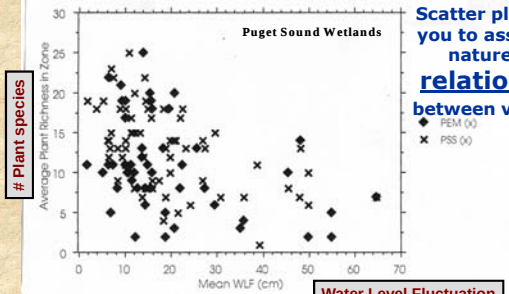


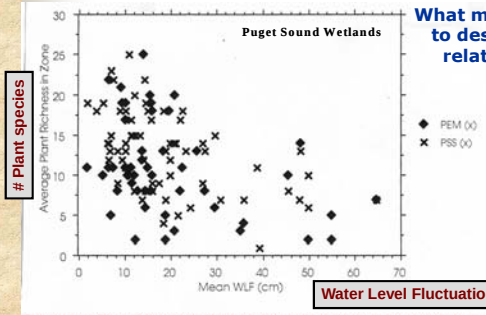
Figure 3-5. Plant richness in the emergent and scrub-shrub communities related to mean annual WLF.

(Cooke & Azous 2001)

Scatter plots help you to assess the nature of a relationship between variables

The value of examining raw data

Example Problem: Do fluctuating water levels affect the ecology of urban wetlands?

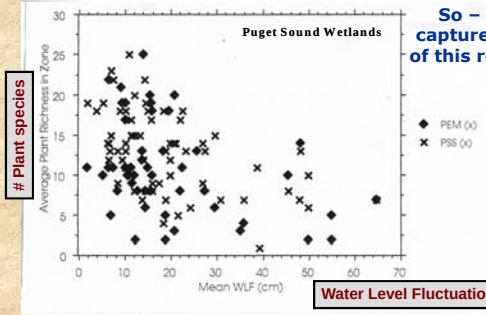


What might you do to describe this relationship?

Figure 3-5. Plant richness in the emergent and scrub-shrub communities related to mean annual WLF.

The value of examining raw data

Example Problem: Do fluctuating water levels affect the ecology of urban wetlands?



So – does this capture the nature of this relationship?

Figure 3-5. Plant richness in the emergent and scrub-shrub communities related to mean annual WLF.

Describing Scientific Data

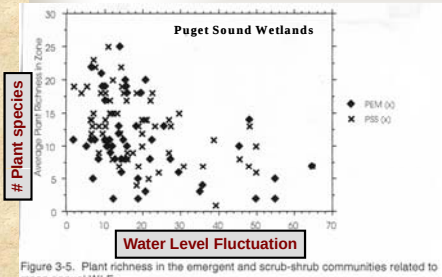


Figure 3-5. Plant richness in the emergent and scrub-shrub communities related to mean annual WLF.

Describing Scientific Data

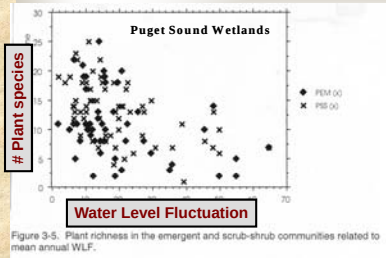


Figure 3-5. Plant richness in the emergent and scrub-shrub communities related to mean annual WLF.

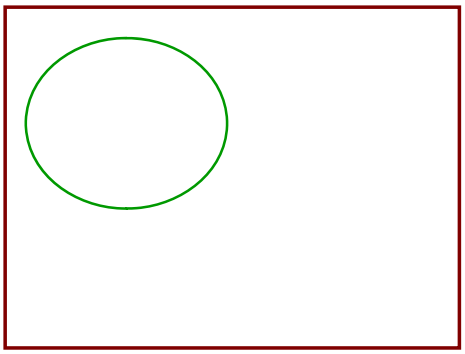
BOTTOM LINE

Sampling the World

Rarely can we measure everything – Instead we usually SAMPLE the world

The "entire world" is called the "population"

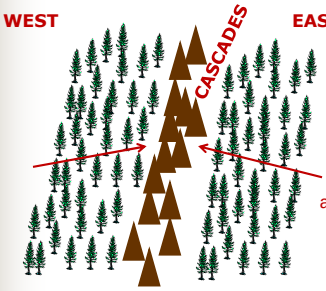
What we actually measure is called the "sample"



Populations & Samples

QUESTION

Does Douglas-fir grow taller on the east slopes of the Cascades or on the west slopes of the Cascades?

