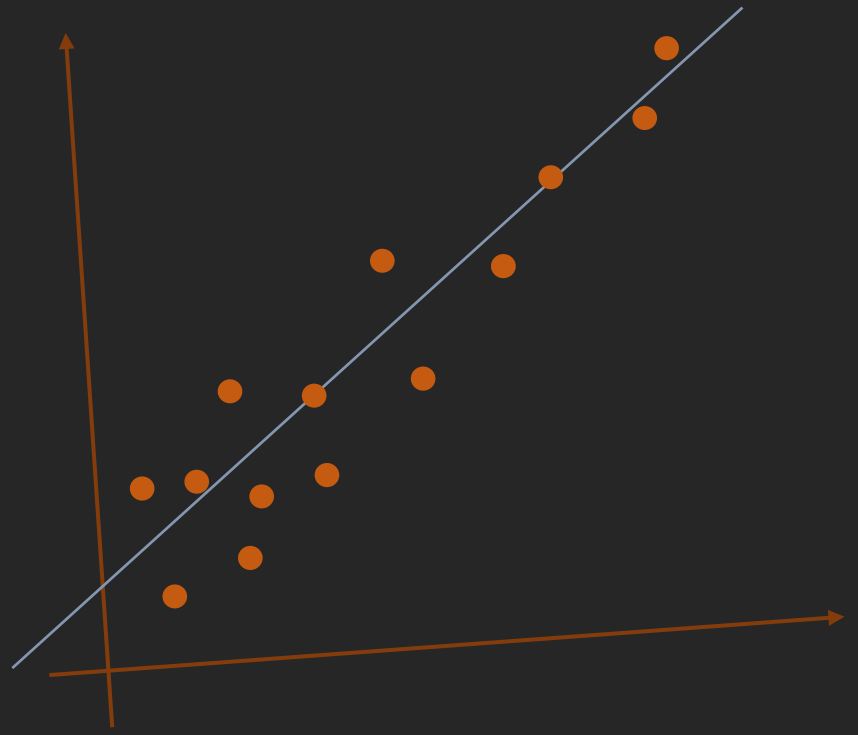
Applied Regression Analysis

STAT 4043/ STAT 5543



Review of statistical inference

Pratyaydipta Rudra

Quick review of statistical inference

Chapter 1 of the textbook

Quiz 2 will be based on pre-requisite topics

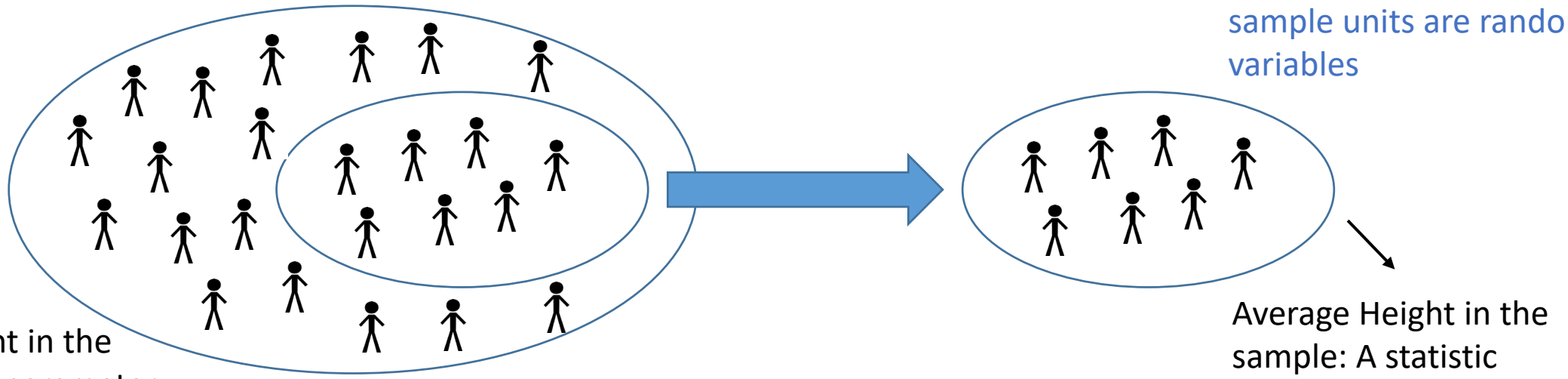
Topics for you to review

- Basic concepts of statistics: data and variables; different types of variables (continuous, discrete, categorical - nominal and ordinal); designed experiment vs observation studies; basic visualization such as bar plots, histograms and boxplots; measures of location such as mean, median, mode; measures of dispersion such as range, variance and standard deviation; population and sample; statistical inference
- Summation notation

- Basic concepts of probability and the common probability distributions (mainly normal, t, F and binomial). Concept of sampling distribution.
- Expectation notation: E(X)
- Estimation - concepts of precision, standard error, confidence intervals
- Hypothesis testing - concepts of null and alternative hypothesis, type-I and type-II errors, statistical power, p-value, level of significance, how to accept or reject the null hypothesis and how to write the results from a hypothesis testing problem.

