AP Biology Calculations:

Descriptive Statistics, Standard Deviation and Standard Error

SBI4UP

The Scientific Method & Experimental Design

Scientific method is used to explore observations and answer questions.

Scientists will use labs, field investigation, models, simulations and data sets to test these observations and prove whether or not a hypothesis is correct.

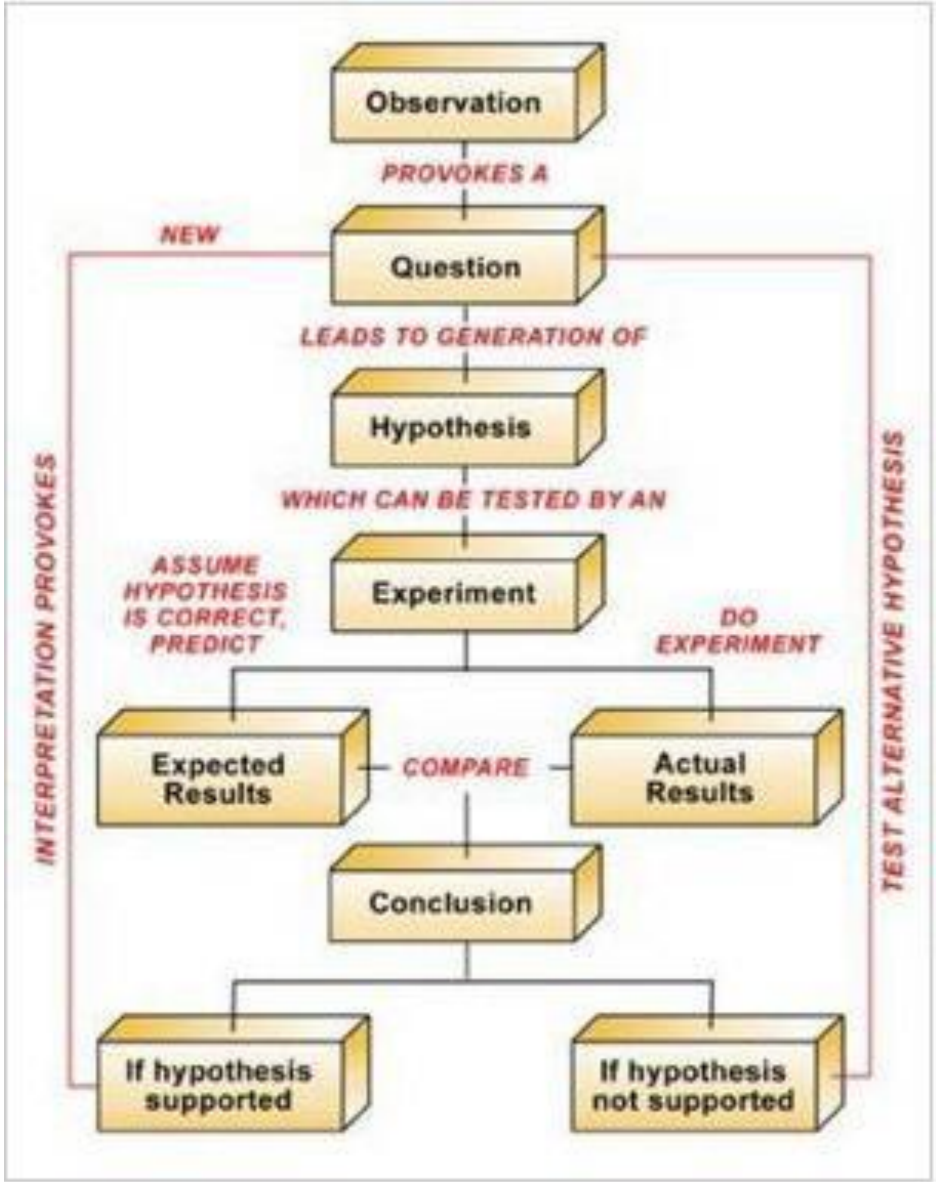
Hypothesis: is the first step in scientific method and it is based on prior knowledge and observation.

The Scientific Method:

1. Make Observation
2. Statement of problem, ask a question
3. Hypothesis: propose a tentative answer
4. Design & conduct an experiment

   (Use quantifiable data → math is extremely important)
5. Use statistical tests to evaluate the significance of your results

   (X^2 test, null hypothesis)
6. Acceptance or rejection of hypothesis.


Observation
PROVOKES A
NEW
Question
LEADS TO GENERATION OF
Hypothesis
WHICH CAN BE TESTED BY AN
Experiment
ASSUME HYPOTHESIS IS CORRECT, PREDICT
DO EXPERIMENT
Expected Results
COMPARE
Actual Results
Conclusion
If hypothesis supported
If hypothesis not supported
INTERPRETATION PROVOKES
TEST ALTERNATIVE HYPOTHESIS

