ENVIRONMENTAL SAMPLING

By C. Kohn
Agricultural Sciences
Waterford, WI

Data Collection

- **Environmental scientists often must go into the "field" to collected data.**
  - In this case, a "field" could be any outdoor environment, including forests, prairies, underwater ecosystems, and even Mars!
- **Scientists must collect as much data as possible in order to be accurate.**
  - However, it is impossible for scientists to collect information on every living organism in a habitat.
  - In order for scientists to feasibly measure an ecosystem, they must measure a portion of the living organisms in a habitat.
  - Even a sample as small as 1% could provide an accurate estimate of the total biodiversity of an ecosystem.

Representativeness

- **How the data is collected will affect the representativeness of that sample.**
  - <u>Representativeness</u> is a measure of how much the data sample that is collected actually reflects what is happening in an ecosystem.
  - The more representative the data, the more the data is similar to the actual conditions in a habitat.
  - For example, if a representative sample indicates that 12% of tree species are sugar maples, this should also be the case over the entire ecosystem being observed.

Representative

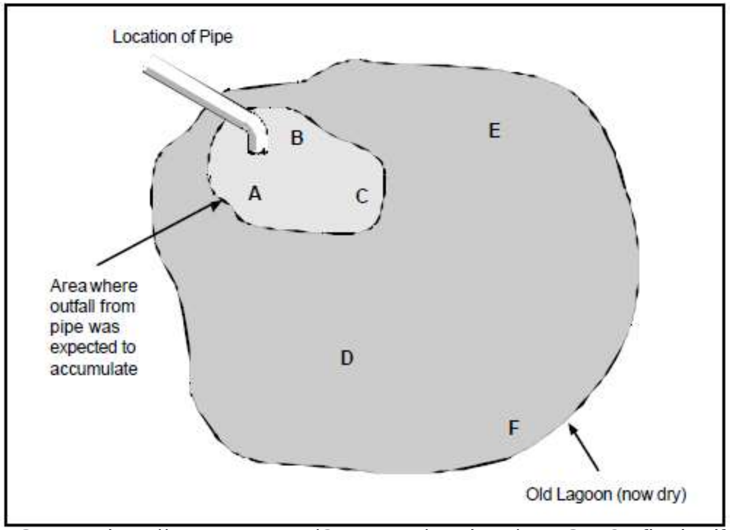
Not Representative

Total population

Example

- **Ensuring accurate representativeness can be extremely difficult in highly variable environments.**
  - For example, imagine a lagoon with a drainage pipe.
  - If all of the samples were collected in A, B, and C, then the data would indicate that the lagoon is actually much more polluted than it really is.
  - However, if data were collected in D, E, and F, then the lagoon would appear to be much less polluted than it actually is.

- **In order for data to be representative of what is actually occurring in an ecosystem, it cannot be skewed or biased.**



Source: http://www.epa.gov/QUALITY/qs-docs/g5s-Sou8rcfinal.pdf

Terms

- **The target population is the population of species, the habitat, the ecosystem, or whatever a scientist is measuring in its entirety.**
  - E.g. a wildlife biologist might collect a representative sample of fish to make a determination about all the fish in a particular lake.
  - All of the fish in that specific lake would be the target population.

- **The sampled population would be portion of the target population from which the sample was taken.**
  - E.g. if you were measuring in a forest, you might have collected samples from specific marked spots on the trail.

- **The sample is the portion of the target population that was actually measured and observed.**
  - E.g. if you wanted to determine if a field had enough nutrients in the soil, you would take a collection of samples that would on average reflect the conditions throughout the entire field.

Terms

- **A measurement protocol is the specific procedure for collecting data from the sample.**
  - This includes procedures for collecting the physical sample, handling and preparing the sample, type of analysis of the data collected, etc.

- **Measurement protocols can be either probability-based or judgmental.**
  - Probability-based measurements depend on random selection of data.
    - *E.g. if you randomly caught fish and averaged their size to make decisions about a target population of those fish.*

  - Judgmental measurements are conclusions based mostly on the observations of an expert in a particular area.
    - *E.g. if a fish biologist used his or her own knowledge about a specific target population of fish to make decisions about those fish.*

Selection of a measurement protocol

- **Accurate selection of an environmental measurement protocol is crucial in order to collect data that is accurate and representative of what is actually occurring in an ecosystem.**
  - Use of the wrong protocol could result in a skewed and inaccurate sample of data that would cause inappropriate decisions to be made when managing an ecosystem.