Population and sample, parameter and statistic

- **Population**: A data set representing the entire entity of interest.
  - **Parameter**: A numerical summary of a population.
- **Sample**: A data set consisting of a portion of the population.
  - **Statistic**: A numerical summary of a sample.



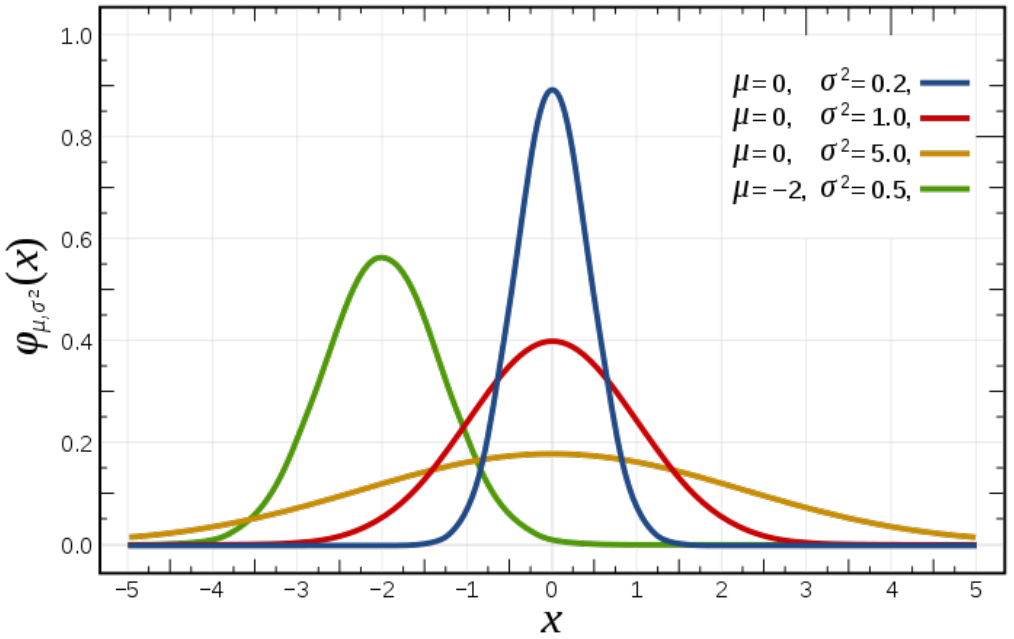
Random variables have probability distributions

- We usually write in the following way:
  - "Suppose $X_1, X_2, \ldots, X_n$ is a random sample…"
- For a random sample from infinite population, $X_1, X_2, \ldots, X_n$ are iid random variables.
- The probability distribution of $X_1, X_2, \ldots, X_n$ will depend on which population they are drawn from.

E( ) and Var( ) are usually used to denote the expectation (i.e. mean) and variance of a random variable.

Normal distribution

- The density curve of normal distribution is a bell-shaped curve. The range of the random variable is from $-\infty$ to ∞.
- The distribution has 2 parameters, the mean μ and the variance σ^2.



A normal distribution is called standard normal distribution when $\mu = 0$ and $\sigma^2 = 1$.

The corresponding random variable is often denoted by Z

Source: wikipedia

Some useful distributions

- Binomial distribution.
- Normal distribution.
- χ^2 distribution.
- t distribution.
- F distribution.

- How do these distributions look like?
  - Symmetric/skewed?
  - What is the range of values?
  - What are the parameters characterizing the distribution?
- How are these distributions useful in statistical inference?

Estimation and hypothesis testing

- Point estimation: Make a guess about a population parameter.
- Interval estimation: Propose an interval that will contain a population parameter with a certain probability.
- Hypothesis testing:
  - Propose a null hypothesis (H_0), i.e. a statement about a population parameter.
  - Test if the hypothesis seems to be true based on the data.

Quickly visit some useful terms

- Standard error.
- Level of significance.
- P-value.