Hypothesis is Scientific Method

A scientific hypothesis must have two important qualities

- *It must be testable*
- *It must have the potential of being rejected*

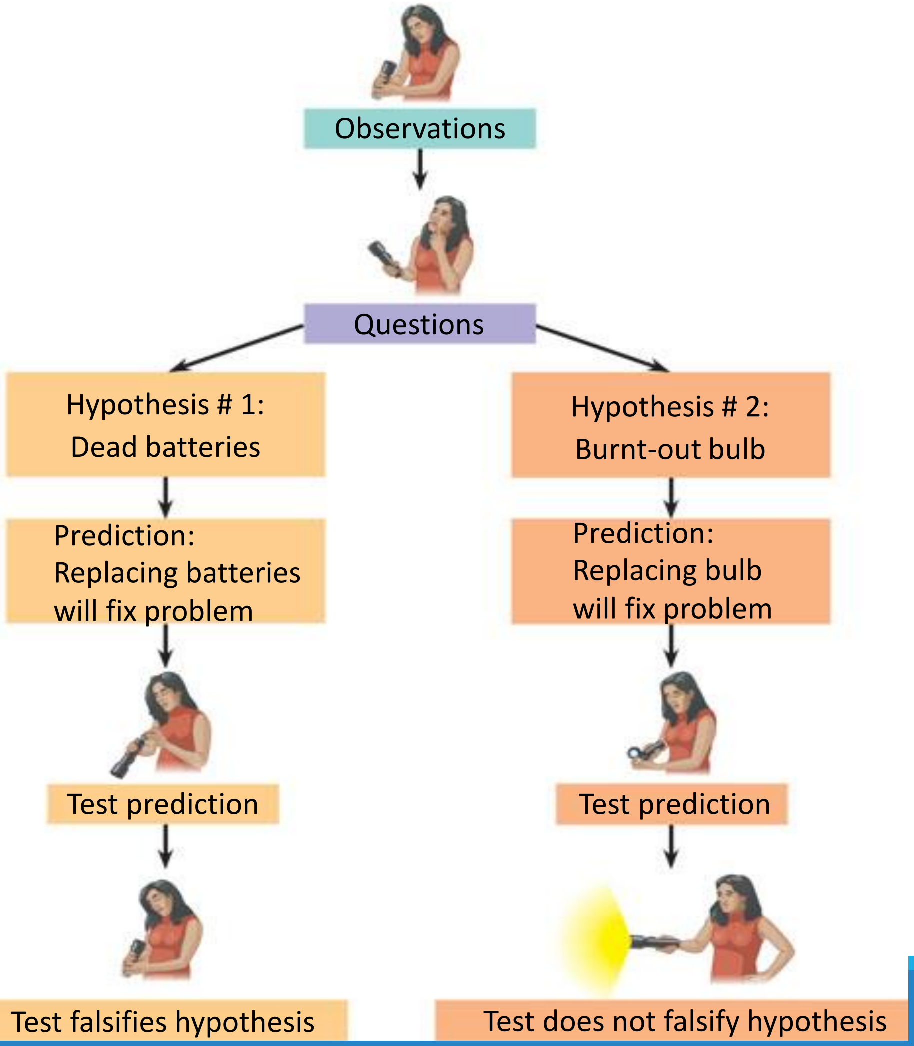


Figure 1.25

Designing Controlled Experiments

In ***controlled experiments*** we start with two (or more) groups that are as similar as possible and we devise a method to manipulate only ONE variable.

Independent Variable =the variable that is manipulated

Dependent Variable = the response that is measured

Data Collection

When you work with data during an experiment, you need to:

- Make accurate and precise measurements.
- Account for error in measured values
- Develop techniques and consistency for collecting data
- Understand the units and properties of the data.
    - Make observations of trends and patterns in the data.
    - Produce visual representation of data, GRAPHS and CHARTS.

Graphical Representation

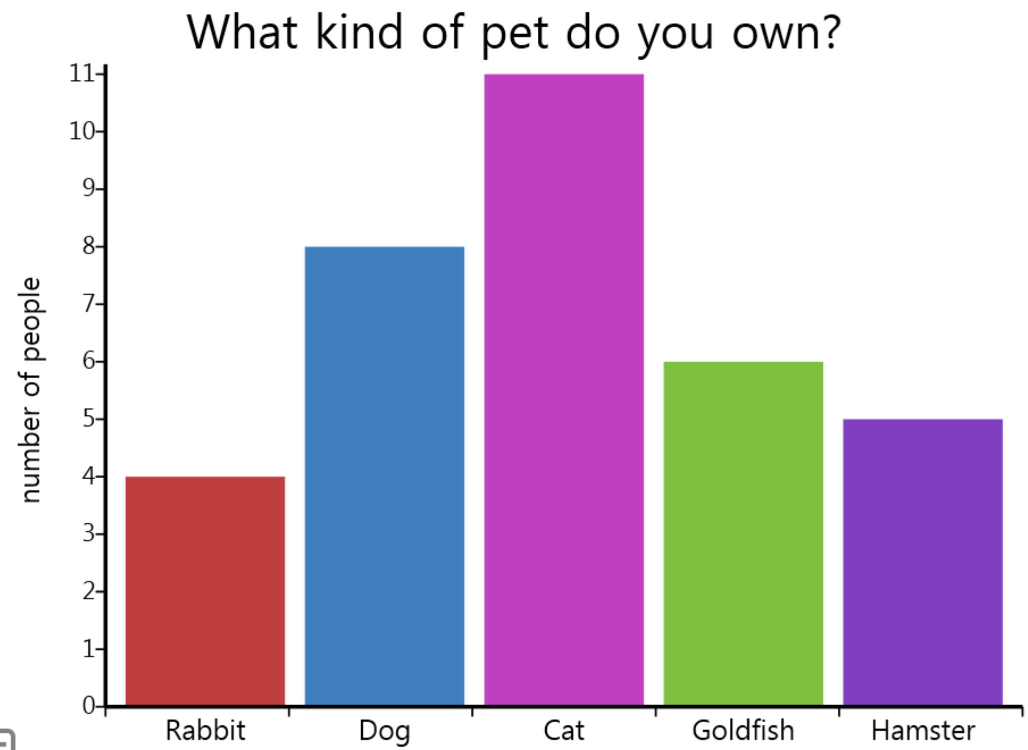
Depending on the data collected, you can represent your findings using a variety of graphs. You must select a graph that can best illustrate your findings.

Types of graphs:

- Bar graphs *
- Histograms
- Pie Graphs *
- Line graphs *
- Scatter Plots *

Bar Graphs

Bar graphs are most commonly used to represent data that does not have a numerical value and when a person needs to compare various categories side by side.

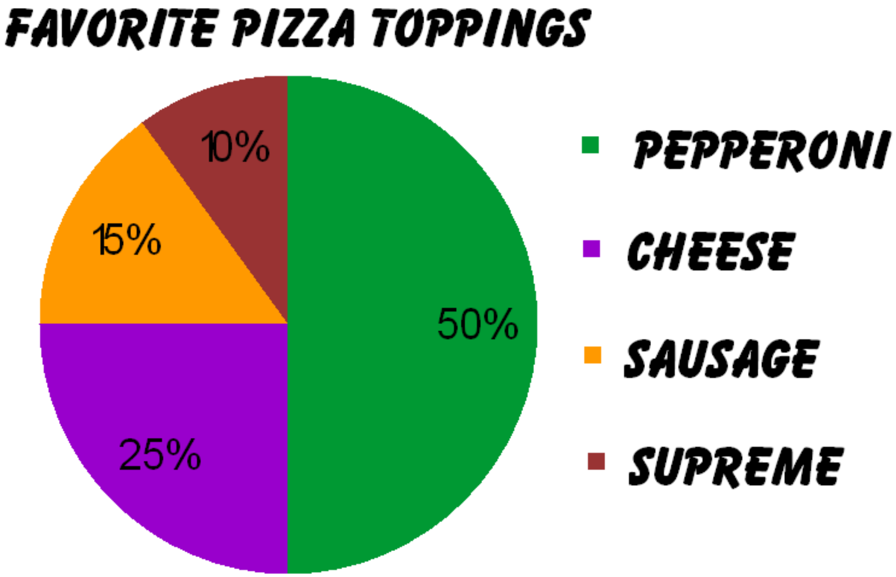


Pie Graphs

These can be used instead of bar graphs when you want the proportions of the data to have more of an impact when communicating the findings.

