ECO 372: Introduction to Econometrics

Spring 2025

Lecture 2: Review of Probability and Statistics

Sakib Bin Amin, Ph.D.

Associate Professor in Economics

Outline

- Introduction
- What is Probability
- Some Useful Concepts
- Probability Distribution
- Conditional Probability
- Joint, Marginal and Conditional Distributions
- Moments
- Correlation and Covariance
- Normal and Standard Normal Distributions
- Some Test Statistics
- Population Vs. Sample
- CLT and LLM

What is Probability?

- Probability deals with the chance that an event will occur:
  - *What is the probability that a candidate will be elected?*
  - *What is the probability that a medical treatment will be effective?*
  - *What is the probability that it will rain tomorrow?*
- Standard Definition: The probability of an event A, Pr (A), is the likelihood that it will occur.
  - *Probability of an event A happening is represented as Pr (A).*
  - *Probabilities are always between 0 and 1, inclusive. They may be represented as percentage. Example, Pr (A) = 0.2 or 20%.*

Some Useful Concepts

- **Experiment:** any action or process whose outcome is subject to uncertainty. Flipping one fair coin is an example of an experiment.
- **Outcome:** is the result of an experiment.
- **Population:** The group or collection of all possible entities of interest (very large)
- **Sample Space:** collection of all possible outcomes (or elements) of the experiment (set S).
- **Event:** collection of elements (subset A, B, etc.) contained with in the sample space (S).
- **Random Variable:** numerical summary of a random outcome. A variable whose value is unknown or a function that assigns values to each of an experiment's outcomes.

-Two types - discrete random variables (DRV) and continuous random variables (CRV). DRV take on a continuous number of distinct values (flipping coins). CRV take any value within a specified range of interval (height)

- **Independent Random Variable:** Two random variables are independent if knowing the value of one gives no information of the other.

Example: Toss a Fair Coin 3 times

- The sample space is {TTT, TTH, THT, HTT, THH, HTH, HHT, HHH}
- Let X = the number of heads you get when you toss three coins
- Then X can take the values: x = 0, 1, 2, 3
- X is the RV; x is the realization of the RV.
- Because you can count the possible outcomes of X and the outcomes are random, X , is a discrete random variable

$$\Pr(A) = \frac{\#\ \text{of outcomes of an event}}{\#\ \text{of outcomes of an event}}$$

Note: $Pr(S) = 1$

Let E be the event of getting 2 heads, i.e. X = 2.