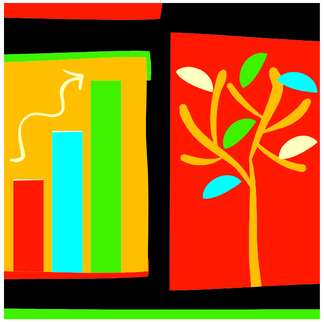
- **The selection of a sampling technique in the environment depends on the type of data needed, the size of the sampling site, the number of people available to collect data.**
  - When using probability-based measurements, the collected data can provide specific information that can be assessed for accuracy using statistical methods.
    - *However, this is a slow and arduous process.*
  - Use of judgmental techniques provides quick results but requires access to a proven expert with extensive training.
    - *This data is also subjective and can be less reliable.*
    - *This option may not be available for many kinds of situations, especially those involving complex relationships, large amounts of data, or large areas of land.*

Kinds of Probability-based Measurements

- **There are many kinds of Probability-based measurements. Some examples include:**

- **Simple Random Sampling: samples are chosen at random.**
  - Each data point from a sampled population has an equal chance of being selected.

- **Simple random sampling has three key benefits. It…**
  1. Provides unbiased averages.
  2. Is easy to understand and implement.
  3. Statistical analysis is straightforward and easy.



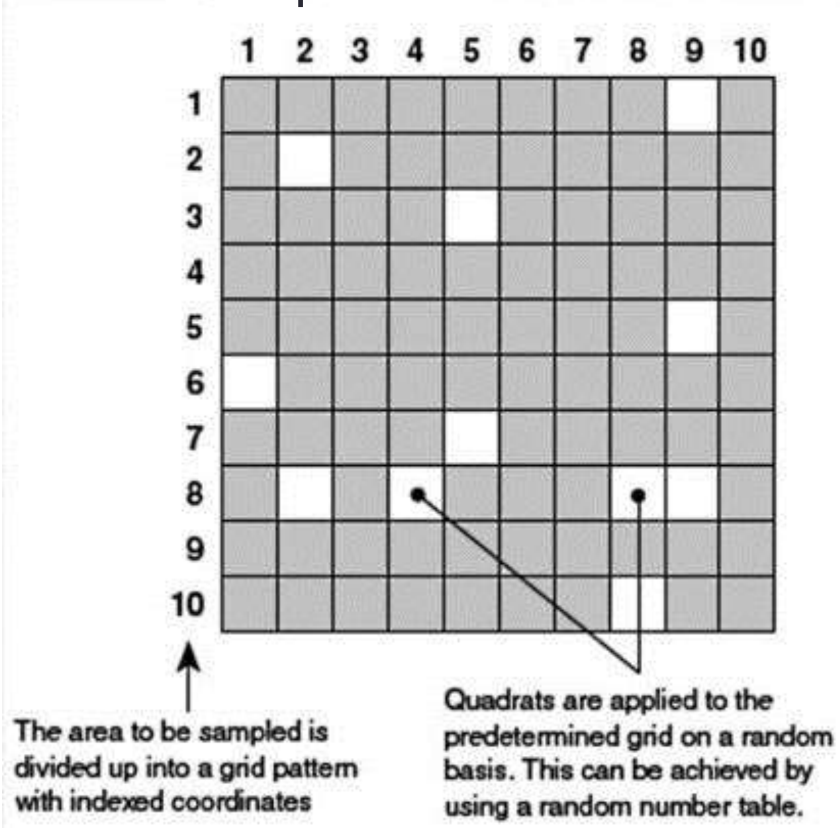
Simple Random Sampling

- **So why is simple random sampling *not* always used? Sometimes it is not feasible.**
  - For example, some parts of a habitat may be isolated or inaccessible to researchers.
  - Sometimes a random sample may be more costly or time-consuming than other methods.

Source: www.wildlifetrusts.org

Quadrat Sampling

- **One of the most commonly-used examples of simple random sampling is called <u>Quadrat Sampling</u>.**
  - In quadrat sampling, the entire area to be sampled is divided into equally-sized parcels.
    - *For example, you could divide a one acre prairie into squares sized 1x1 meters.*
    - *Each 1x1 square would be a quadrat.*
  - Each quadrat is assigned a number and the quadrats to be sampled are chosen at random.

