Stats for AP Biology

AP BIOLOGY EQUATIONS AND FORMULAS

Statistical Analysis and Probability

Mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Standard Deviation*

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

Standard Error of the Mean*

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Chi-Square

$$\chi^2 = \sum \frac{(o-e)^2}{e}$$

Chi-Square Table

p value	Degrees of Freedom							
	1	2	3	4	5	6	7	8
0.05	3.84	5.99	7.82	9.49	11.07	12.59	14.07	15.51
0.01	6.64	9.21	11.34	13.28	15.09	16.81	18.48	20.09

$\bar{x}$ = sample mean

n = size of the sample

s = sample standard deviation (i.e., the sample-based estimate of the standard deviation of the population)

o = observed results

e = expected results

Degrees of freedom are equal to the number of distinct possible outcomes minus one.

SLIDE SHOWS MODIFIED FROM:
http://apbiowarde.weebly.com/uploads/3/8/3/0/38303939/stats_pp.pptx

https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved=0ahUKEwiBgJ6ntcTXAhWHTSYKHUT_DWAQFggvMAE&url=http%3A%2F%2Fwww.bville.org%2Ftfiles%2Ffolder427%2Fap_standard_deviation_and_standard_error.pptx&usg=AOvVaw3J7ssjoCq2jMZGlQ3dcAmW

Bozeman Biology videos:

AP Science Practice 2: Using Math Appropriately

Statistics for Science

Standard Deviation

Standard Error

Modified from: Mathisfun.com

http://www.mathsisfun.com/data/standard-deviation.html

Data analysis allows you to…

- **Arrive at conclusions about your data**
- **Make claims about your data**
- **Support arguments using your data**
- **Estimate the reliability of your data**
- **Effectively communicate conclusions about your work to a larger scientific community**

Appropriate descriptive statistics for a data set typically include:

- Mean (average)
- Sample size
- Standard deviation
- Standard error

Mean (average)

- Sum of the numbers in the sample divided by the total number in the sample
- Summarizes the entire sample
- Might provide an estimate of the entire population's (that was sampled) true mean

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Sample size

- How many members of the population are included in the study
- Important when determining confidence that analysis of sample set is representative of entire population
- In formulas, sample size = n

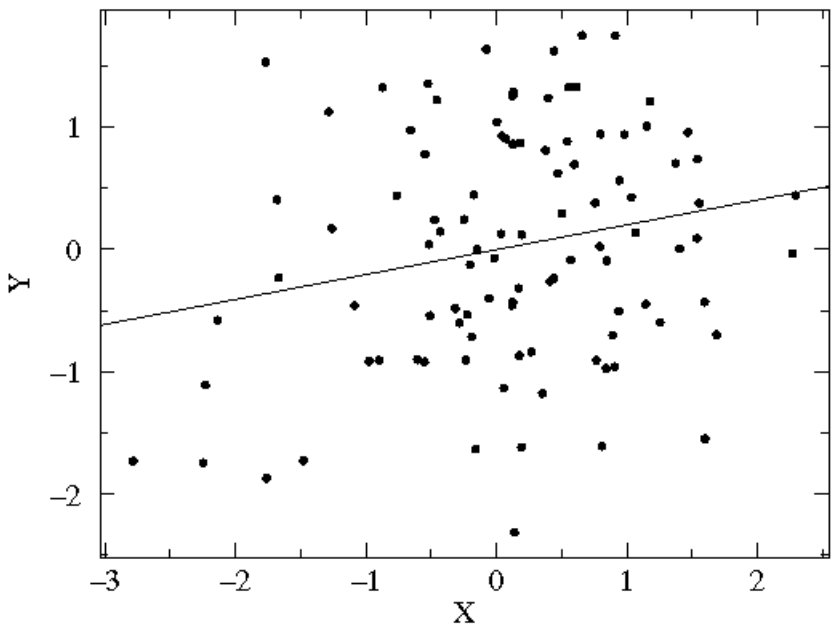
Standard deviation

- Tool for measuring the spread (variance) in the sample population
- Large standard deviation indicates that the data have a lot of variability
- Small standard deviation indicates that the data are clustered close to the sample mean

