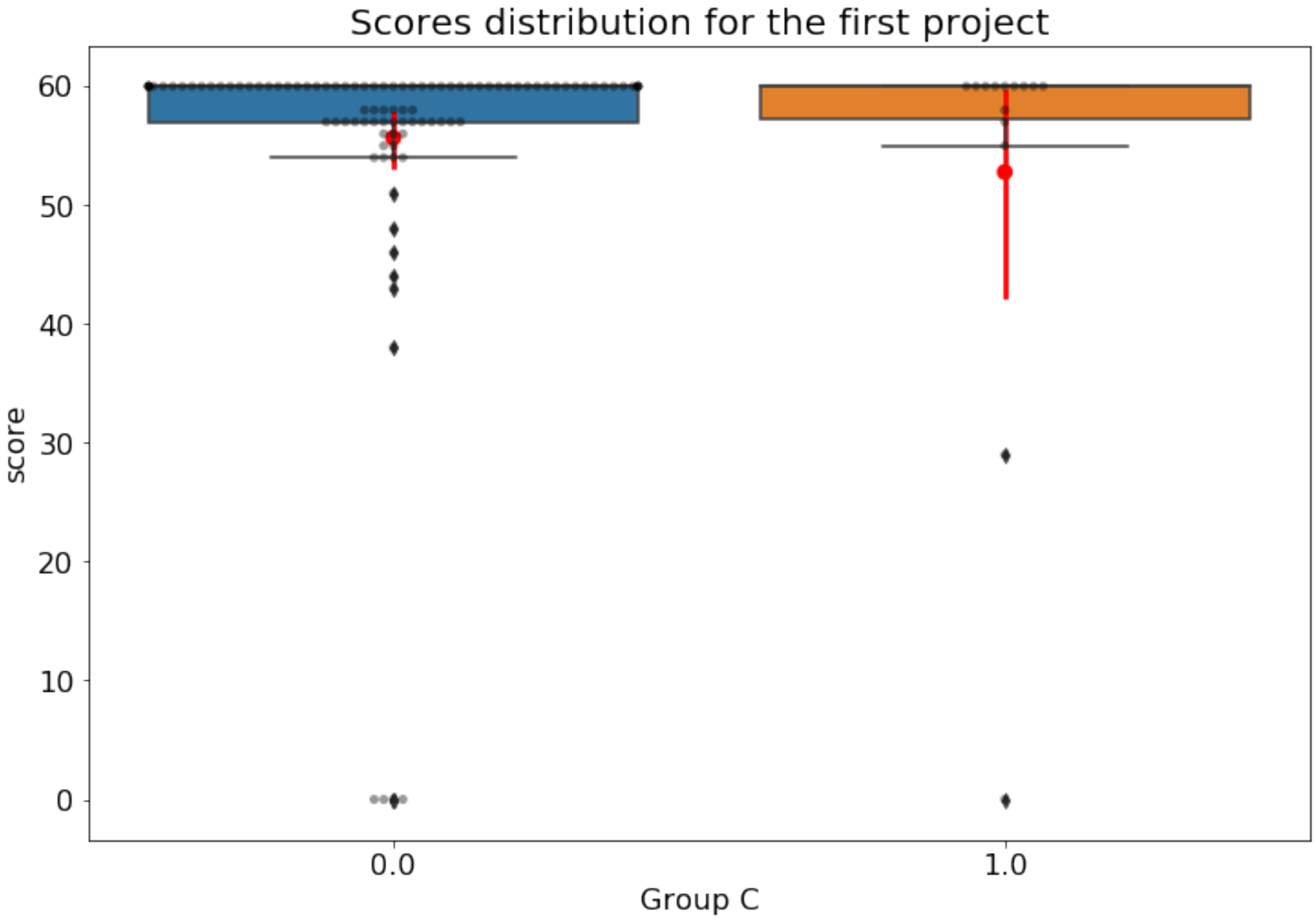