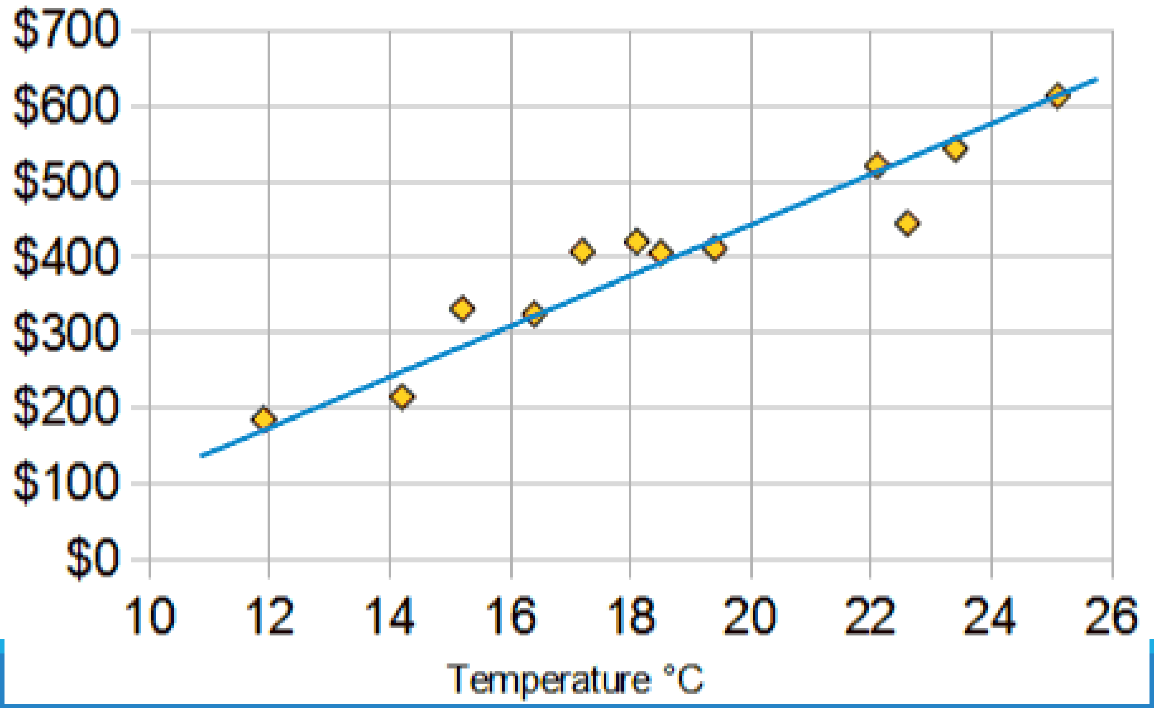
The dependent variable will have a numerical value in the form of a percentage.

Need to have 6 categories or less.



Line Graphs

Line graphs are the most commonly used graphs in scientific experiments. Most scientific experiments will have a dependent and independent variable, that can be easily depicted onto a line graph. By looking at the direction of the line one can determine the relationship of both variables.



Making a Line Graph

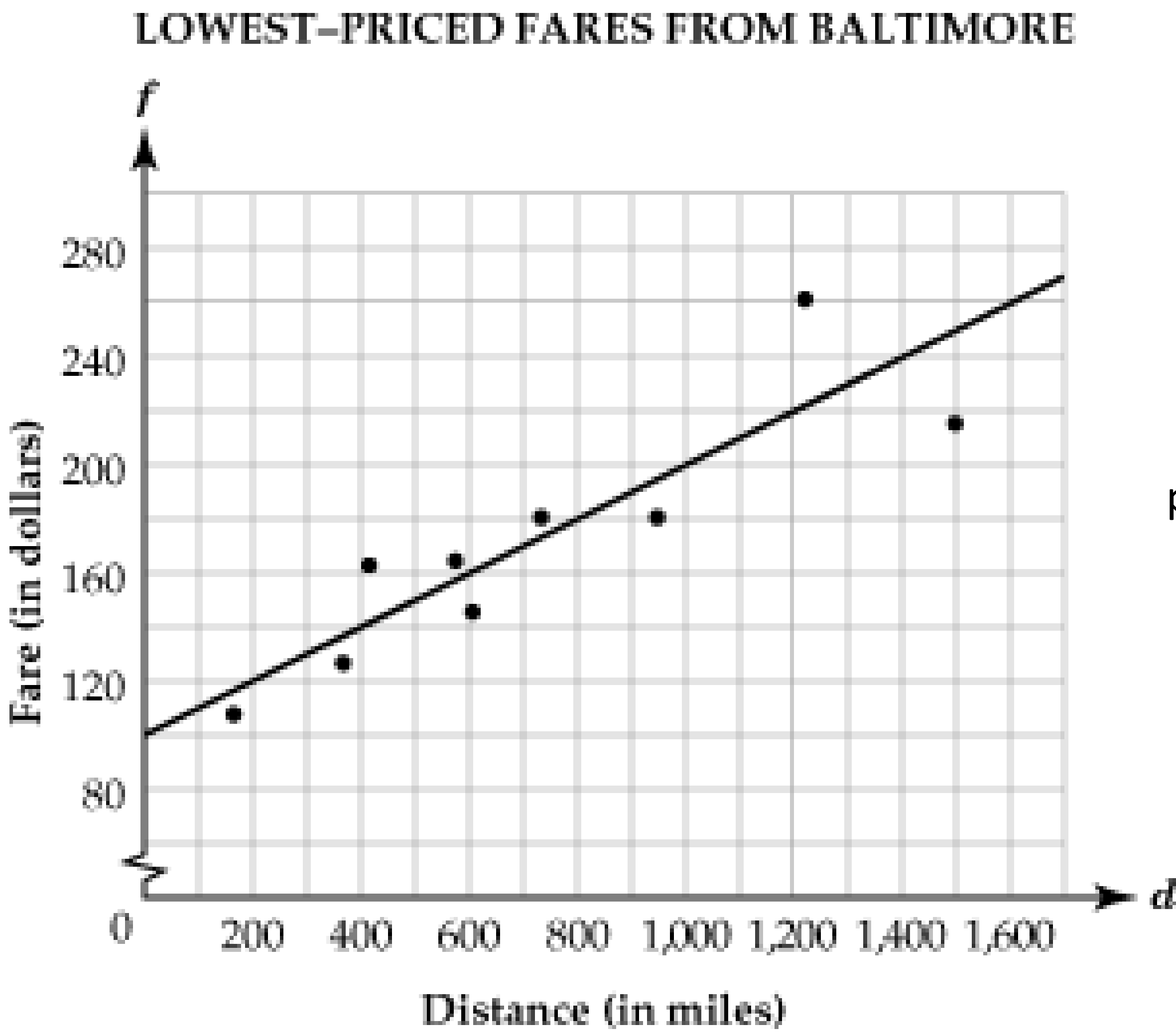
- Identify the independent and dependent variable.
- The independent variable is on the x-axis.
- The dependent variable is on the y-axis.
- Choose your scale carefully (must be consistent intervals).