So, Pr (E) = Pr (X = 2) = 3/8 = 0.375

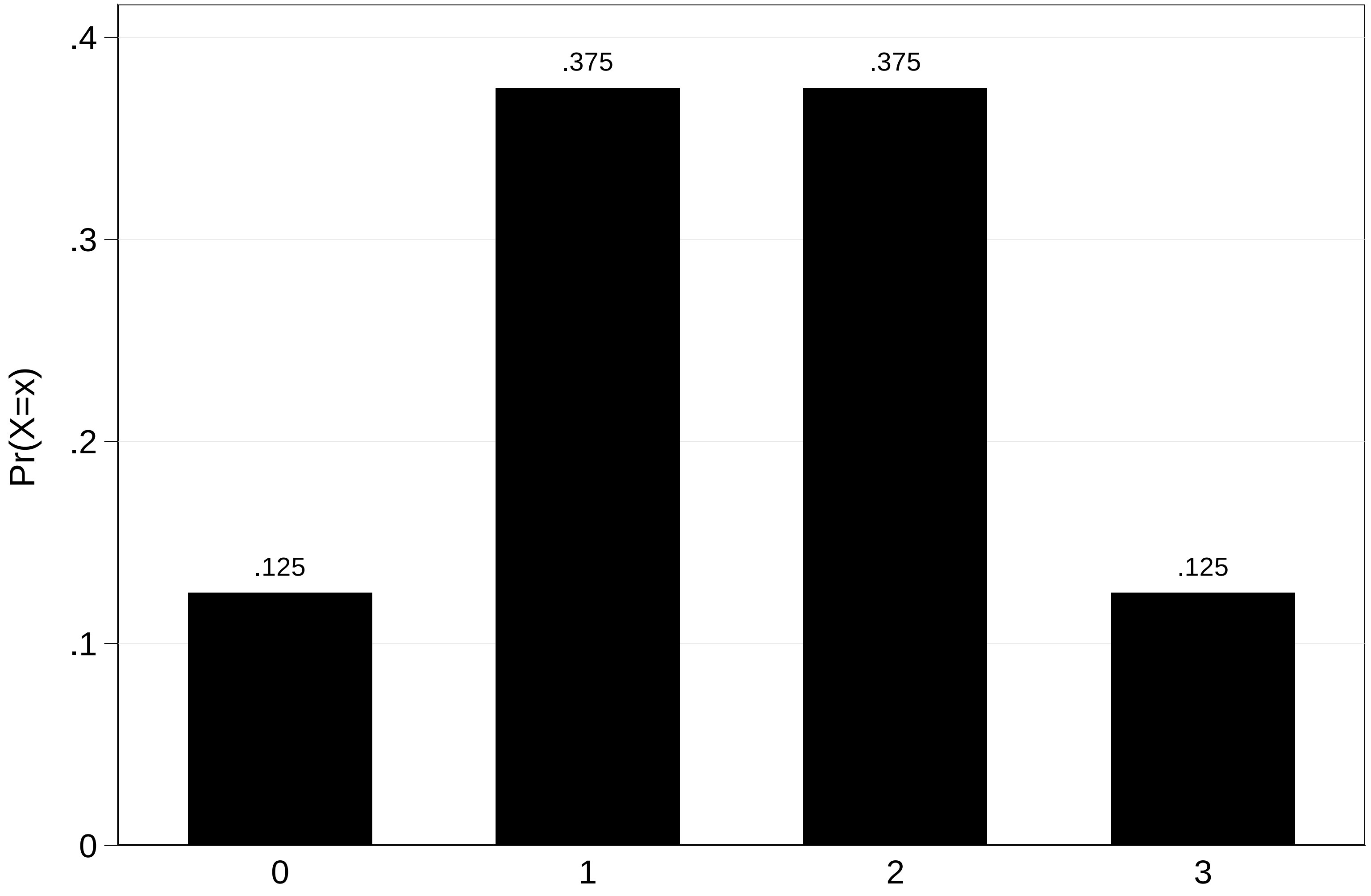
Similarly we can find the probabilities for all values of X, i.e.

Pr (0), Pr (1), and Pr (3).

Probability Distribution

- A discrete probability distribution is the list of all possible values of a variable and the probability that each value will occur. It has two characteristics:
  1. Each probability is between 0 and 1, inclusive
  2. The sum of the probabilities equals 1.
- For continuous distributions: it is called probability density function (PDF or p.d.f).
- Note: For continuous distributions $\Pr(A = c) = 0$, where c is a constant.
- A cumulative distribution function (CDF or c.d.f) is the probability that the random variable is less than or equal to a particular value.
- Written as: $\Pr(X \leq c)$, where c is a constant.

Probability Distribution



Probability Distribution

X	$Pr(X = x)$	Cumulative Probability, $Pr(X \leq x)$
0	0.125	0.125
1	0.375	0.5
2	0.375	0.875
3	0.125	1

Conditional Probability

- Conditional probability that event A happens given B , written "A|B", is the probability that both events A and B happens divided by the probability that event B happens.