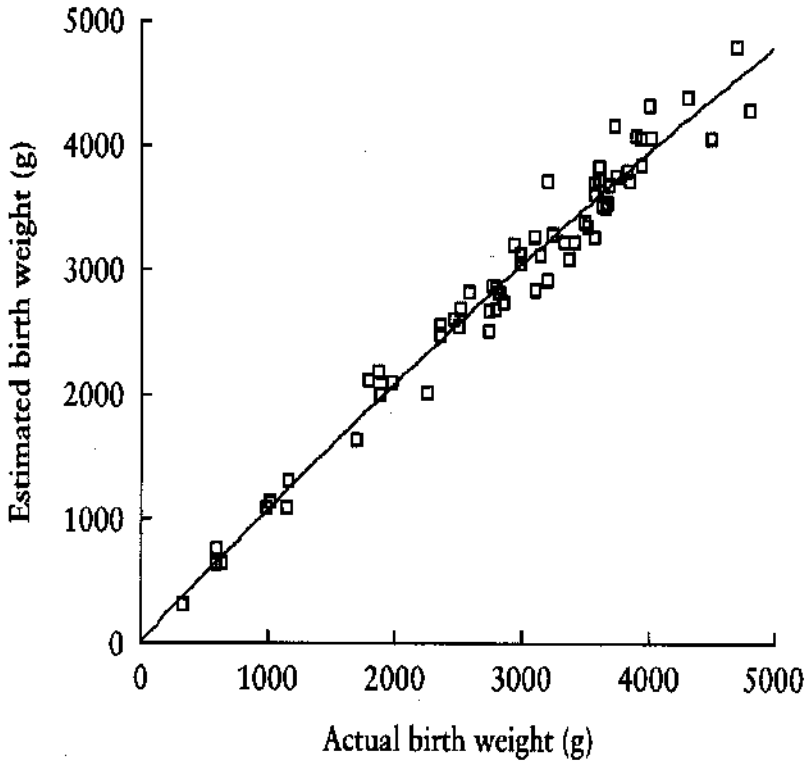
$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

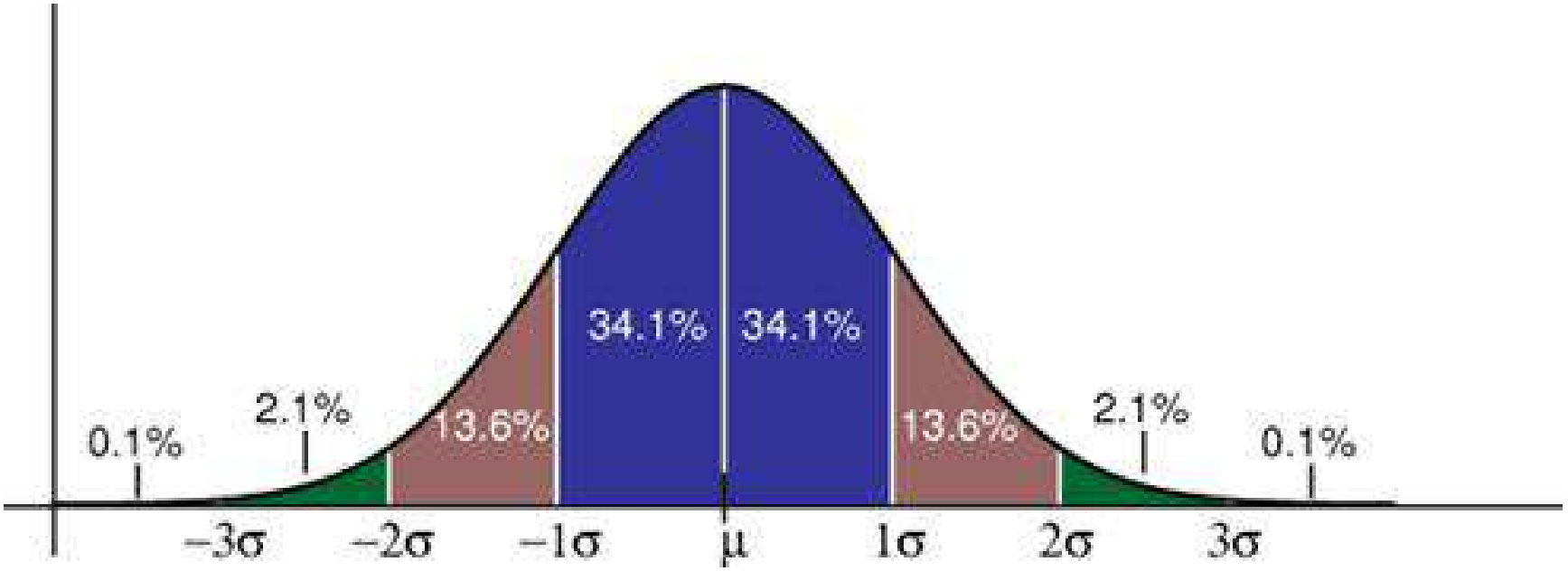
You do NOT have to calculate this on the AP Biology Exam, but you should understand how it is derived and used.

Large standard deviation (spread out from mean)



Small standard deviation (clustered close to mean)