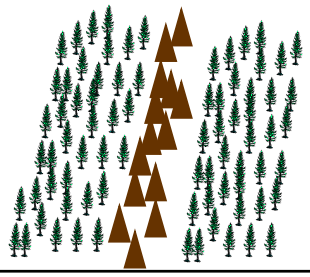


How would you study this?

Populations & Samples

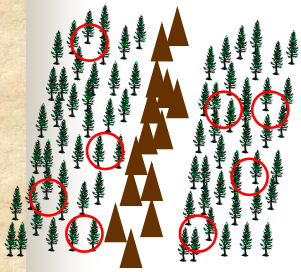
Does Douglas-fir grow taller on the east slopes of the Cascades or on the west slopes of the Cascades?

1.



Populations & Samples

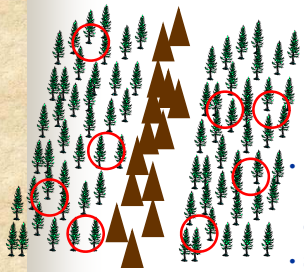
Does Douglas-fir grow taller on the east slopes of the Cascades or on the west slopes of the Cascades?



2.

Populations & Samples

Does Douglas-fir grow taller on the east slopes of the Cascades or on the west slopes of the Cascades?



3. A major research question:

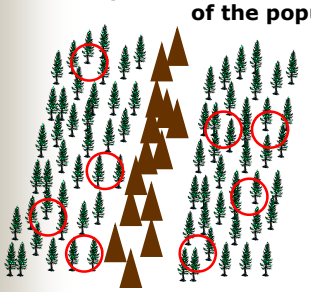
How many samples do you need to accurately describe (1) each population and (2) their possible difference?

- **Depends on sample variability** (standard error – see textbook pages 150-152) **and degree of difference between populations**
- **Sample size calculations before study are important** (statistics course for more detail)

Populations & Samples

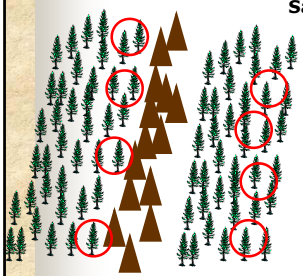
Does Douglas-fir grow taller on the east slopes of the Cascades or on the west slopes of the Cascades?

4. Samples are taken that are representative of the population



Populations & Samples

Does Douglas-fir grow taller on the east slopes of the Cascades or on the west slopes of the Cascades?

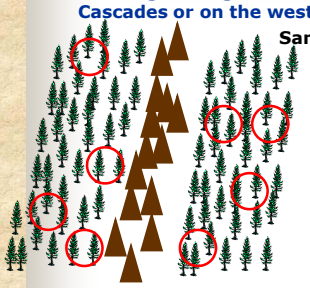


Samples are taken representative of the population

What criteria are used to locate the samples?

Populations & Samples

Does Douglas-fir grow taller on the east slopes of the Cascades or on the west slopes of the Cascades?

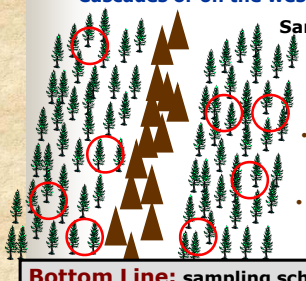


Samples are taken representative of the population

What criteria are used to locate the samples?

Populations & Samples

Does Douglas-fir grow taller on the east slopes of the Cascades or on the west slopes of the Cascades?



Samples are taken representative of the population

What criteria are used to locate the samples?

- **Randomization**
  - ✓ Commonly, but not in all studies (statistics course)
- **Stratification**
  - ✓ Accounts for uncontrolled variables (e.g., elevation, aspect)

Bottom Line: **sampling scheme needs to be developed INTENTIONALLY. It must match the question & situation, not what is "usually done".**

SOME BOTTOM LINES FOR THE LAST 2 DAYS

Describing & Taking Data	Conventional Wisdom	Best Practices
Examine the data	Compare averages	Look at raw data as well as summaries
Summarize the data	Take averages	Frequency distributions, Means, etc.
Describe variation	Use Standard Deviation	Use measure appropriate to need
Use variation data	As an adjunct to means (testing for differences)	Use for understanding system as well as for statistical tests
Describing relationships	Fit a line to data	Use approach appropriate to need; examine raw data
Creating a sampling scheme	Take random samples	Use approach appropriate to need