Online Review Course of
Undergraduate Probability and Statistics

Review Lecture 1
What is Statistics?

Chris A. Mack
Adjunct Associate Professor

Course Website: www.lithoguru.com/scientist/statistics/review.html

Two basic tasks in statistics

- Descriptive statistics
  - Describe or summarize a large set of data with just a few numbers
- Statistical inference
  - Use measurements on a sample (a subset of the population) to infer some property of the population
  - Inference is an extrapolation (or generalization) and we must always assess how confident we are in the extrapolated result
- An important mathematical tool: probability

Some Definitions

- Statistics
  - A branch of mathematics that provides procedures for describing and reasoning from data
  - Less formal: a set of tools designed to help one move from data to decisions
- Statistics is:
  - A method of reasoning
  - A collection of mathematical procedures
  - A philosophy of uncertainty and knowledge
- Statistics is required because of a fundamental characteristic in the universe: variation

Two Approaches to Statistics

- Frequentist (classical) approach
  - Based on the frequency approach to probability (law of large numbers)
- Bayesian approach
  - Based on Bayes' theorem
- The difference is mostly based on how one treats unknown parameters

Population and Sample

- Population
  - The entire collection of items of interest
  - For many measurements in science, the population is effectively infinite
- Sample
  - A subset of the population selected for study
  - Goal: the sample should be "representative" of the population

What is Data?

- Data = the results of a measurement
  - Definition of the thing being measured (measurement model)
  - Measurement value (number + unit)
  - Experimental context (method + environment)
  - Uncertainty estimate (including context uncertainty)
- Don't take any of these aspects of data for granted!

Kinds of Data

- Data set types
  - Univariate: only one variable is considered
  - Bivariate: two related variables are considered
  - Multivariate: three or more variables considered
- Observational data
  - The influence of the observer is considered negligible (environment is uncontrolled)
- Experimental data
  - Researcher manipulates (controls) the environment to make interpretation of the data easier
  - Confounding variable: a variable that is not manipulated and not controlled but impacts the measured values
- Sample planning and design of experiments (DOE)
  - Maximize the quality of results for minimum cost

Kinds of Data (2)

- Categorical data
  - Define categories, then count how many fall into which category
  - All categories are arbitrary; some categories are useful
- Quantitative data
  - The output of a measurement, the quantity of a thing
  - Generally a continuous variable (temperature, speed, time, etc.)
- Type of data is classified by its measurement scale

Measurement Scales

- Measurement
  - The assignment of numbers to objects or events according to rules
- Measurement scales
  - Nominal: count how many fall into which category
  - Ordinal: categories have a natural order (rank)
  - Interval: differences can be compared numerically, but no natural zero point (origin)
  - Ratio: both differences and ratios can be compared; there is defined and natural origin
- The statistics that can apply depend on the measurement scale

Measurement Scale Examples

- Nominal
  - Number of 2014 UT graduates by degree, by gender
- Ordinal
  - Hardness of minerals, grade or quality of leather
  - Pain rating on a scale of 1-10 (maybe)
- Interval
  - Non-absolute temperature (Fahrenheit, Centigrade)
  - Intelligence measures (at least in theory)
- Ratio
  - Most measurements used in science (time, absolute temperature, mass, pressure, velocity, viscosity, density, etc.)

see S. S. Stevens, "On the Theory of Scales of Measurement", *Science*, **103**(2684), p. 677 (June 7, 1946).

Review #1: What have we learned?

- Define "statistics"
- What are the two basic tasks in statistics?
- Define "population" and "sample"
- What are the four parts that make up data?
- Explain the difference between observational and experimental data
- Name and describe the four measurement scales