  Make your graph as large as possible by spreading out the data on each axis. Let each square grid on your axis represent a convenient interval. Do not number every grid on your axis.

- Plot each point as a dark dot

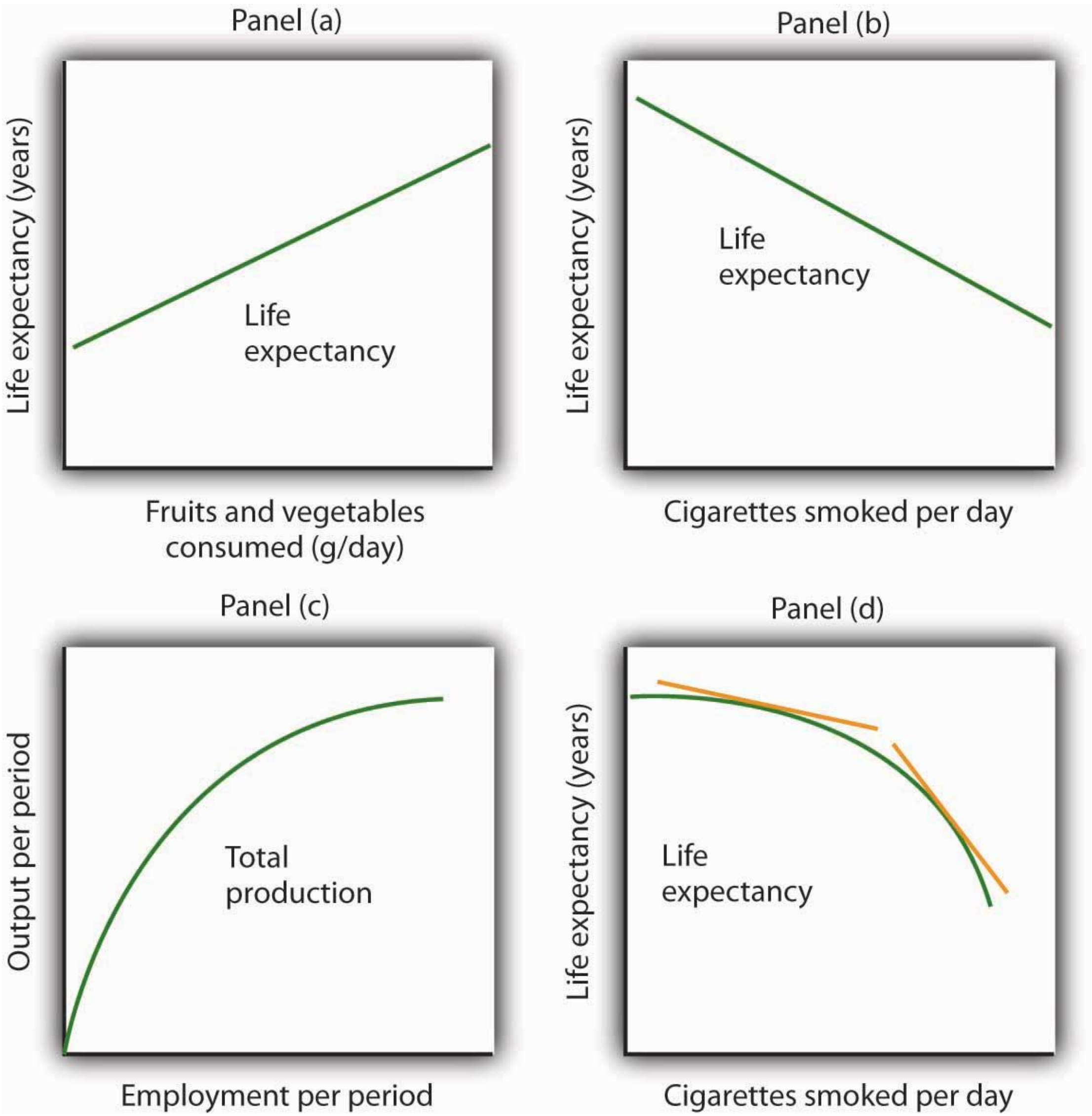
Making a Line Graph

- Label each axis with the name of the variable and the units of measure.
- Title your graph. The title should be short and clearly state the purpose of the graph. *The title could include the independent and dependent variables. General the titles are written as the y variable vs. the x variable.*
- Use a single sheet of graph paper for each graph. Do not use the back of a sheet graph paper.



Line of Best Fit:

Draw the line so that there are equal numbers of data points above and below the line you draw. Try to draw the line so that it comes as close to all data points as possible.

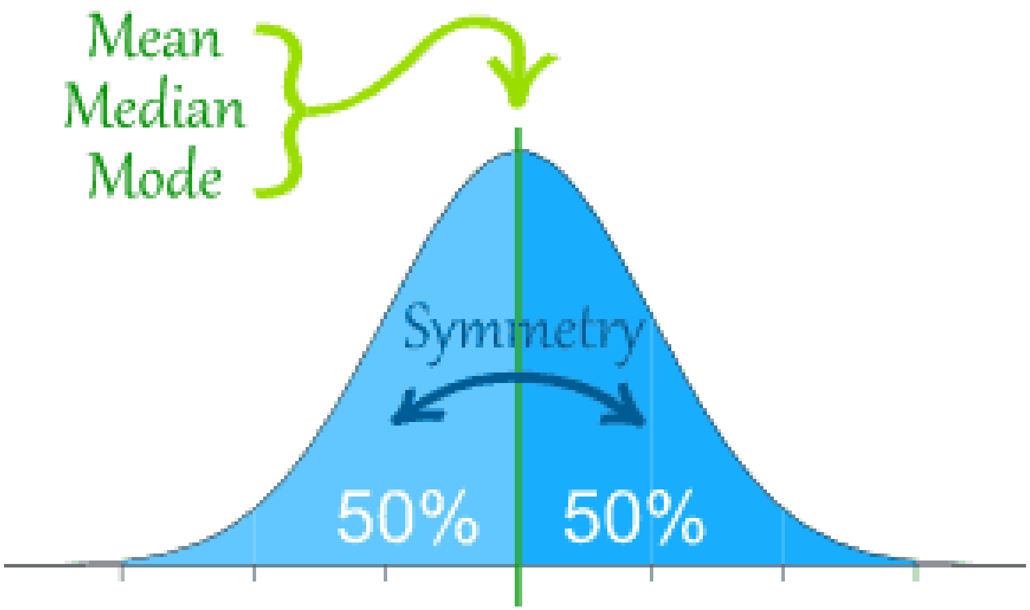


Different type of relationships between the dependent and independent variable.