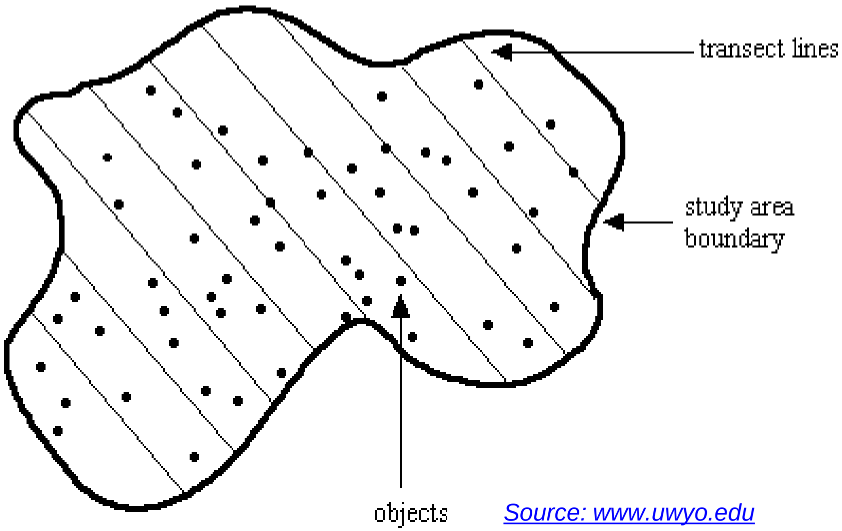


Source: rgsbio09.wikispaces.com

Systematic & Grid Sampling

- **Another type of probability-based measurement is <u>systematic and grid sampling</u>.**
  - In this kind, samples are collected at regularly spaced intervals.

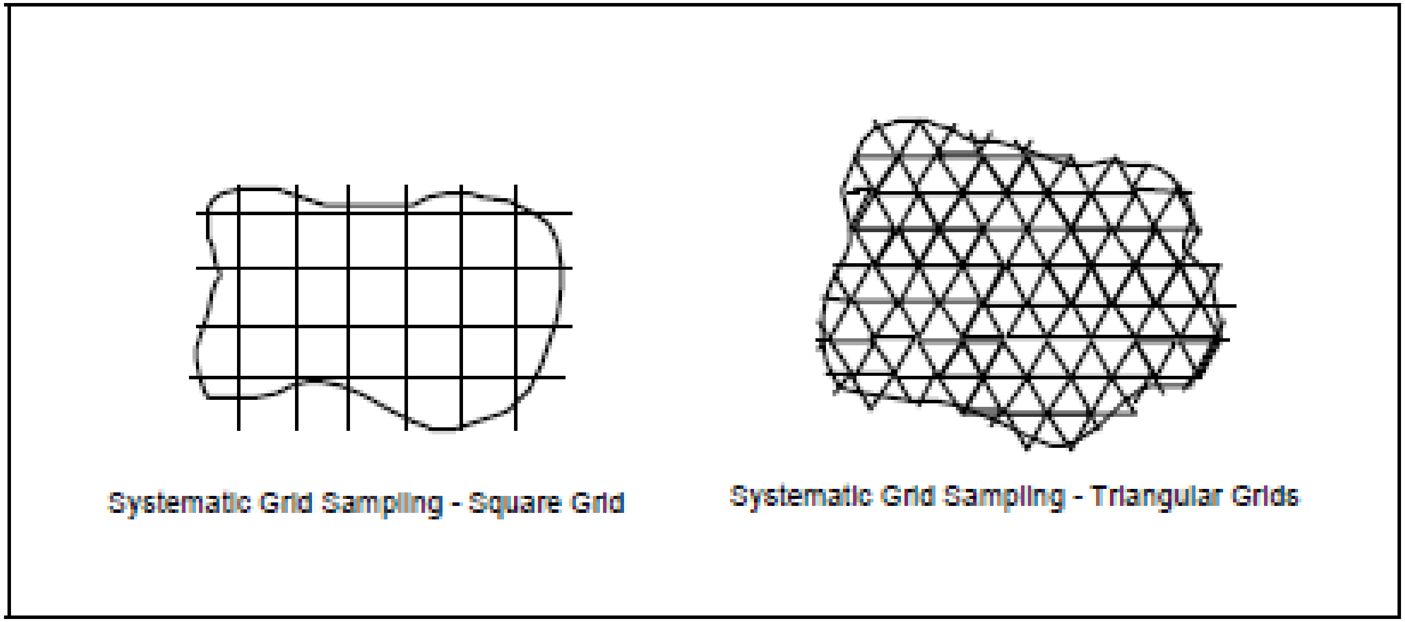
- **Transect Sampling is an example of this type.**
  - In transect sampling, parallel lines are established in a habitat.
  - Data is collected at consistent intervals along each line (e.g. at 5, 10, and 15 meters into the habitat).