In a normal distribution

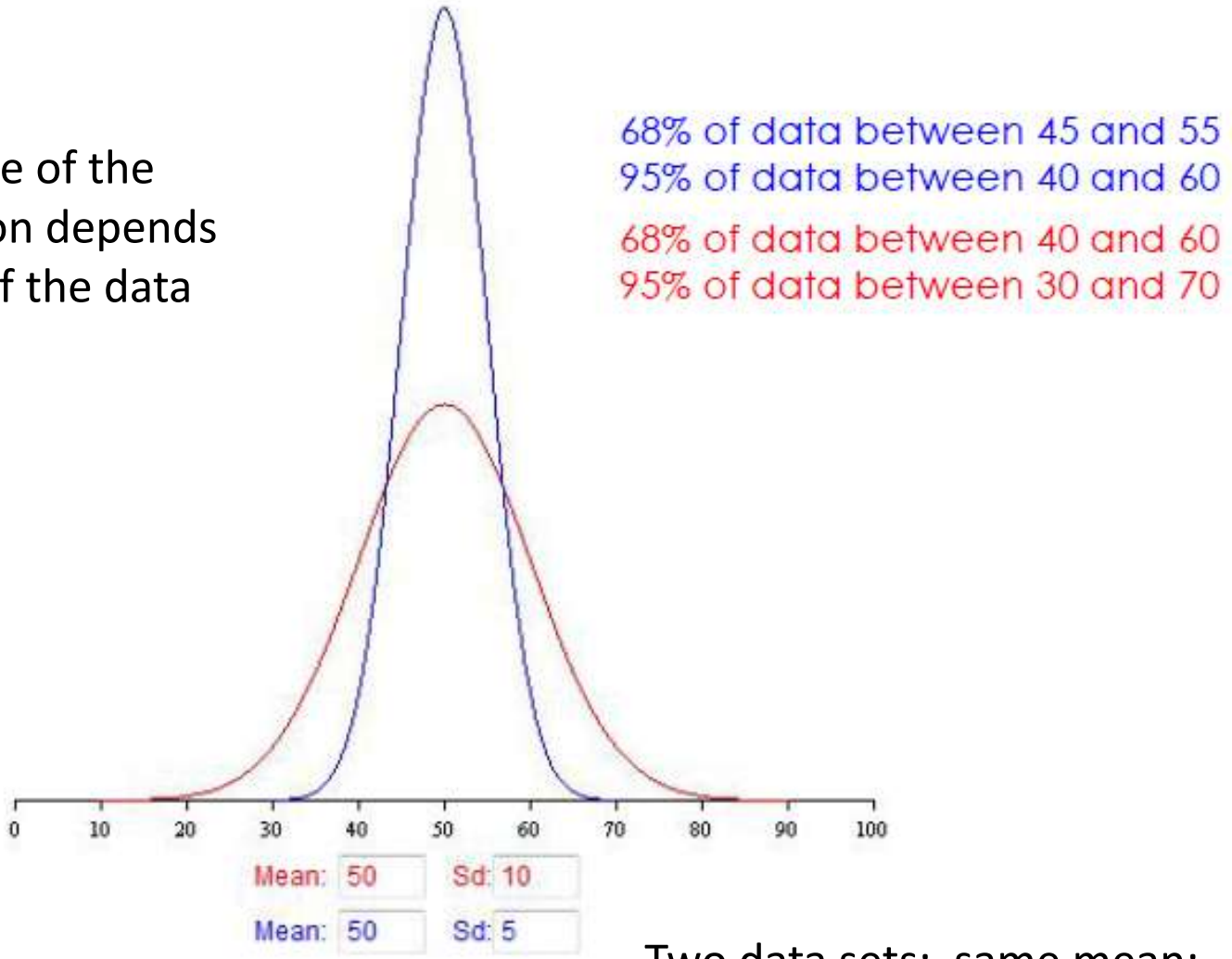
68.2% within 1 standard deviation from mean

