Quantitative Reasoning – LAB 413

# Probability & Sampling Distributions

Week 6 Recap ▪ Lab 5

Instructor: **Subin Na**

October 1, 2025

# Today's Plan

1. **Quick Recap**
   - Why Probability Matters
   - Essential Probability Rules
   - Random Variables
   - Statistical Inference
   - Sampling Distributions
   - Key Theorems
2. **RStudio Practice**

# 01 Why Probability Matters

**The Problem**

How do we know if our estimates are *"real"* or just due to random chance?

**The Solution.** Probability helps us understand random variability and uncertainty.

**Probability**

Know population distribution $\rightarrow$ predict sample outcomes

**Statistical Inference**

Have sample data $\rightarrow$ infer population characteristics

# 02 Essential Probability Rules

## Fundamental Rules

**Range:** $0 \leq P(A) \leq 1$

**Total Probability:** $P(S) = 1$

**Complement Rule:** $P(\text{not } A) = 1 - P(A)$

**Addition Rule:** $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

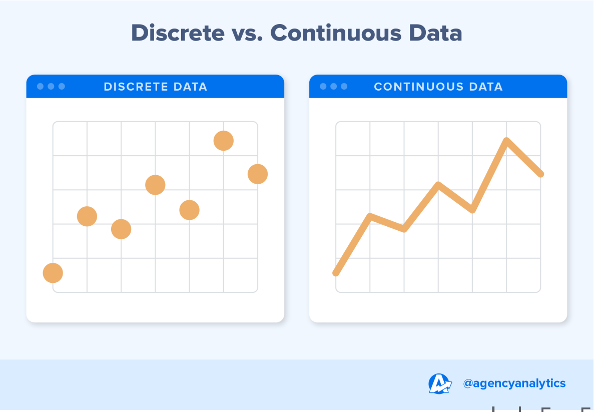
**Multiplication Rule:** $P(A \text{ and } B) = P(A) \times P(B \mid A)$

**Definition**

A random variable **assigns numeric values to outcomes**.

## Two types

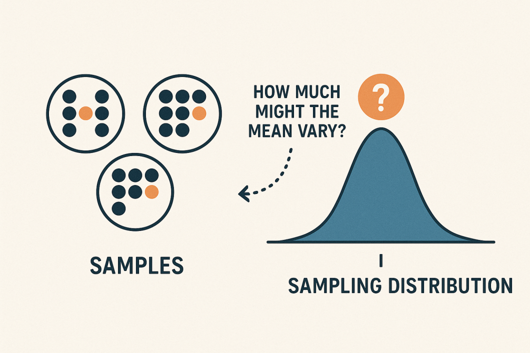
- **Discrete:** countable values (e.g., number of heads in coin flips, survey responses)
- **Continuous:** any real value (e.g., height, income)



# 04 Statistical Inference

**Goal.** Use sample statistics to estimate population parameters.

- **Parameter:** unknown population characteristic ($\mu$, $\sigma$)
- **Statistic:** sample-based estimate ($\bar{x}$, $s$)
- **Sampling Distribution:** how statistics vary across samples



# 05 Sampling Distributions

**Definition**

The distribution of a statistic across **all possible samples**.

## Key Properties

- **Sampling Variability:** statistics vary across samples
- **Bias:** does the statistic center on the true parameter?
- **Variability:** how spread out are the sample statistics?

## Managing Quality

- Reduce **bias** $\rightarrow$ use random sampling
- Reduce **variability** $\rightarrow$ increase sample size

### 1. Law of Large Numbers

Sample mean approaches the population mean as $n$ increases:

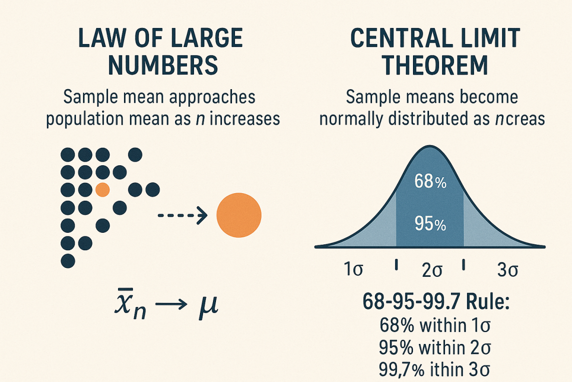
$$\bar{x}_n \to \mu$$

### 2. Central Limit Theorem

Sample means become normally distributed as $n$ increases.

**68–95–99.7 Rule**

68% within $1\sigma$, 95% within $2\sigma$, 99.7% within $3\sigma$.



# RStudio Practice

## Problem Set 2 with the `gapminder` Data

See `Lab5.Rmd` / `Lab5_1001.R`