Descriptive Statistics

- **Descriptive statistics** are used to describe the basic features of the data in a study.
- They provide simple summaries about the sample and the measures.
- Together with graphics analysis, they form the basis of virtually every quantitative analysis of data.

The Normal Curve



Properties of the normal curve:

1) Represents a frequency distribution of a large population
2) The graph is symmetrical and bell-shaped
3) Most of the data occurs around the mean, mode and median
4) A small portion of the data occurs at the tail of the curve

Measures of Central Tendency

Mean: the average of all data points (take the sum of all values and divide by the total amount of values

- **Eg.** 1, 1, 2, 3, 3, 5, 5, 5, 6, 7, 7, 7, 7 ,9, 9
- **Mean** =

Mode: the most frequent observation

- **Eg.** 1, 1, 2, 3, 3, 5, 5, 5, 6, 7, 7, 7, 7 ,9, 9
- **Mode** =

Median: the number in the middle of an ordered series of numbers

- **Eg.** 1, 1, 2, 3, 3, 5, 5, 5, 6, 7, 7, 7, 7 ,9, 9
- **Median** =

Practice Activity

1. 12, 15, 12, 97, 46, 88

   Mean: ________
   Median: ________
   Mode: ________
   Range: ________

2. 33, 76, 37, 92, 92, 88

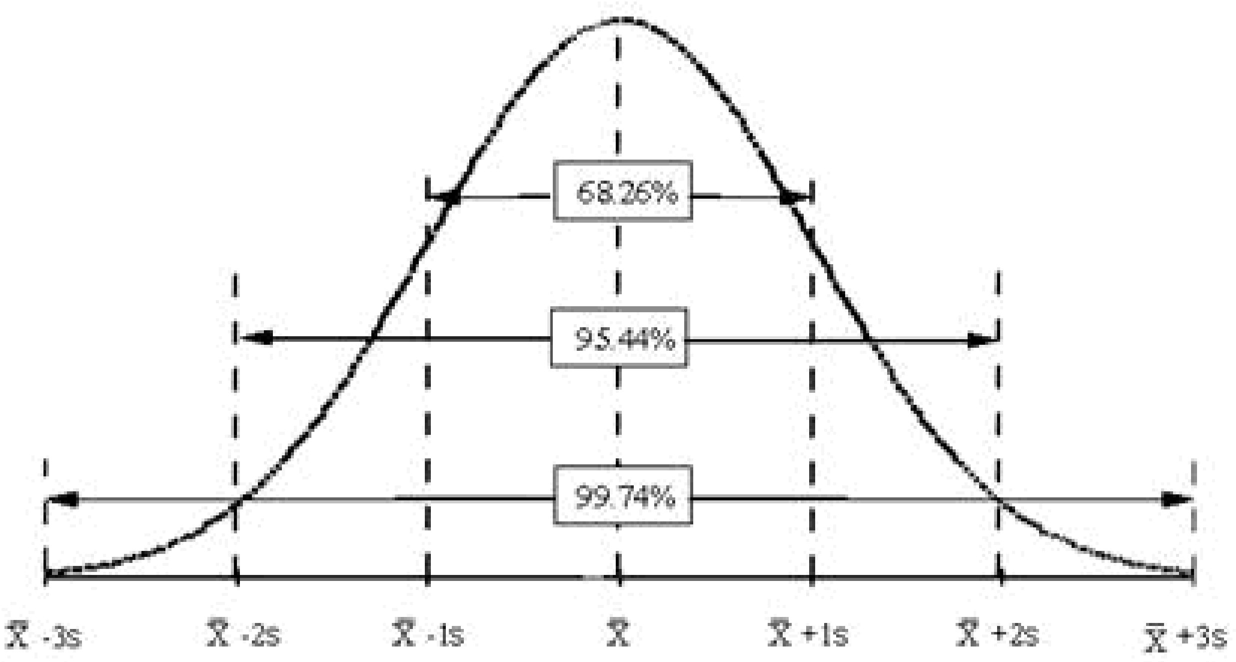
   Mean: ________
   Median: ________
   Mode: ________
   Range: ________

3. 4, 12, 4, 77, 4, 4

   Mean: ________
   Median: ________
   Mode: ________
   Range: ________

Standard Deviation (s)

A measure of **how spread out** the data is from the mean



Standard Deviation (s)