95.4% within 2 standard deviations from mean

99.7% within 3 standard deviations from mean

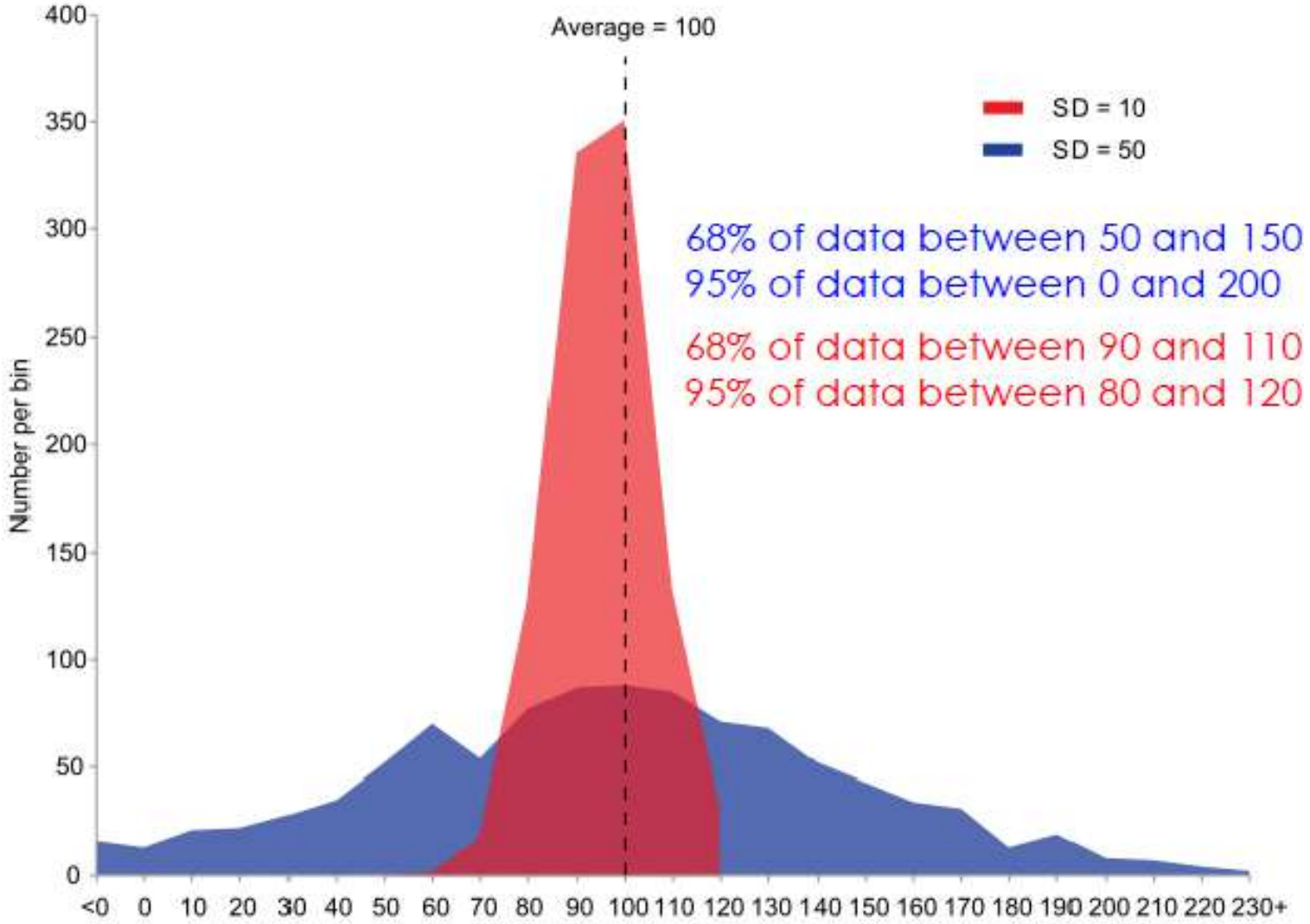
http://www.strath.ac.uk/media/faculties/hass/appliededucationalresearch/mod4unit3/curve1.jpg

The magnitude of the standard deviation depends on the spread of the data set



Two data sets: same mean; different standard deviation

Actual data sets aren't always so pretty...



Standard Error of the Mean (Standard Error)

- Allows you to infer how well the sample mean matches up to the true population mean
- Helps you to determine confidence in the data collected in a sample
- **95% confidence interval = ± 2 SE**

(Random sampling of the population should produce a mean that falls within ± 2 SE 95% of the time.)

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

You do NOT have to calculate this on the AP Biology Exam, but you should understand how it is derived and used.

Example Problem

A student noticed that the ivy leaves growing on the shady side of a building were larger than ivy leaves growing on the sunny side of the same building. The student collected and measured the maximum width, in centimeters, of 30 leaves from each habitat. Use statistical analysis to determine if it's likely that there is a significant difference in leaf size between the shady and sunny ivy plants with 95% confidence (±2 SE).

Collected Data

	Shady Leaves	Sunny Leaves
Mean	7.43	5.88
Standard Deviation	1.63	1.32
N	30	30
Standard Error	0.30	0.24

Collected Data

	Shady Leaves	Sunny Leaves
Mean	7.43	5.88
Standard Deviation	1.63	1.32
N	30	30
Standard Error	0.30	0.24
2 SE	0.60	0.48