Source: www.uwyo.edu

Systematic & Grid Sampling

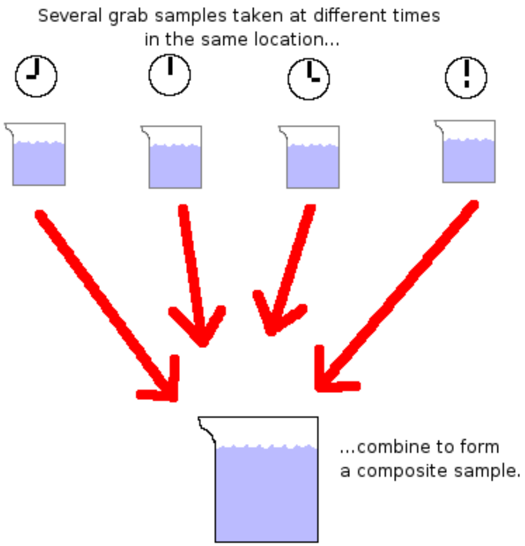
- **In systematic and grid sampling, a habitat may also be broken up by a grid of regularly intersecting lines.**
  - Collection of data at each intersecting line can provide reliable averages and/or provide comparisons at different points in a habitat.



Source: http://www.epa.gov/QUALITY/qs-docs/g5s-Sou8rcfinal.pdf

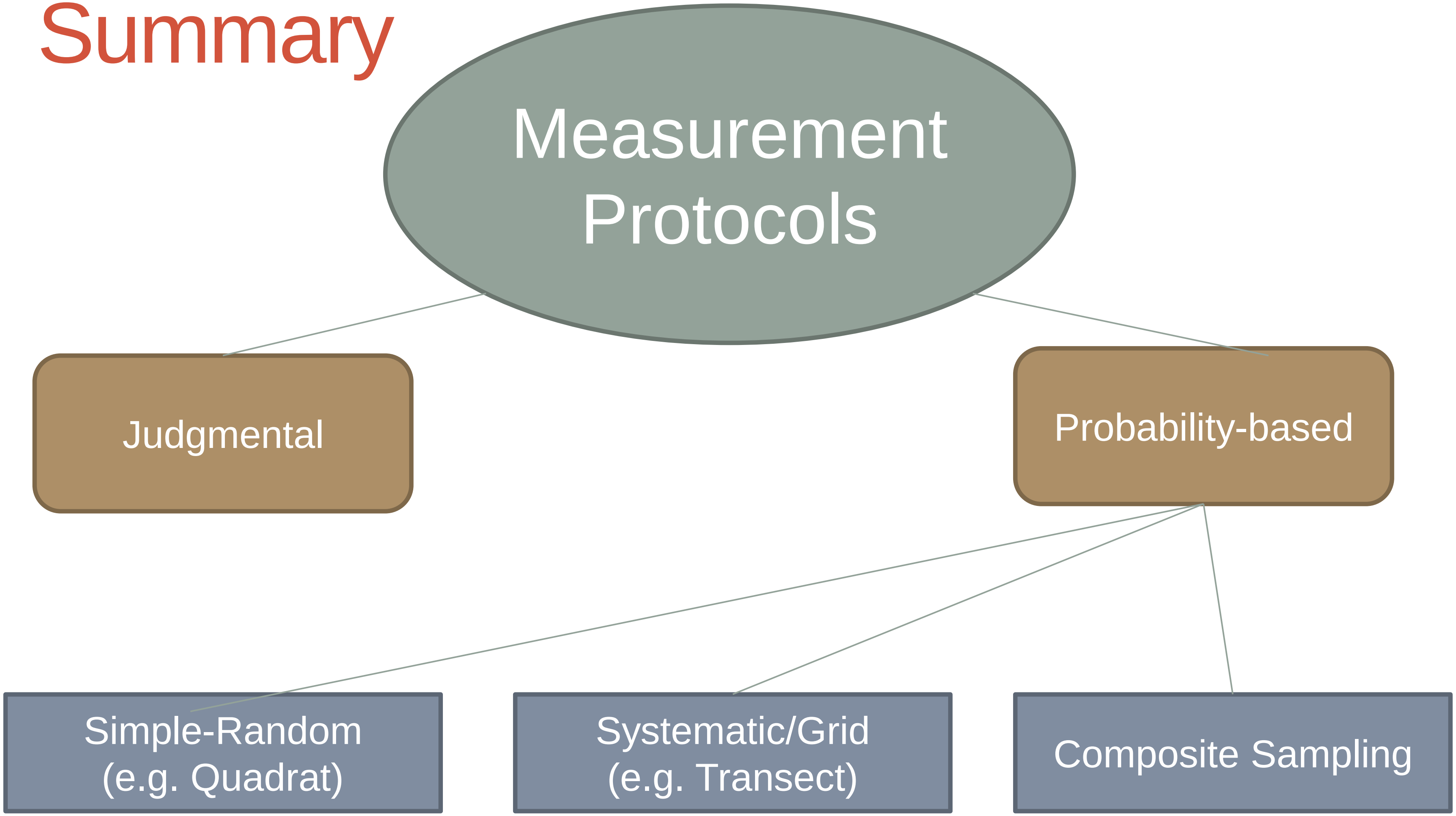
Composite Sampling

- **A final example of probability-based measuring is the <u>composite sampling technique</u>.**
  - In this technique, samples are combined and mixed to provide a representative sample of an area.
  - For example, soil analysis may be conducted in this way; the samples are all mixed together and the combined samples are then tested.
- **This method can be very time-saving and cost-effective.**
  - However, it does not work in all cases.
  - E.g. you cannot do this with non-mixable measurements, like the kind of measurements you might collect for animals.



Source: water.me.vccs.edu

Summary


Measurement Protocols
Judgmental
Probability-based
Simple-Random (e.g. Quadrat)
Systematic/Grid (e.g. Transect)
Composite Sampling

How do we know we know?

- **A major concern in any science is proving that what we have observed would occur again if we repeated the experiment.**
  - Random things can affect our experiments.
  - Your samples might be affected by little things that change or skew your results.
  - The trends you find in your experiment may not occur in a different experiment done in the same way.

- **We must always be prepared to answer the Scientist's Questions:**
  -How do I know I am not wrong?
  -How do I know that this will always occur every time I do this experiment?

opticalres.com

Target Shooting & Statistics

- **Data in research is sort of like target practice.**
  - When you are shooting at a target, you want all of your shots to be close together.
    - *The closer your shots are to each other, the better.*
