What is *statistics*?

the study of the collection, organization, analysis, and interpretation of data.

What is *data*?

*qualitative* or *quantitative* attributes of a variable or set of variables.

(observations (such as measurements, survey responses) that have been collected)

Statistics deals with all aspects of this, including the planning of data collection in terms of the design of *survey* and *experiments*.

The theoretical basis for collection, organization, analysis, and interpretation of data is *Probability*.

To analysis and interpret data, using probability, is called *Inference* (statistical).

Examples:

Ecology: Deers on campus, how many? How to guess?

Health: Is a new drug better? How to compare?

Biology: How to find which part of DNA determines the color of eyes?

Everyday life: How do you get to know a people?
Basic concepts:

Population:
The complete collection of all elements to be studied

Census:
Collection of data from every member of population

Sample:
Subset of census, collect only part of a population

Parameter:
A measurement describing some characteristic of a population

Statistic:
A measurement describing some characteristic of a sample

Type of Data:

Quantitative (numbers)
Continuous vs. discrete

Qualitative (categorical, name of categories)
Nominal vs. ordinal
How to collect data

Observational study

We observe and measure specific characteristics, but we do not attempt to modify the subjects being studied.

Cross-sectional study (deer # on campus)
Retrospective study (Canada goose migration habit)
Prospective study (effect of trauma)

Experiment

We apply some treatment and then proceed to observe its effects on the subjects.

Avoid Confounding

Confusion of variable effects.

Only sickest patient get the drug will make the drug look less effective.

Swiss physician H. C. Lombard: students has the shortest length of life.
**Sampling**

Random sample (every element equal chance) vs. Simple random sample (every size n sample same chance)

**Sampling ERROR**

If sample data are not collected in an appropriate way, the data may be so completely useless that no amount of statistical analysis can salvage them.

Voluntary response sample

**Other Sampling methods**

Convenience, systematic, stratified, cluster sampling