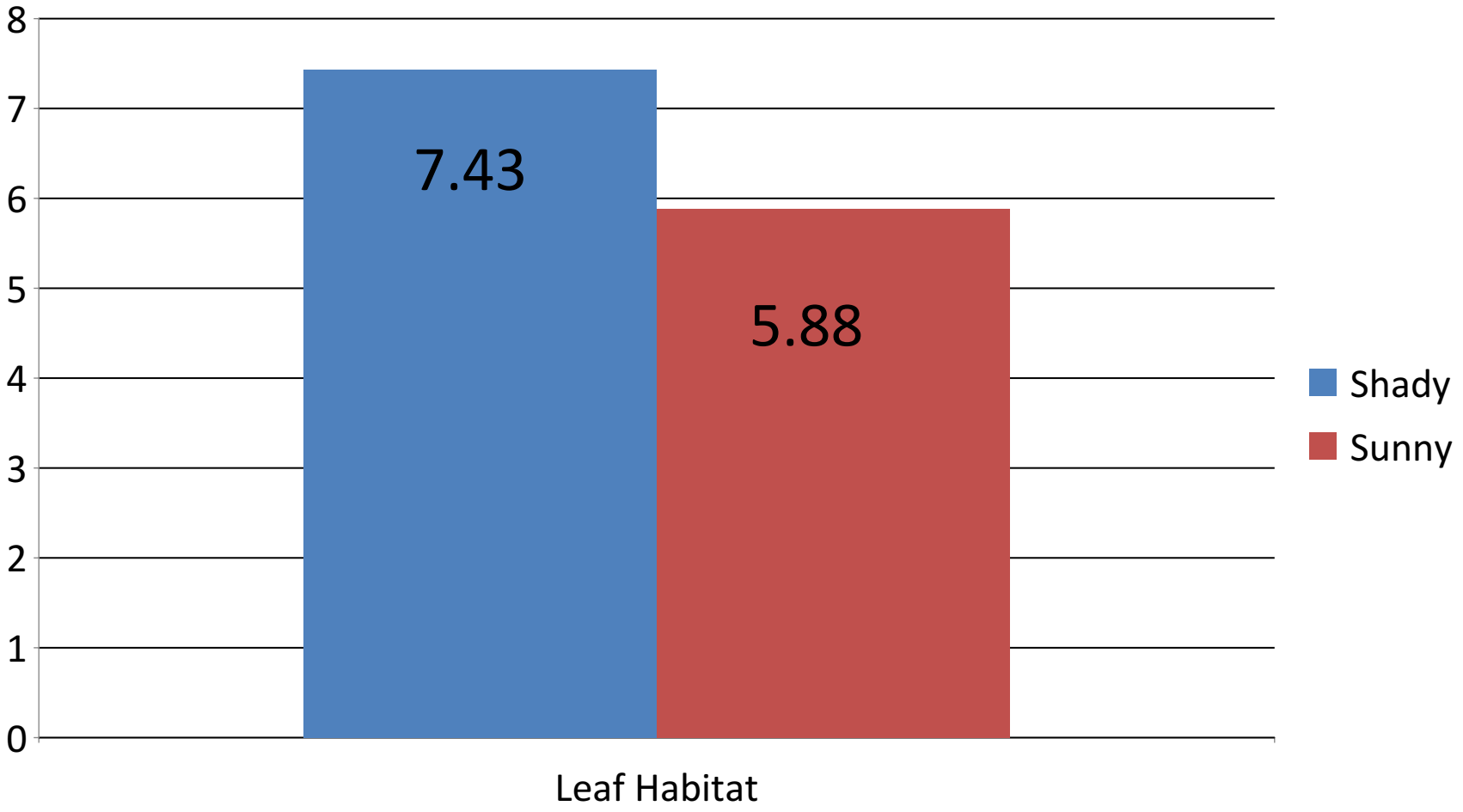
95% confidence

Collected Data

Graph Mean

	Shady Leaves	Sunny Leaves
Mean	7.43	5.88
Standard Deviation	1.63	1.32
N	30	30
Standard Error	0.30	0.24
2 SE	0.60	0.48