  - You also need to take a lot of shots in order to be accurate.
    - *The more times you shoot at a target, the more accurate of a shooter you are.*
- **Statistics are the same: the more data we have, and the more similar each number is to each other, the better.**

Source: topendsports.com

Science & Statistics

- **In science, we can use statistical equations to determine whether or not we can be confident in our results.**
  - In other words, the use of statistics can tell us whether our experimental results are reliable.

- **If we are likely to see similar results every single time, this means that our results are reliable.**
  - On the other hand, if we get very different results each time we do an experiment, our data is less reliable.

- **The more variable our data, the less reliable it is.**

Less reliable

- **The less our data varies, the more reliable it is.**

More reliable

labellecuisine.com

Factors that Affect Data Reliability

- **Things that affect the *reliability* of our data include:**
  - How **similar** our data is:
    - *The more similar the data, the more reliable our average will be.*
      - E.g. if all of our students are between 5’10” and 6’1”, we would have more reliable data than if the range of the data was greater (such as if the range was between 4’5” and 7’1”)
  - The **amount** of data we have:
    - *The more data we have, the more reliable our average will be.*
      - E.g. if you flip a coin 3 times, you might get 2 heads, 1 tail.
      - If you flip a coin 10 times, you might get 6 heads and 4 tails.
      - If you flip a coin 100 times, you might get 49 heads, 51 tails
        - Each time we get closer to the “real” average of 50/50

goldenstateofmind.com

Source: istockphoto.com

Standard Deviation

- **<u>Standard Deviation</u> is a measure of variability.**
  - We want our data to be very similar;
    we don't want it to be spread out.
  - Data is like butter – we want it in a tight form, like a stick. We don't want it to melt and spread everywhere.