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$$

- The sample space, S = {1, 2, 3, 4, 5, 6}.
- Let E be the event of rolling an number that is at least a 5
  - Pr(E)=2/6
- Let F be the event of rolling a odd numbers, F = {1, 3, 5}
  - Pr(E)=3/6
- Probability of both E and F happening: Pr (E ∩ F)=1/6
- Conditional probability: (1/6)*(6/3)=1/3

Joint, Marginal, and Conditional Distributions

- Joint probability distribution is the probability that the random variables simultaneously take on certain values, Pr (X = x, Y = y).
- Marginal probability distribution is the probability distribution of one variable.

$$\Pr(Y = y) = \sum_{i=1}^{I} Pr\,(X = x_i, Y = y_i)$$

- The distribution of Y, conditional on X, taking on a specific value.

$$\Pr(Y = y|X) = x\,\frac{\Pr(X=x_i, Y=y_i)}{\Pr(X=x_i)}$$

	Rain (X = 0)	No Rain (X = 1)	Total
Long Commute (Y = 0)	0.15	0.07	0.22
Short Commute(Y = 1)	0.15	0.63	0.78
Total	0.30	0.70	1.00

- Joint Distribution: Pr (X = 1, Y = 0) = 0.07
- Marginal Distribution of Y:
Pr (Y = 0) = Pr (Y = 0, X = 1) + Pr (Y = 0, X = 0) =0.15 + 0.07 = 0.22
- Conditional Distribution of Y=0 |X=1: 0.07/0.70=0.10

Moments

- **Mean** = Expected value (expectation) of Y

  = $E(Y)$

  = $\mu\mathrm{B}_{Y\mathrm{B}}$

  = long-run average value of Y over repeated realizations of Y

Variance = $E(Y - \mu\mathrm{B}_{Y\mathrm{B}})^2$

  = σ_Y^2

  = measure of the squared spread of the distribution

- **Standard Deviation**= $\sigma\mathrm{B}_{Y\mathrm{B}}$

- Skewness = measure of asymmetry of a distribution= $\frac{E[(Y-\mu_Y)^3]}{\sigma_Y^3}$

  - skewness = 0: distribution is symmetric

  - skewness $>$ ($<$) 0: distribution has long right (left) tail

Moments

Kurtosis = measure of mass in tails

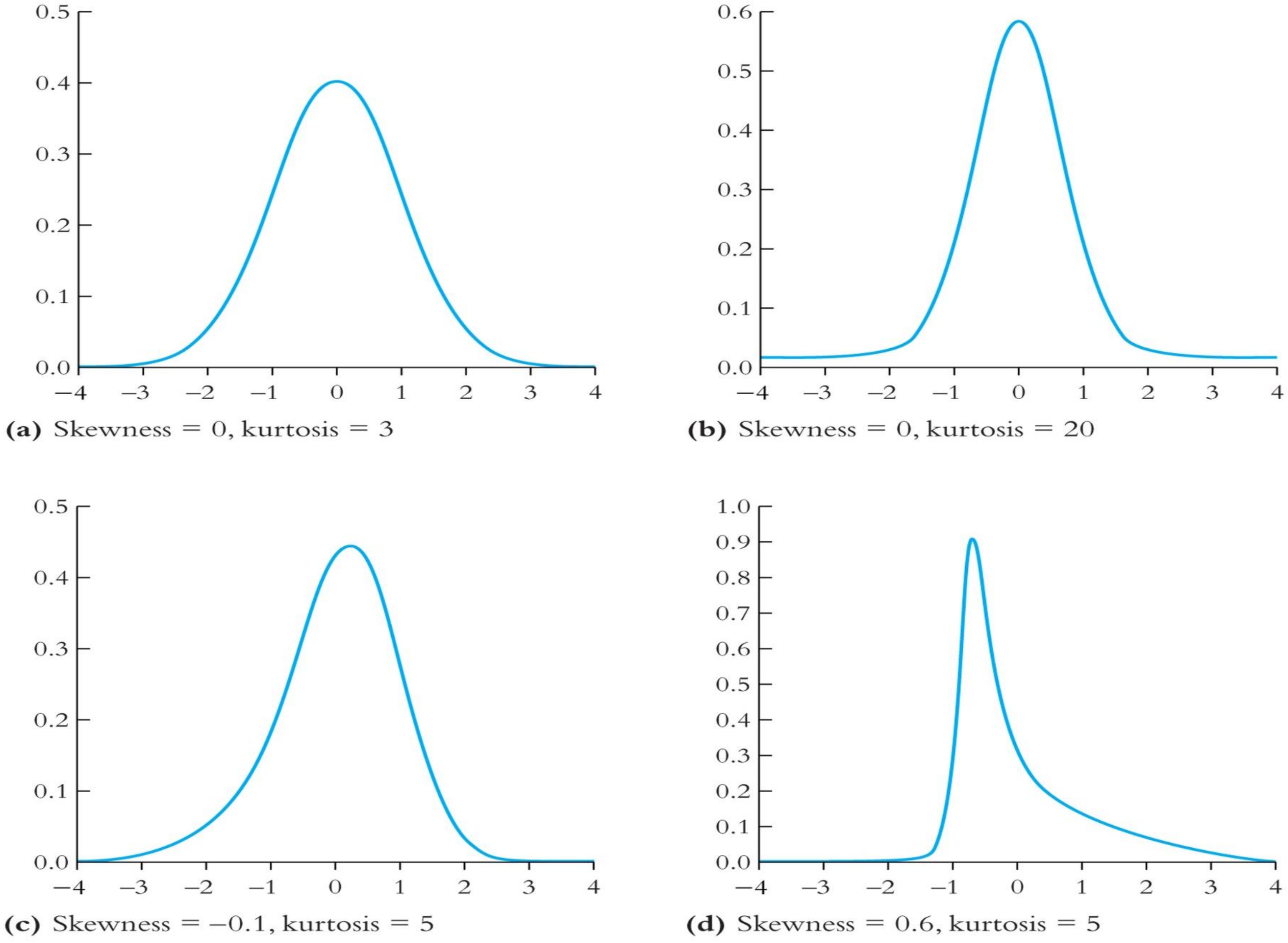
= measure of probability of large values