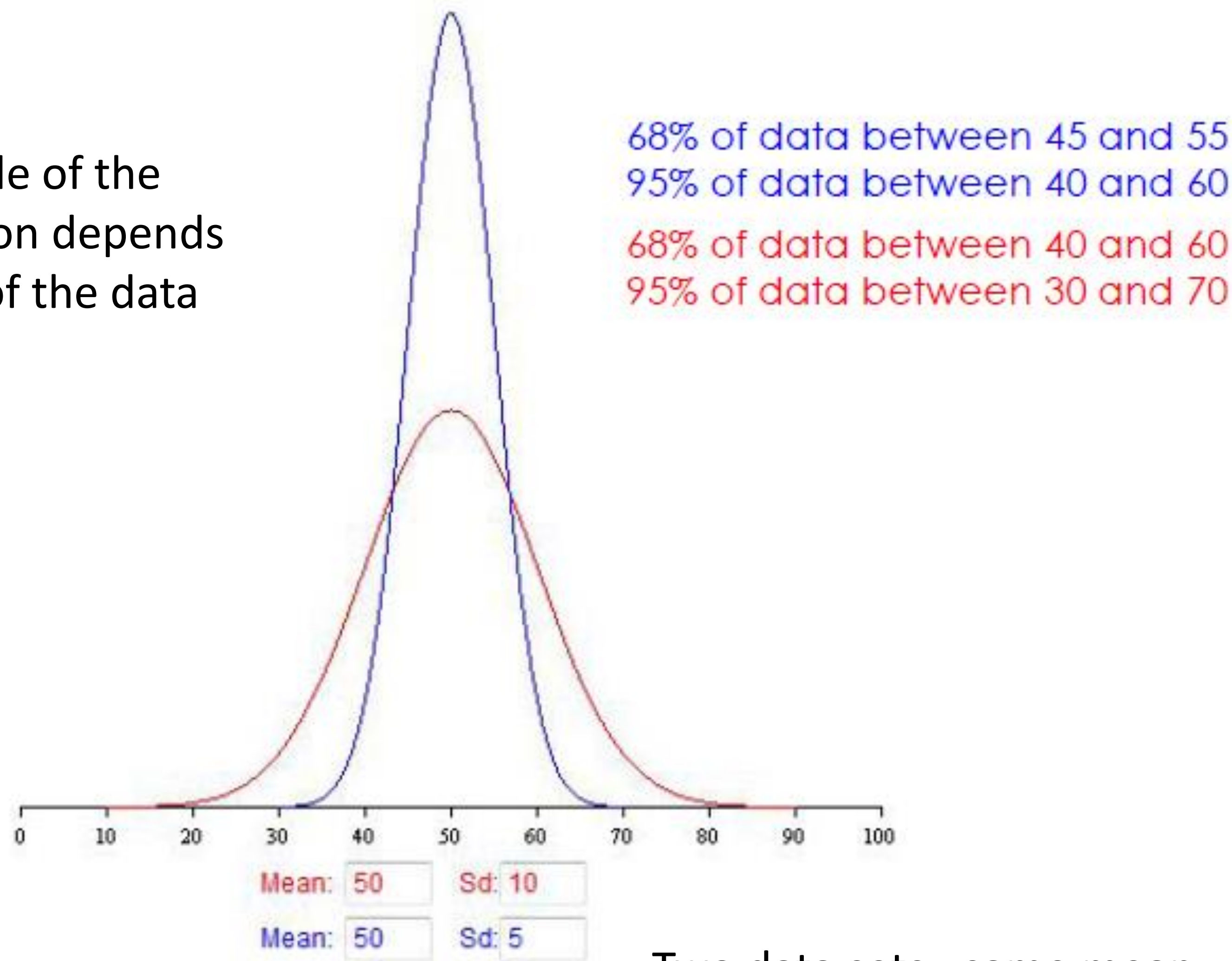
Lower standard deviation:

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

Higher standard deviation:

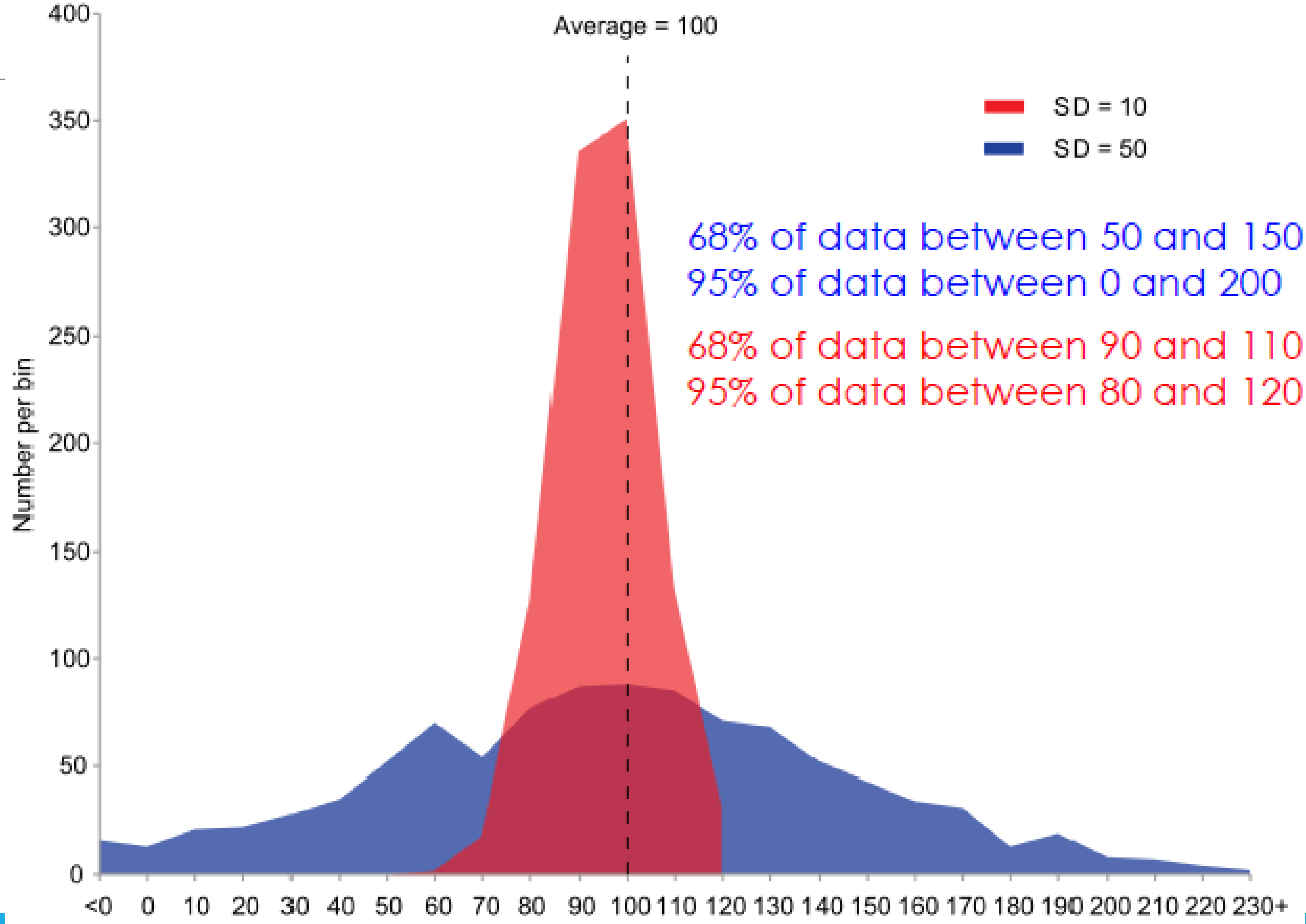
- Data is more **spread out from the mean**
- More likely factors, other than the independent variable, are influencing the dependent variable

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...