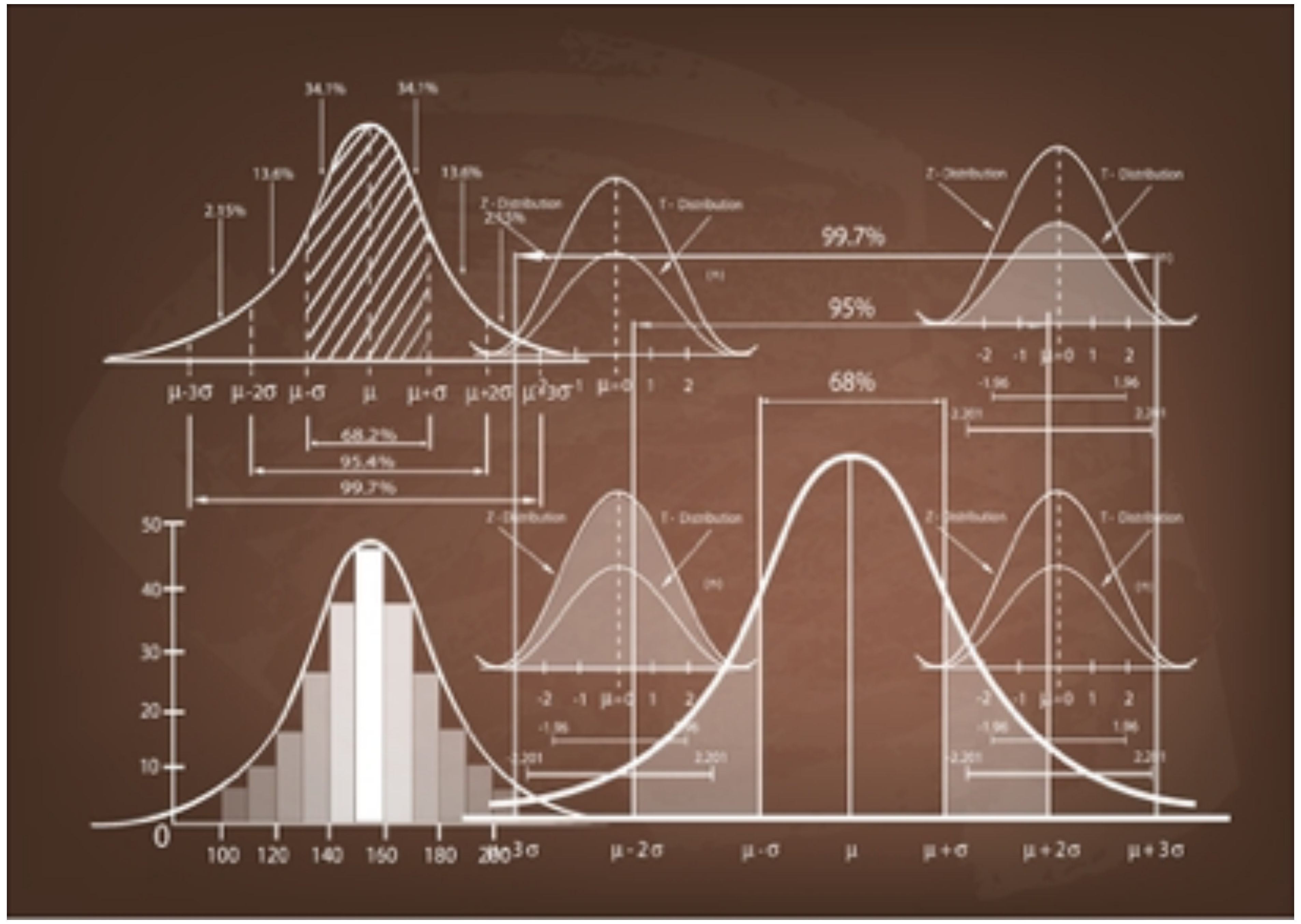
Project 1 results