- **Standard Deviation can be used to find the Margin of Error (or our "range of accuracy").**
  - <u>Margin of Error is usually equal to 2x the Standard Deviation on either side of the mean (*average*).</u>
    - *As you will soon see, we have a better way to calculate the margin of error.*
  - The margin of error shows us all of the possible results for our research (if we repeated it) with a 95% accuracy.

Standard Error

- **<u>Standard Error </u>is a measurement of reliability of a data sample; it involves both the size of your data sample and the variance of your data.**
  - Standard Error is calculated by dividing your Standard Deviation by the square root of your sample size.

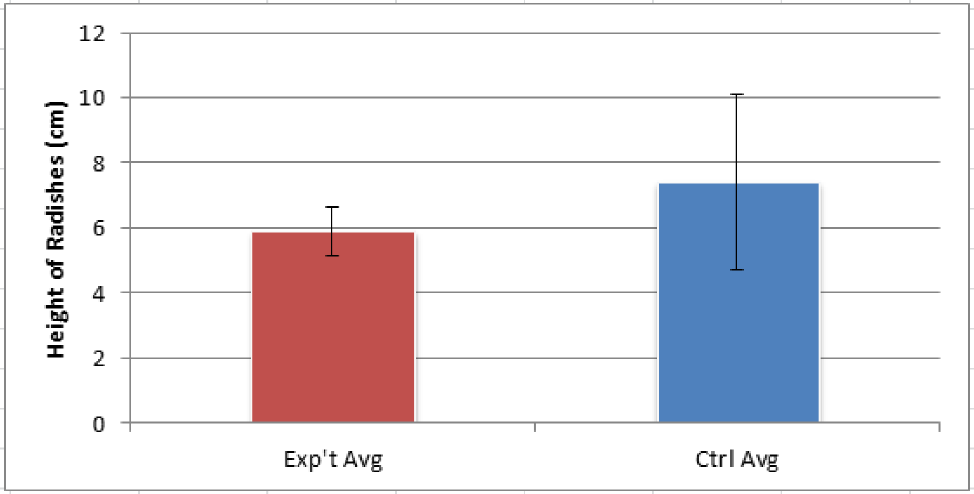
- **<u>Standard Error is a measure of the reliability of your data.</u>**
  - *It uses both the **size** of the data sample and the **variability** of the data.*

Error Bars

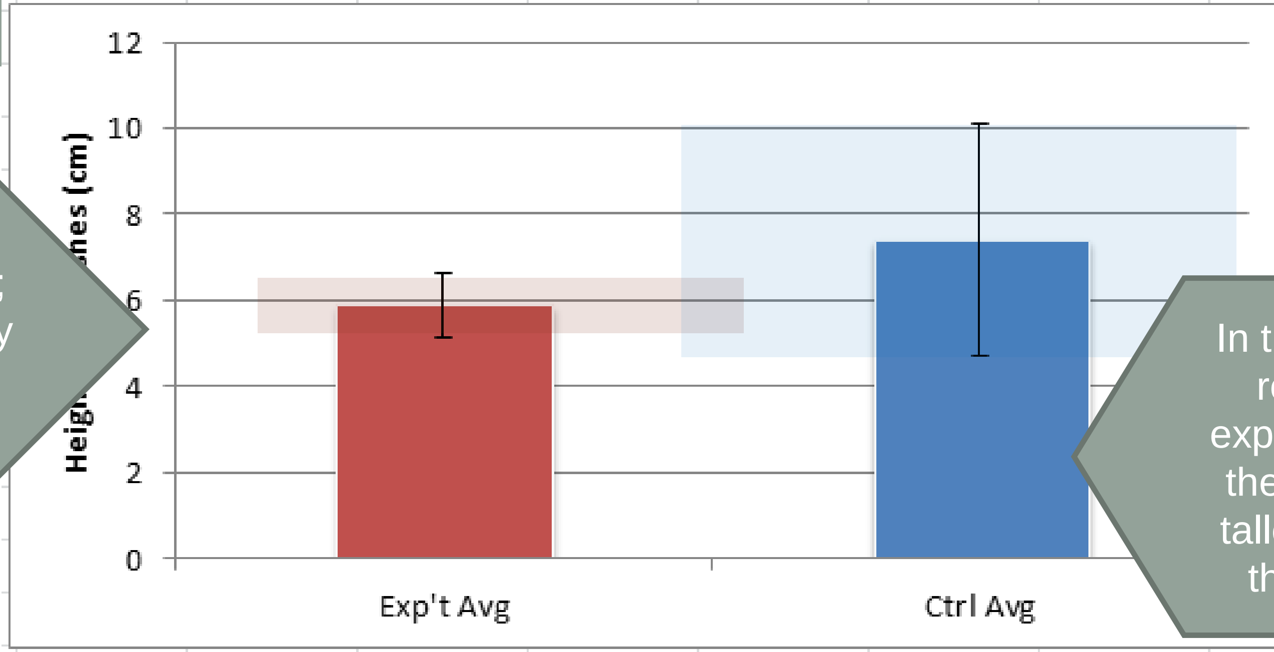
- **In this example, the control has an average height (or *mean* height) that is over a full centimeter taller than the experimental average.**
  - It looks as if the blue average is noticeably greater than the red average.