Calculating standard deviation (s)

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

1. Calculate the mean (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

Practice Activity

Sunflower Plant	**Stomata** (per examination area)	$(x_i - x)^2$
1	88	
2	93	
3	90	
4	92	
5	75	
6	78	
n=	Mean (X)=	$\sum =$

Calculations:

Standard Error ($SE_{\bar{x}}$)

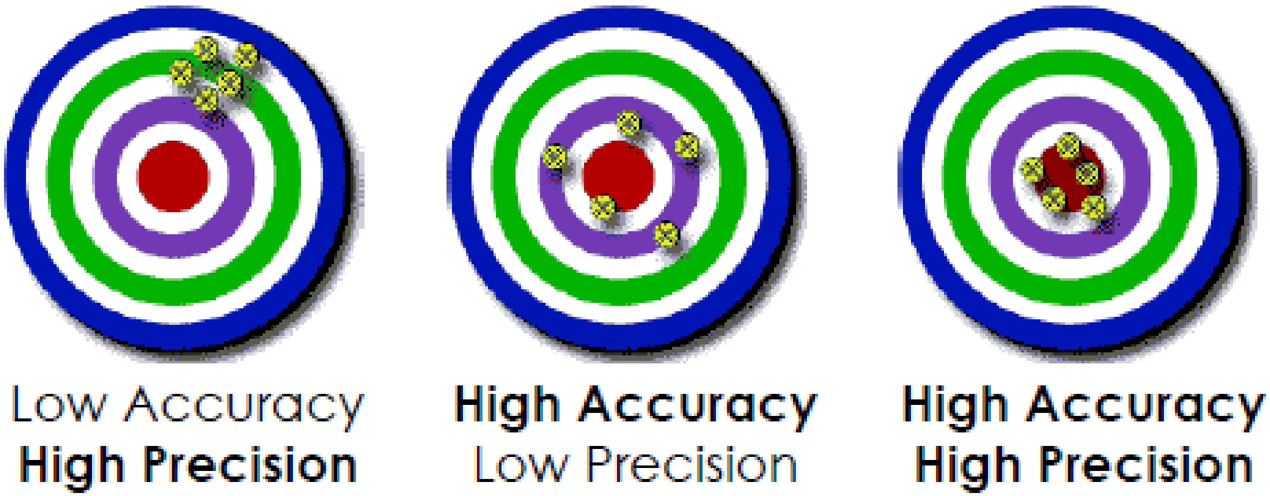
Indication of **how well the mean of a sample ($\bar{x}$) estimates the true mean of a population** (μ)

Measure of accuracy, if the true mean is known

Measure of precision, if true mean is not known

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

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



Calculating Standard Error, SE

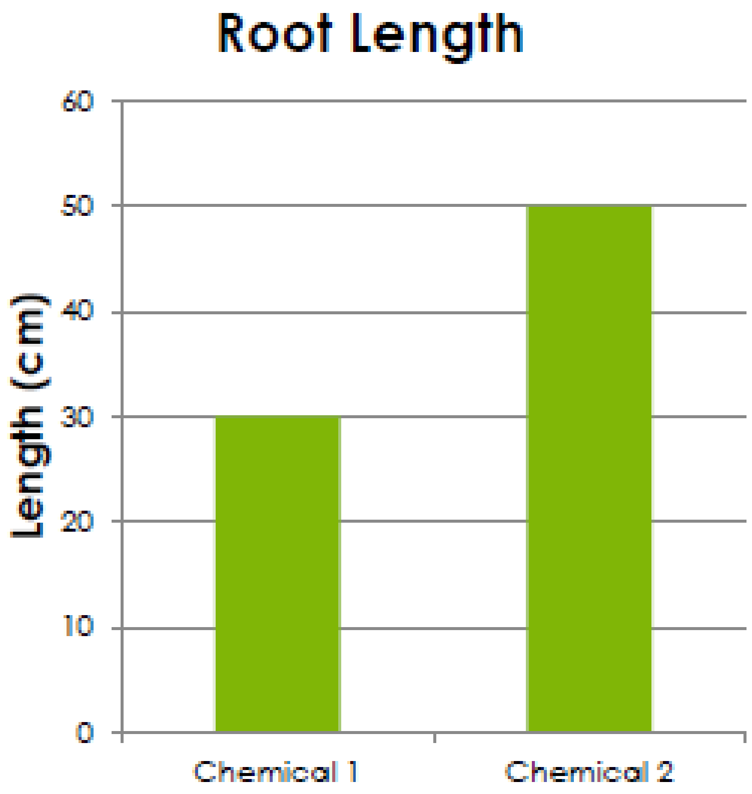
$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

1. Calculate standard deviation
2. Divide standard deviation by square root of sample size

How do we use Standard Error?

Create bar graph:

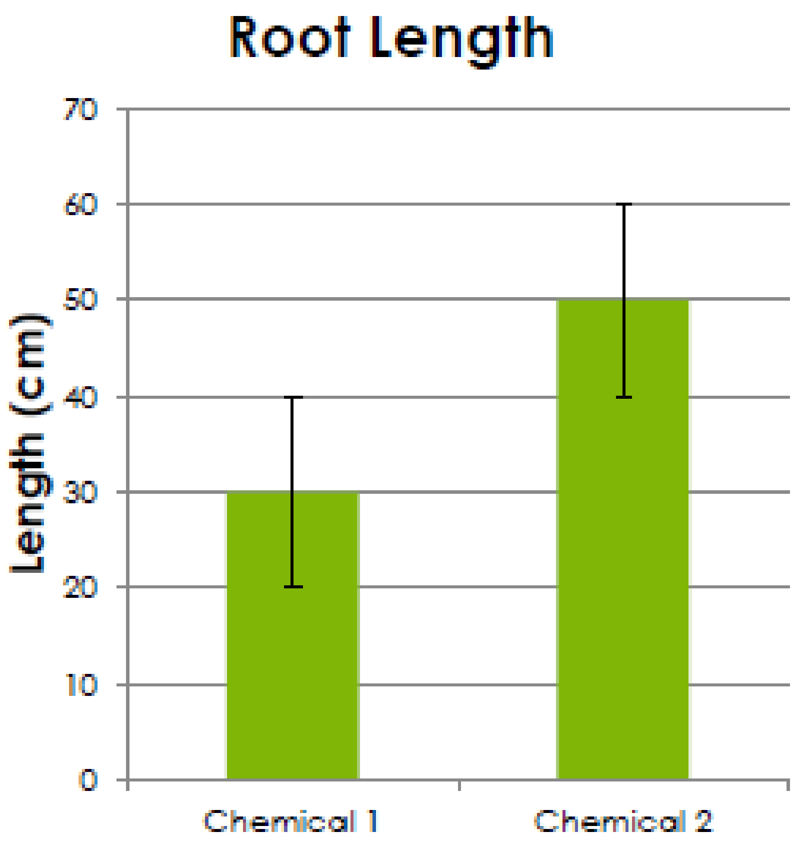
- mean on Y-axis
- sample(s) on the X-axis



Add error bars!