$$= \frac{E[(Y-\mu_Y)^4]}{\sigma_Y^4}$$

kurtosis = 3: normal distribution

kurtosis > 3: heavy tails (“leptokurtotic”)

Moments



(a) Skewness = 0, kurtosis = 3

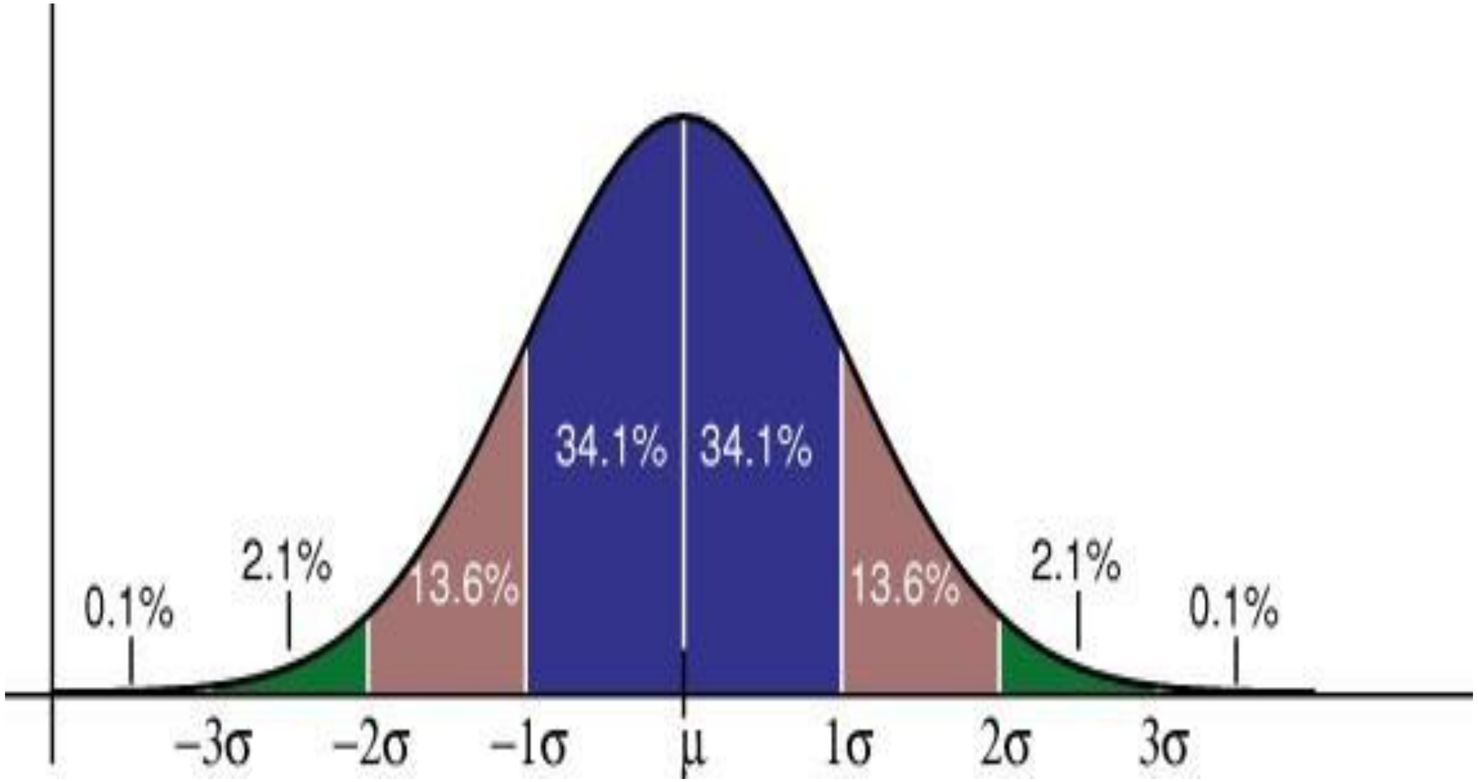
(b) Skewness = 0, kurtosis = 20

(c) Skewness = −0.1, kurtosis = 5

(d) Skewness = 0.6, kurtosis = 5

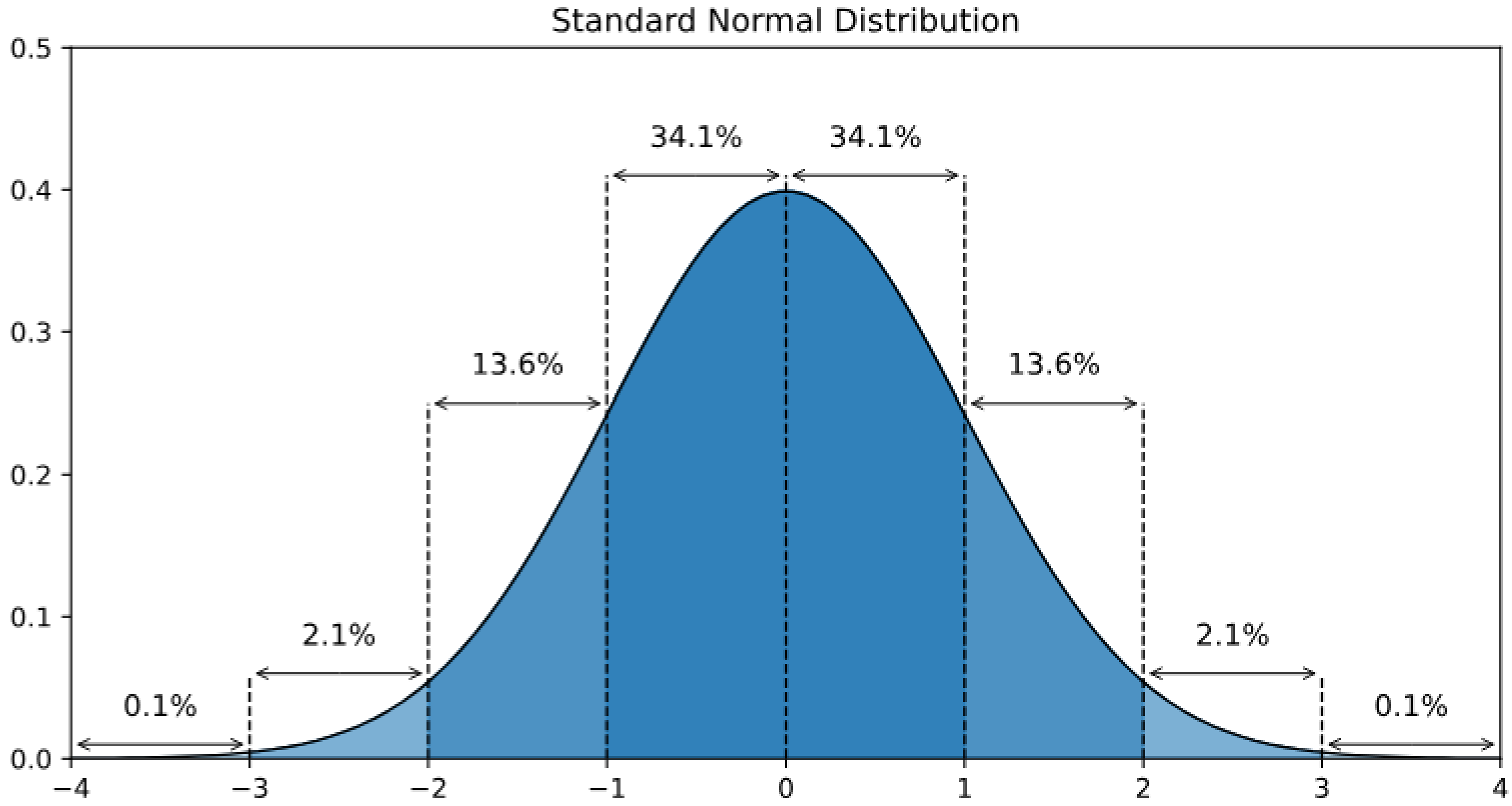
Normal Distribution

- Normal distribution is a continuous distribution.
- It has a bell shaped density function.
- $X \sim N(\mu_x, \sigma_x^2)$



Standard Normal Distribution

- Standard Normal distribution is a special normal distribution.
- $X \sim N(0,1)$



Chi-Squared, Student t, and F Distributions

- Chi-squared distribution is the sum of m independent standard normal distributions. Where, 'm' is called the degrees of freedom (df).