Graph Means

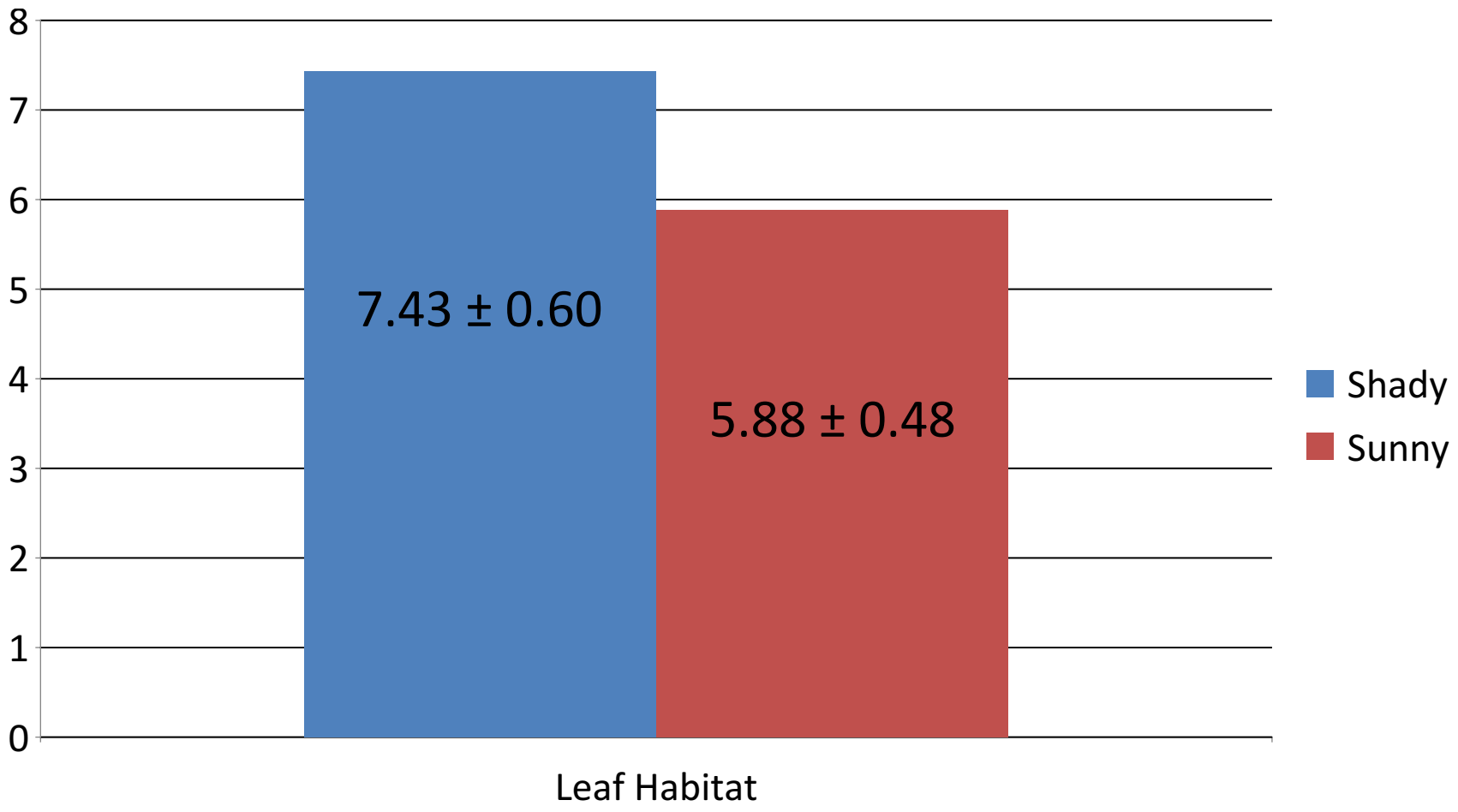


Collected Data

	Shady Leaves	Sunny Leaves
Mean	7.43	5.88
Standard Deviation	1.63	1.32
N	30	30
Standard Error	0.30	0.24
2 SE	0.60	0.48

Add ± 2 SE Error Bars

Add ± 2 SE Error Bars



Conclusions

- **±2 SE Error bars (95% confidence intervals) do NOT overlap between sunny and shady means (7.43 - 0.60 > 5.88 + 0.48; 6.83 > 6.36)**
- **Strongly suggests that the two populations are indeed statistically significantly different from one another**

If the error bars/confidence intervals did overlap between the groups, you could not claim a statistically significant difference.

- **Lower standard deviation:**
  - Data is **closer to the mean**
  - Greater likelihood that the independent variable is causing the changes in the dependent variable

- **Higher standard deviation:**
  - Data is more **spread out from the mean**
  - More likely factors, other than the independent variable, are influencing the dependent variable

Calculating standard deviation, s

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

1. Calculate the mean ($\bar{X}$)
2. Determine the difference between each data point, and the mean
3. Square the differences
4. Sum the squares
5. Divide by sample size (n) minus 1
6. Take the square root

Standard Error:

- Indication of **how well the mean of a sample ($\bar{x}$) estimates the true mean of a population** (μ)

- Measure of accuracy, if the true mean is known
- Measure of precision, if true mean is not known

STANDARD DEVIATION

the **standard deviation** of the sample is the degree to which individuals within the sample differ from the sample mean.

$$s = \sqrt{s^2}$$

Bozeman Biology Standard Deviation

STANDARD ERROR OF THE MEAN

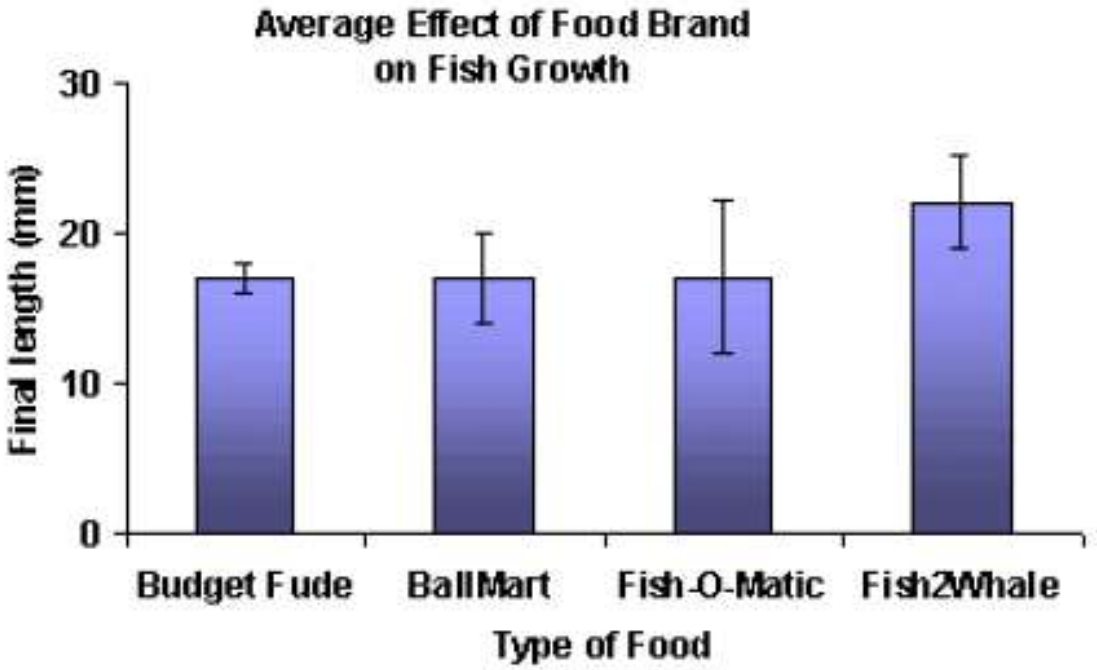
the **standard error of the mean** is an estimate of how far the sample mean is likely to be from the population mean