± SE

Indicate in figure caption that error bars represent standard error (SE)

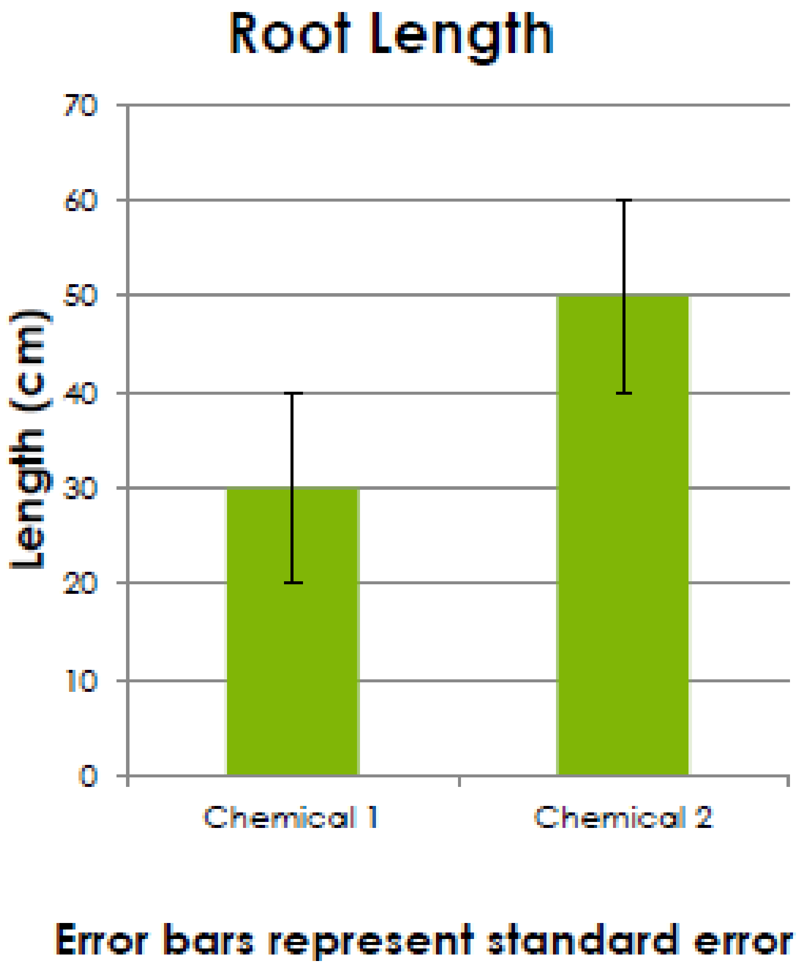


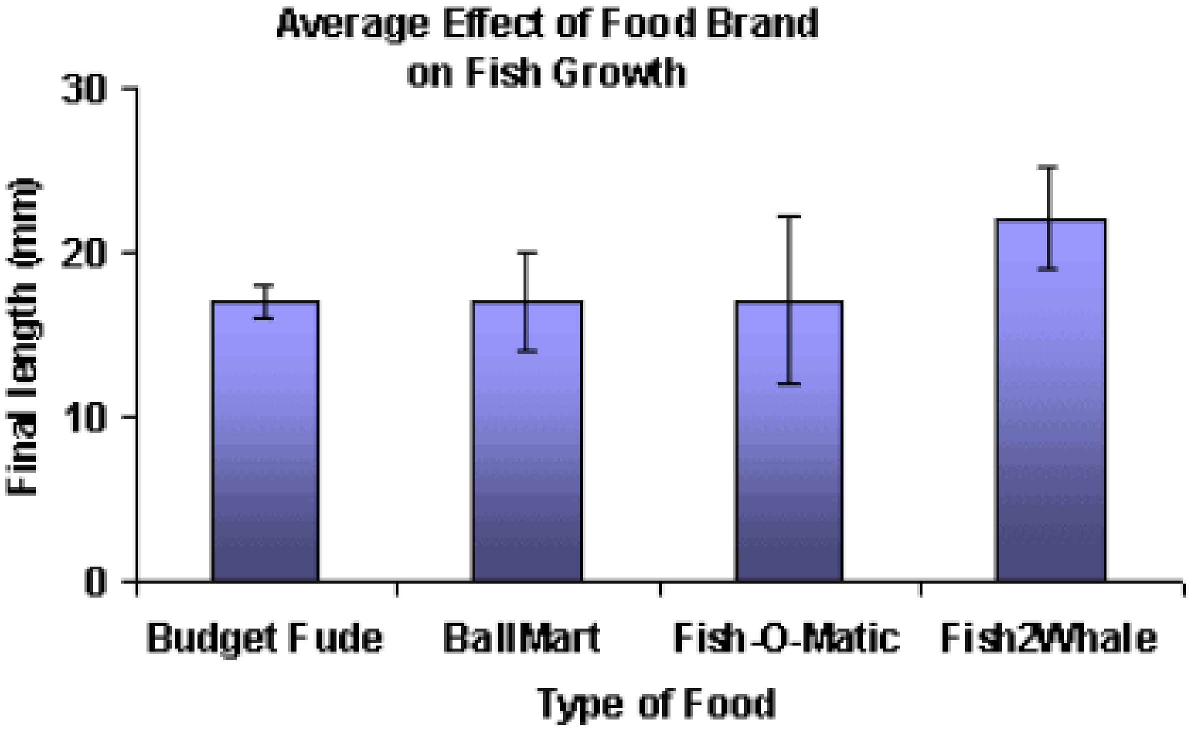
Error bars represent standard error

Analyze!

Look for overlap of error lines:

- **If they overlap**: The difference is not significant
- **If they don't overlap**: The difference <u>may</u> be significant





Average Effect of Food Brand
on Fish Growth
30
20
10
0
Final length (mm)
Budget Fude
BallMart
Fish-O-Matic
Fish2Whale
Type of Food

Standard Deviation vs. Standard Error

Standard deviation – measures how each individual data point deviates from the sample mean.

Standard error – measures how the sample mean deviates from the population mean.

95 % Confidence Interval

When conducting a study, most samples should be compared to a 95% interval of confidence ($\pm$2 SD).

When using the 95% Confidence Interval you must refer to the **t-table** and determine **degrees of freedom (df).**

$$\overline{X} \pm t \frac{s}{\sqrt{n}}$$

** This formula is used only when the population is less than 30.*

95 % Confidence Interval (t-table)

df	PROPORTION IN TWO TAILS COMBINED					
	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878