Business Statistics

MLS 1 - Inferential Statistics Foundations



Adel Benlagra
July 10th, 2022

Topics covered in the module

- Week 0: Pre-work on Probability Basics
- Week 1: Inferential Statistics Foundations
- Week 2: Estimation and Hypothesis Testing
- Week 3: Common Statistical Tests
- Week 4: Project Debrief

Session objectives

Mentored Learning Session 1: Inferential Statistics Foundations

Learning Objectives

- Understand the concepts of probability and random variables
- Understand the basics of probability distributions
- Illustrate the importance of distribution in daily life through the Medicon Use case using scipy.stats

Concepts covered in week 1

1. Intro and agenda
2. Introduction to inferential statistics
3. Fundamental terms in distributions
4. Binomial distribution theory
5. Binomial distribution hands-on
6. Uniform distribution theory
7. Continuous uniform distribution hands-on
8. Normal distribution theory
9. Normal distribution hands-on
10. Z-score hands-on

Probability

- Probability refers to chance or likelihood of a particular event taking place.
- An event is an outcome of an experiment.
- An experiment is a process that is performed to understand and observe possible outcomes.
- Set of all outcomes of an experiment is called the sample space.

Probability : Example

A company is conducting a telephone survey of randomly selected individuals to get their overall impressions of the year 2020. So far, 1646 people have been surveyed. The frequency distribution shows the results. What is the probability that the next person surveyed has a positive overall impression of 2020?

Response	Number of times
Positive	201
Negative	903
Neither	235
Don't know	307
Sum	1646

P(response is positive) = #Positive Responses / #Total Responses

P(response is positive) = $201/1646 \approx 0.12$

Random Variable

A **random variable** assigns a value of each outcome of an experiment. It assumes different values with different probabilities.

There are **two types** of random variables:

- **Discrete random variable**: it can take a finite number of values. For example: Number of employees getting promoted in an organization
- **Continuous random variable**: it can take an infinite number of values in a given range. For example: Speed of an aircraft

Suppose there are 1000 students in the university.

Passing an exam is an example of a discrete random variable: pass or fail

Bernouilli and Binomial distributions

- The **Bernouilli distribution** describes a random variable with **two possible outcomes**, referred to as *Success or 1* and *Failure or 0*, **in a single trial**

$$X = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } 1-p \end{cases}$$

- The **Binomial Distribution** describes the number of successes *x* in *n* repeated trials each with two possible outcomes. It generalizes the Bernouilli distribution. Its probability mass function writes

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$$

- The Binomial distribution assumes the following
  - Trials are independent of each others
  - The probability of success *p* is the same for each trial
  - The number of trials is fixed
- The Bernouilli and Binomial distributions are implemented in Python with the *scipy.stats.bernouilli* and *scipy.stats.binom* classes respectively

Uniform distribution

- The **uniform distribution** describes a random variable with mutually exclusive outcomes having an equal probability of occurrence
- For a **discrete uniform distribution** with *m* possible outcomes, the probability mass function is

$$P(X = x) = \begin{cases} \frac{1}{m}, & \text{is x is in the possible outcomes} \\ 0, & \text{else} \end{cases}$$

- For a **continuous uniform distribution** with values in the [a,b] range, the probability density function is

$$P(X = x) = \begin{cases} \frac{1}{b - a}, & \text{if } \; a \leq x \leq b \\ 0, & \text{else} \end{cases}$$

- The continuous and discret uniform distributions are implemented in Python with the *scipy.stats.uniform* and *scipy.stats.randint* classes respectively

Normal distribution

- The **normal distribution** is a continuous distribution that is widely used in inferential statistics
- As will be seen in the Central Limit Theorem, the normal distribution is the asymptotic distribution of the normalized sum of independent variables regardless of their distribution
- The normal distribution is entirely characterized by its two parameters μ, σ and its probability density function writes

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

- The normal distribution has a symmetric bell-shaped curve and equal mean, median and mode
- The Z-score is a standard normal variable that measures the number of standards aviations above or below the mean

$$Z = \frac{X_{norma} - \mu}{\sigma}$$

- The normal distribution is implemented in Python with the *scipy.stats.norm* class

