307102
Descriptive Statistics for Business

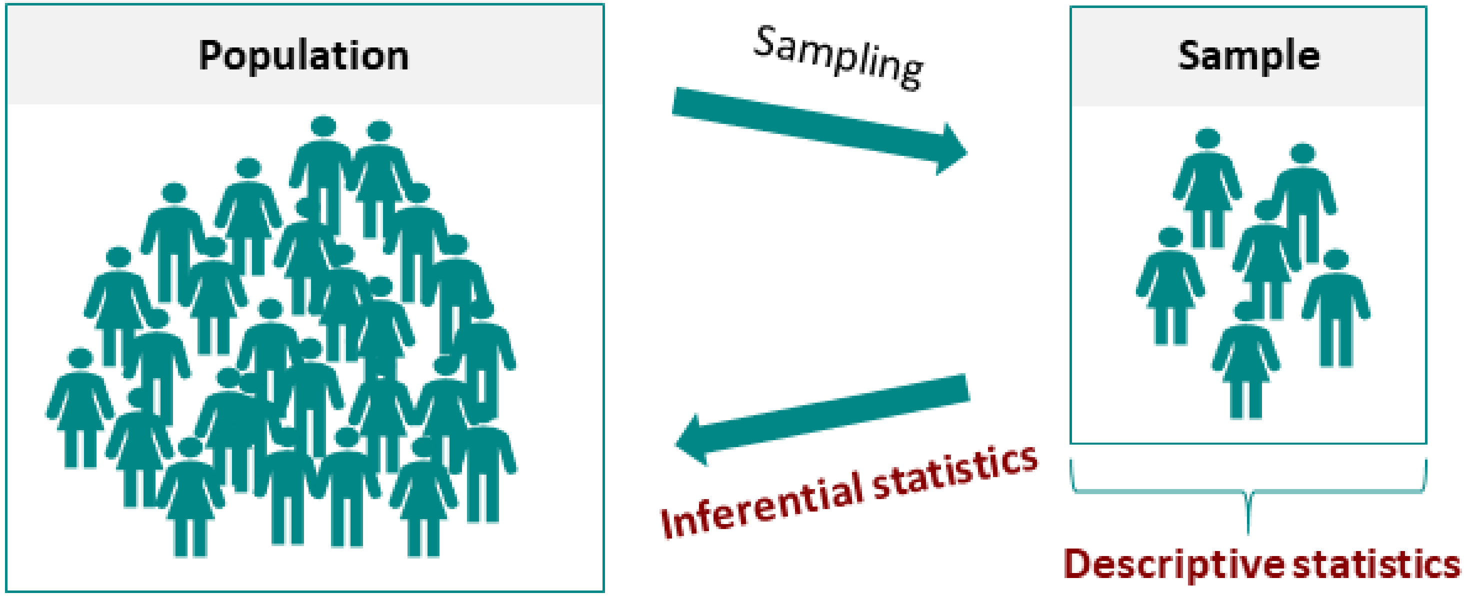
Introduction to Probability Distributions

Content

- Introduction to Inferential Statistics
- Random Variables and Probability Distributions
  - Discrete Random Variables
  - Continuous Random Variables
- The Normal Distribution
- The Empirical Rule
- The Standard Normal Distribution and the Standard Scores (Z-Scores)
- Using the Z-Tables

Introduction to Inferential Statistics

- **Definition**: Inferential statistics is the branch of statistics that involves making inferences or predictions about a population based on sample data.
- **Objective**: The main goal of inferential statistics is to draw conclusions, make predictions, or test hypotheses about a population using information obtained from a sample.



Inferential Statistics Uses

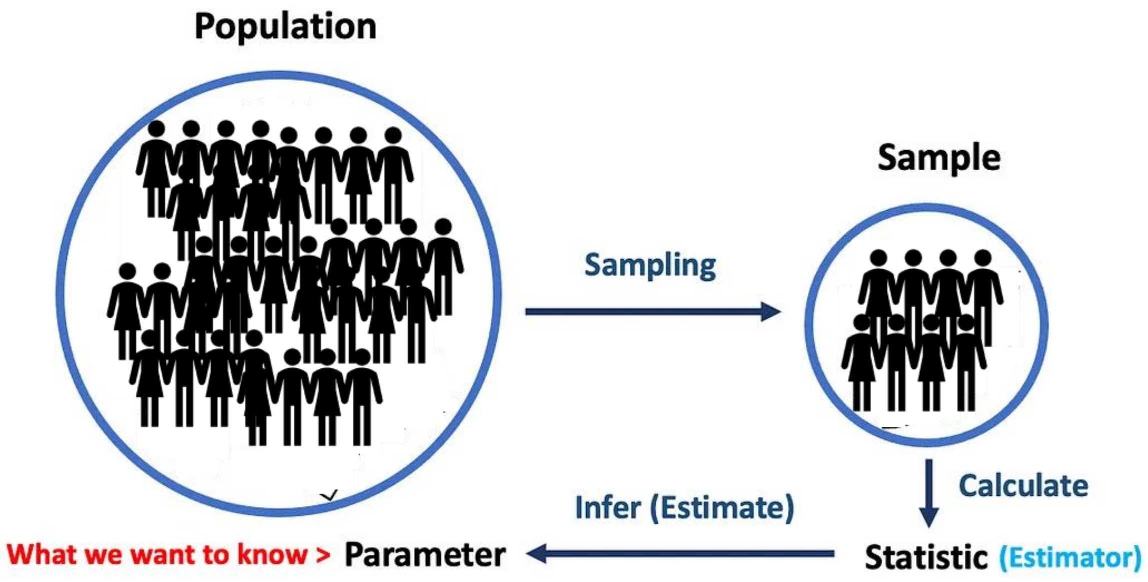
Inferential statistics is used in various fields and contexts to make predictions, draw conclusions, and make inferences about populations based on sample data. Here are some common areas and applications where inferential statistics is used:

- **Scientific Research**: Inferential statistics is fundamental in scientific research across disciplines such as biology, physics, chemistry, and environmental science. Researchers use statistical tests to analyze data and draw conclusions about hypotheses.
- **Business and Economics**: Businesses use inferential statistics for market research, sales forecasting, quality control, and decision-making. Econometric models are employed to analyze economic data and make policy recommendations.
- **Healthcare and Medicine**: Medical researchers and healthcare professionals use inferential statistics to study the effectiveness of treatments, analyze patient data, and draw conclusions about disease prevalence. Clinical trials rely heavily on inferential statistics.
- **Education**: In the field of education, inferential statistics are used to assess the effectiveness of teaching methods, evaluate standardized test scores, and make policy decisions about educational programs.
- **Market Research and Data Analysis**: Market researchers use inferential statistics to make predictions about consumer preferences, market trends, and the impact of marketing campaigns.
- **Finance and Investment**: In finance, inferential statistics are used to assess investment risk, analyze stock market data, and estimate future asset prices. Portfolio optimization and risk management rely on statistical modeling.
- **Criminal Justice and Criminology**: Researchers and law enforcement agencies use inferential statistics to analyze crime data, study crime patterns, and evaluate the effectiveness of criminal justice programs.
- **Sports and Athletics**: In sports analytics, inferential statistics are used to analyze player performance, predict game outcomes, and make strategic decisions in sports management.

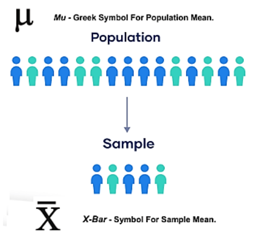
Introduction to Inferential Statistics

Key Concepts:

- **Population**: The entire group of individuals or items of interest that we want to study.
- **Sample**: A subset of the population from which data is collected and analyzed.
- **Parameter**: A numerical characteristic of a population (e.g., mean, proportion).
- **Statistic**: A numerical characteristic of a sample (e.g., sample mean, sample proportion).



https://medium.com/towards-data-science/what-is-p-value-370056b8244d



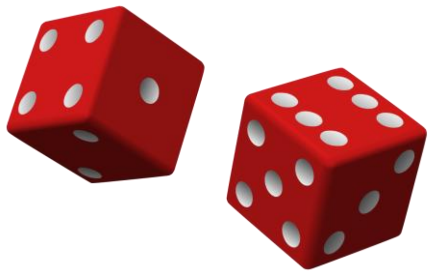
Types of Inferential Statistics

1. Estimation
   - Point Estimation: Provides a single best guess about a population parameter (e.g., the mean, proportion).
   - Interval Estimation (Confidence Intervals): Gives a range of values for a population parameter with a certain level of confidence.
2. Hypothesis Testing
   - A statistical method that uses sample data to evaluate a hypothesis about a population parameter, determining whether there is sufficient evidence to accept or reject a given hypothesis about that population.
3. Regression Analysis
   - Assesses the linear relationship between one continuous dependent variable and one or more independent variables.

Introduction to Random Variables

Understanding the Basics of Probability in Statistics

What is a Random Variable?



- A random variable is a variable whose possible values are outcomes of a random process.
- A random process is a process or event or where the outcome cannot be predicted with certainty, and each possible outcome has an associated probability.
- It can be classified as discrete or continuous depending on the range of values that it assumes.
- A discrete random variable assumes a countable number of distinct values, whereas a continuous random variable is characterized by measurable (uncountable) values.
- Example: Flipping a coin (discrete), measuring the amount of rain in a day (continuous)

Discrete Random Variables

A <u>Discrete Random Variable</u> is a variable which can take on only a countable number of distinct values like 0, 1, 2, 3, 4, 5…100, 1 million, etc.