Bozeman Biology Standard Error

$$\mathrm{SE}_{\bar{x}} = \frac{s}{\sqrt{n}}$$

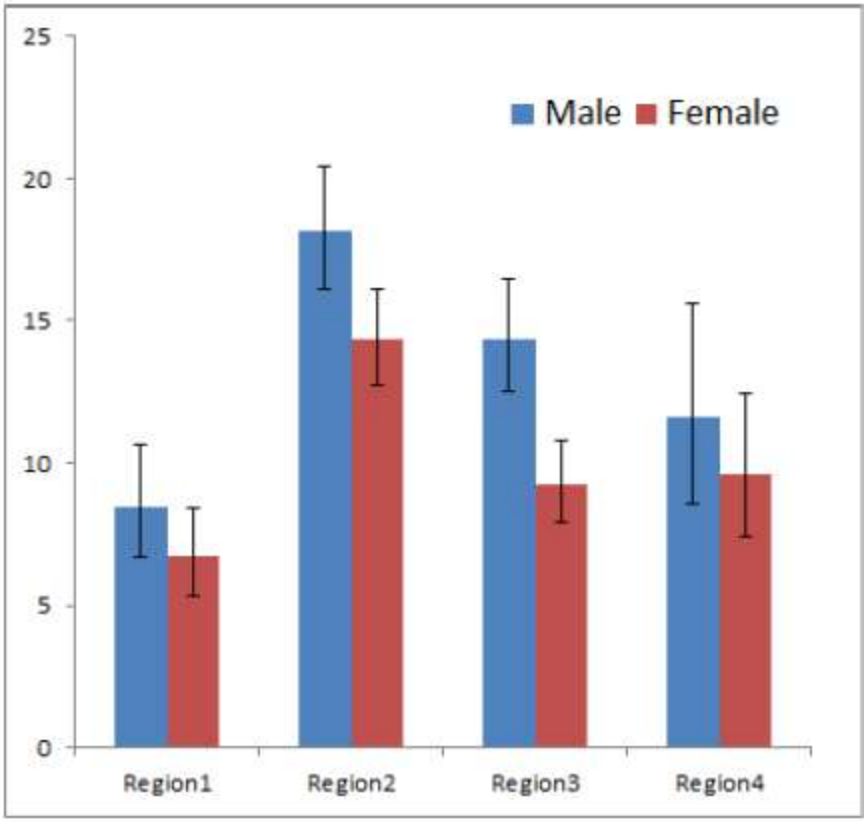
- **Accuracy** – How close a measured value is to the **actual (true) value**

- **Precision** – How close the measured values are **to each other**.



Which is a valid statement?

- Fish2Whale food caused the most fish growth
- Fish2Whale food caused more fish growth than did Budget Fude



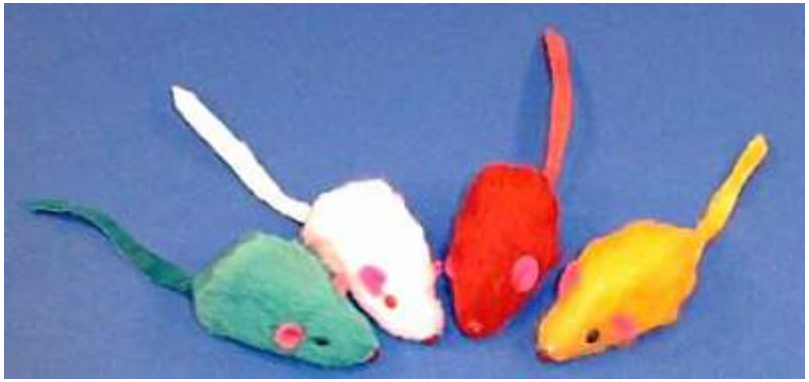
Statements:

- In all four regions, more males exhibited the trait measured than did females.

- More males in region 3 exhibited the measured trait than did females

Does the mutagen have an impact on mean tail length ($\overline{x}$) in *Mus musculus*?

- A group of mice was fed a diet with a mutagen added that was thought to affect tail length in mice.
- Another group of mice was fed the same diet w/o the mutagen added.
- H_O ?



Mutagen mice idea from Kristen Dotti
Catalyst Learning Curriculum

Image from: http://www.petdiscounters.com/assets/images/product_images/image/d_4491.jpg

MEASURING MICE

- Determine the tail lengths of the mice in the sample provided to your group.