$$\chi_m^2 = \sum_{i=1}^{m} Z_i^2$$

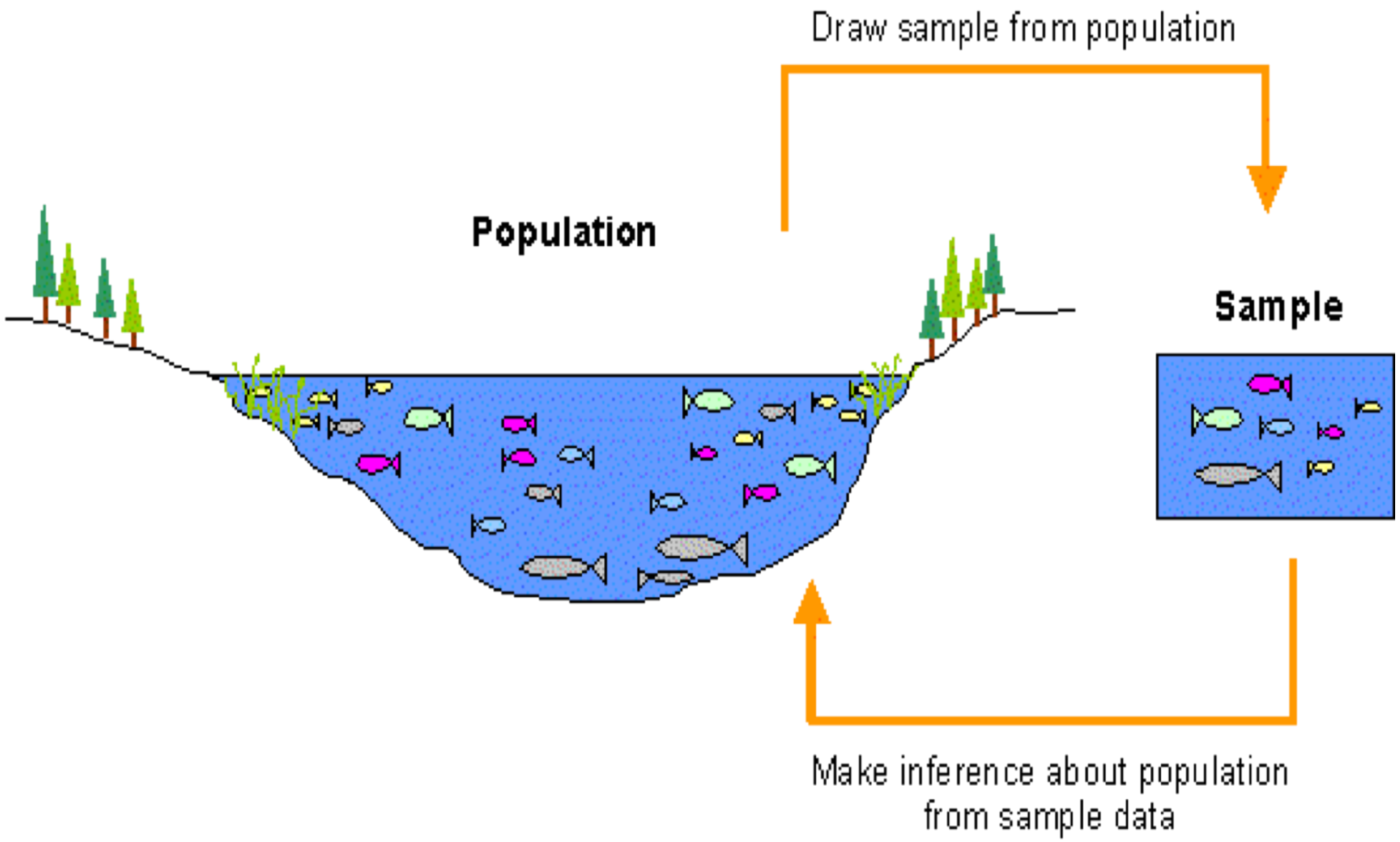
- Student t distribution is the ratio of a standard normal and the square root of an independent Chi-squared distributions with m df , divided by m. Or, one can use the estimated regression coefficients.

$$t_m = \frac{Z}{\sqrt{\chi_m^2/m}} \text{ or } t = \frac{\hat{\beta}_j - \hat{\beta}_{j,0}}{SE\ (\hat{\beta}_j)}$$

- F distribution is the ratio of two independent Chi-squared distributions with degrees of freedom m and n, divided by the respective df .

$$\frac{\chi_m^2/m}{\chi_n^2/n}$$

Population Vs. Sample



Sample Mean and Variance

- Let n be the number of observations in the sample, then sample mean:

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

- Given the same setup, sample variance:

$$Var(\bar{Y}) = \sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{n}$$

- Sampling distribution is a distribution of sample means ($\bar{Y}$)

Central Limit Theorem (CLT) and Law of Large Number (LLM)

- CLT sates that regardless of the population distribution model, as the sample size increases, the sample mean tends to be normally distributed around the population mean, and its standard deviation shrinks as n increases
- Certain conditions must be met to use the CLT
  - The samples must be independent
  - The sample size must be "big enough"
- LLM states that, under general conditions, $\bar{Y}$ will be near μY with high probability as n is very large.

Sampling Distribution