- **However, the Error Bars (+/- 2 Standard Errors) overlap.**
  - If your error bars overlap, this means that there is no statistically significant difference between the control and the experimental average.

  - You must treat them as if they are the same.

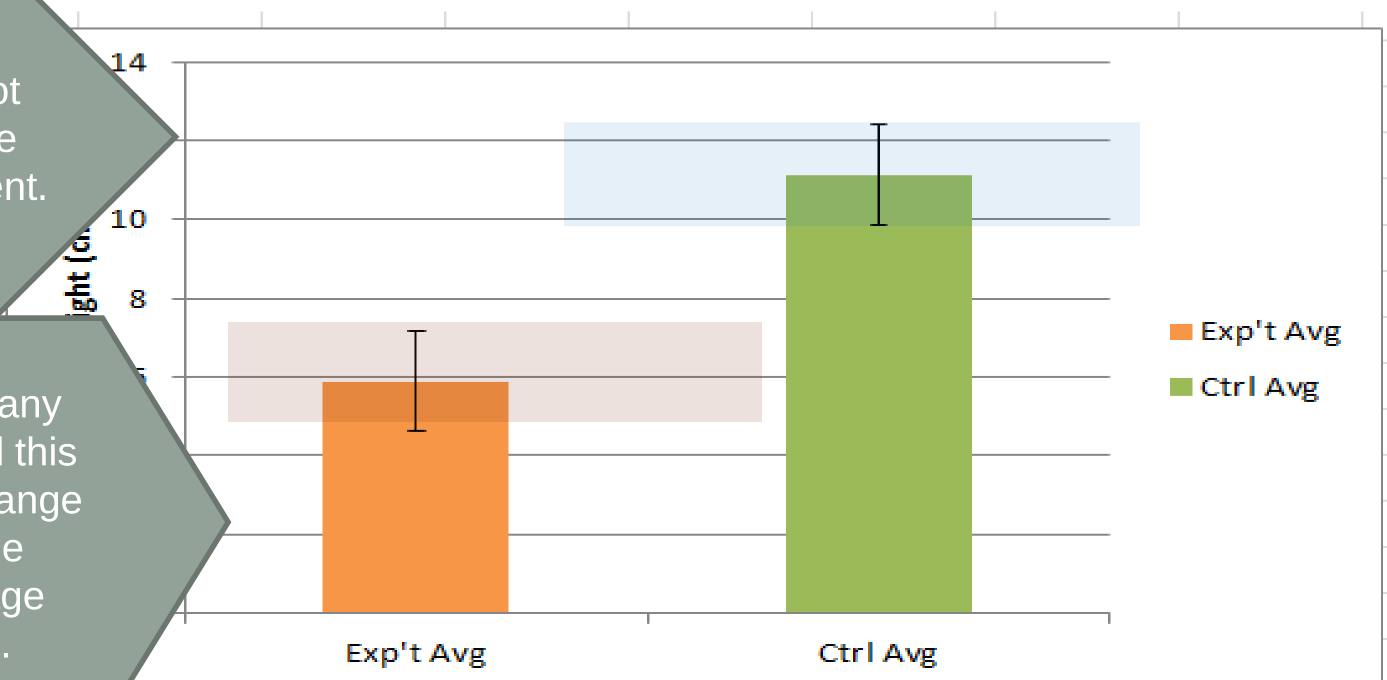


Error Bars overlap; they are statistically the same.



In this case, if we repeated the experiment again, the red *could* be taller on average than the blue.

Error Bars do not overlap; they are statistically different.



No matter how many times we repeated this experiment, the orange would always be shorter on average than the green.