- Determine RANGE, MEAN, STANDARD DEVIATION, STANDARD ERROR of the MEAN, 95% CONFIDENCE for the mice in your sample.
- Post your measurements on class data chart

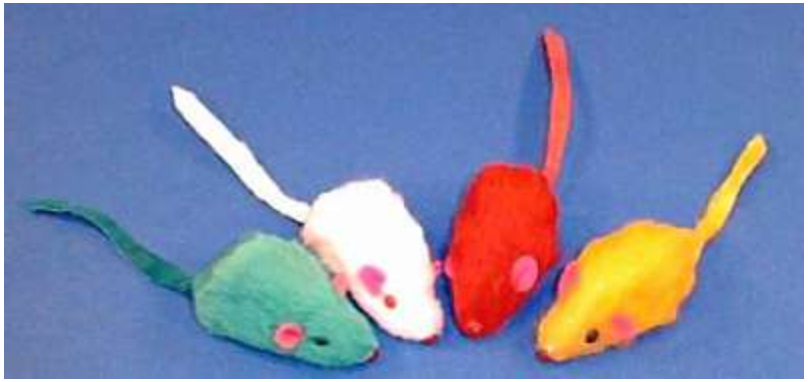
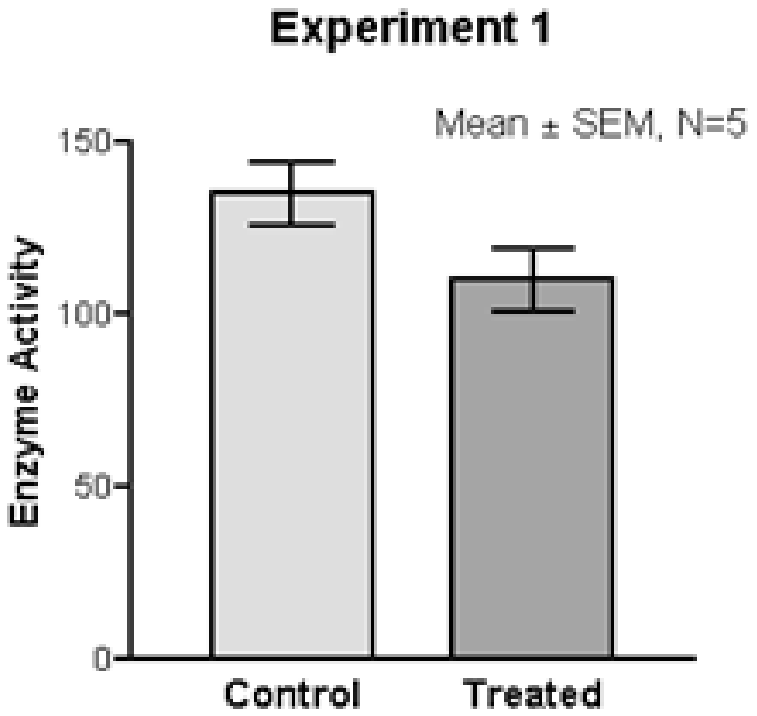
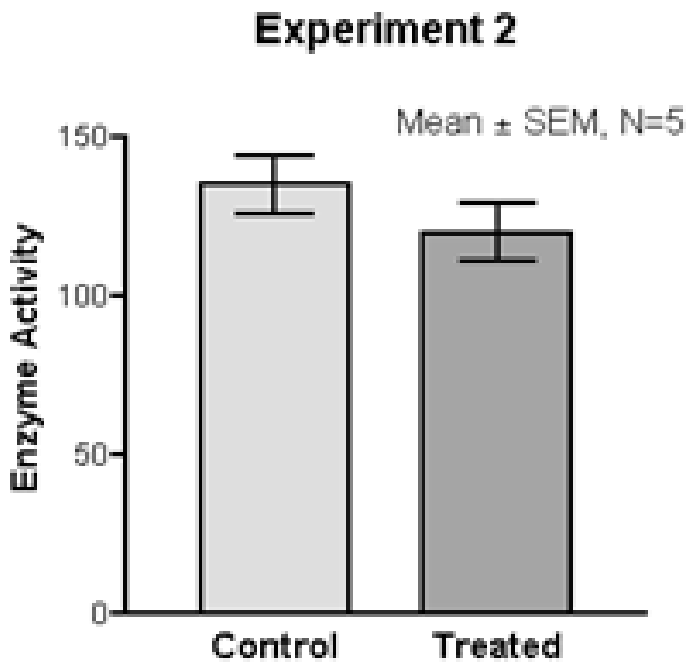


Image from: http://www.petdiscounters.com/assets/images/product_images/image/d_4491.jpg





What can you conclude about Experiment 1?

When standard error (SE) bars do not overlap, it suggests there might be a difference between the two means, but you cannot be sure. More statistics needs to be done.

What can you conclude about Experiment 2?

When SE bars overlap you can be sure there is NO statistically significant difference between the two means.