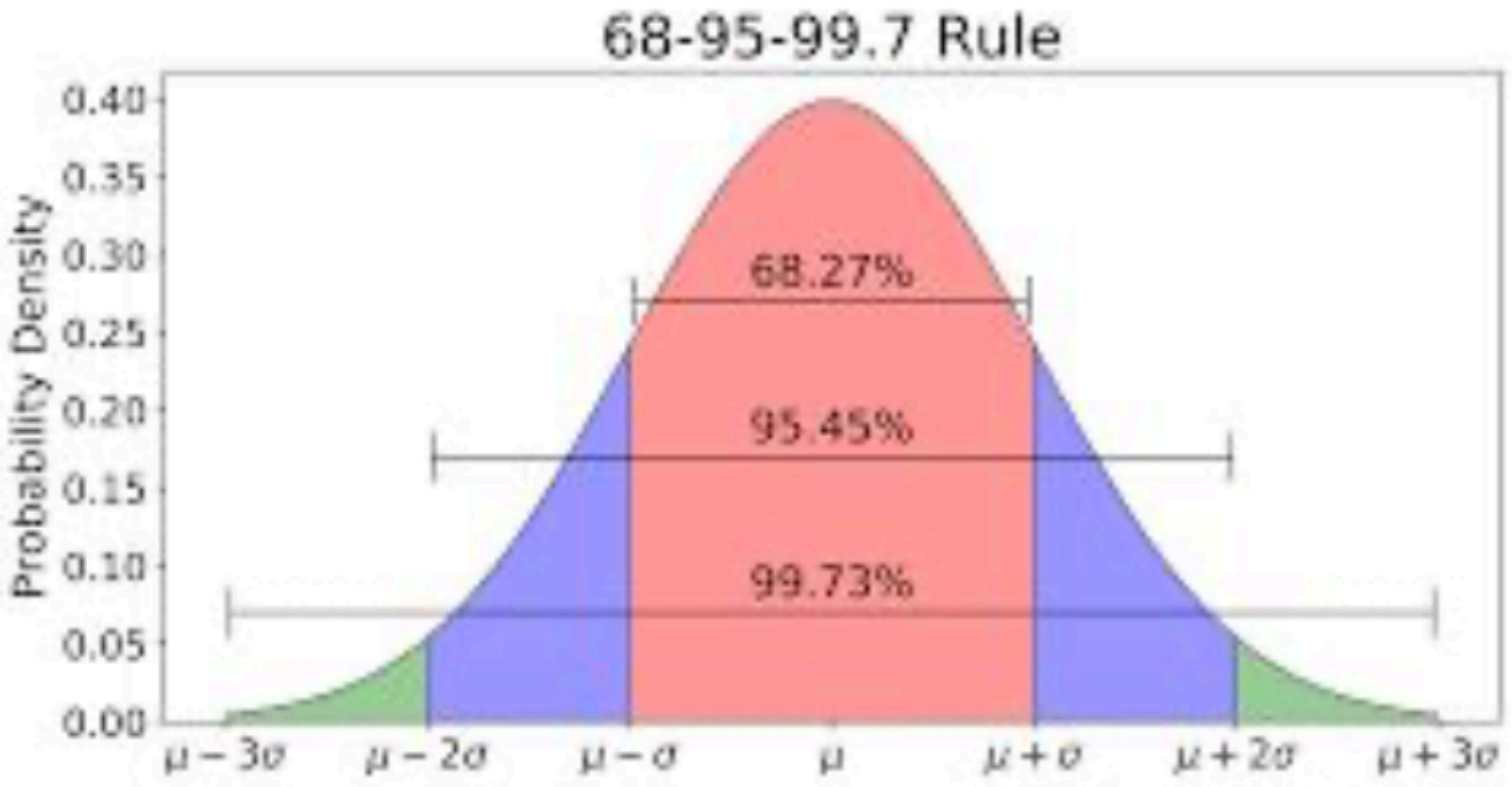
Normal distribution

- The **normal distribution** is a continuous distribution that is widely used in inferential statistics
- As will be seen in the Central Limit Theorem, the normal distribution is the asymptotic distribution of the normalized sum of independent variables regardless of their distribution

- The normal distribution has a symmetric bell-shaped curve and is entirely characterized by its two parameters μ, σ
- The Z-score is a standard normal variable that measures the number of standards aviations above or below the mean

$$Z = \frac{X_{norma} - \mu}{\sigma}$$

- The normal distribution is implemented in Python with the *scipy.stats.norm* class



Annexe: Statistics of distributions

Distribution	Parameters	Mean	Median	Mode	Standard deviation
Bernouilli	p	p	$\lfloor p \rfloor$ or $\lceil p \rceil$	$\lfloor 2p \rfloor$ or $\lceil 2p \rceil - 1$	p(1-p)
Binomial	n, p	np	$\lfloor np \rfloor$ or $\lceil np \rceil$	$\lfloor (n+1)p \rfloor$ or $\lceil (n+1)p \rceil - 1$	np(1-p)
Uniform (Continuous)	a, b	$\frac{1}{2}(a+b)$	$\frac{1}{2}(a+b)$	Any value in (a, b)	$\sqrt{\frac{1}{12}}(b-a)$
Normal	μ, σ	μ	μ	μ	σ

$\lfloor . \rfloor$ is the floor function and $\lceil . \rceil$ is the ceil function