Examples of discrete random variables:

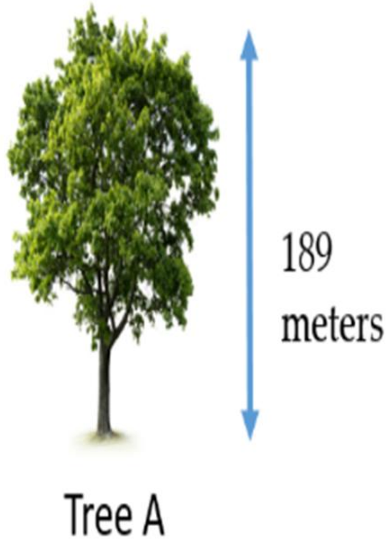
- The number of cars that passes through a toll booth everyday in a given day.
- The number of times a coin lands on tails after being flipped 20 times.
- The number of defective widgets in a box of 50 widgets.
- Number of Calls Received by a Call Center.
- Number of Emails Received in a Day

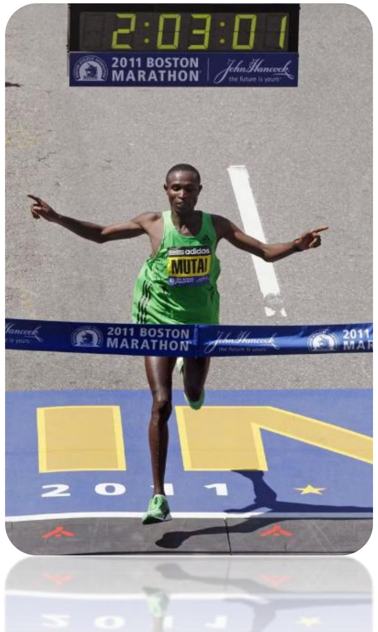
Continuous Random Variables

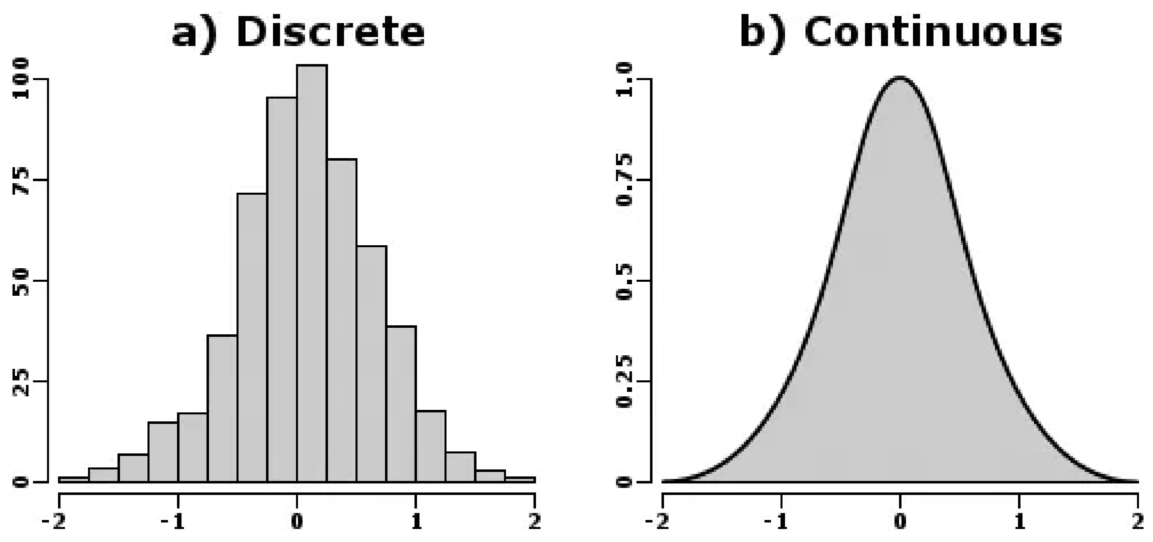
A Continuous Random Variable is a variable which can take on an infinite number of possible values.

Examples of continuous random variables:

- Business profits.
- The height of a tree.
- The weight of an animal.
- Arrival times of airlines.
- Time required to run a marathon.
- The weight of a chocolate bar on a production line.







Introduction to Probability Distributions

Discrete and Continuous Probability Distributions

Discrete Probabilities Distributions

Introduction to Probability Distributions

- Probability distributions are mathematical functions that describe the likelihood of various outcomes in a statistical experiment.
- They provide a mathematical framework for modeling random phenomena.

Discrete vs. Continuous Probability Distributions

- Discrete distributions: Outcomes are countable and have finite or countably infinite values.
- Continuous distributions: Outcomes are uncountable and have an infinite range of values.

- For a probability distribution to be valid, it must satisfy the following two criteria:
  - The probability for each outcome must be between 0 and 1.
  - The sum of all the probabilities must add up to 1.

Discrete Probability Distributions
The Discrete Uniform Distribution

Discrete Uniform Distribution

- **Definition:**
  Equal probability is assigned to each of k possible outcomes.
- **Notation:**

$$X \sim \mathcal{U}(a, b), \quad \text{where } a, b \in \mathbb{Z},\ a \leq b$$

- **Probability Mass Function (PMF):**

$$P(X = x) = \frac{1}{b - a + 1}, \quad \text{for } x \in \{a, a+1, \dots, b\}$$

- **Mean:**

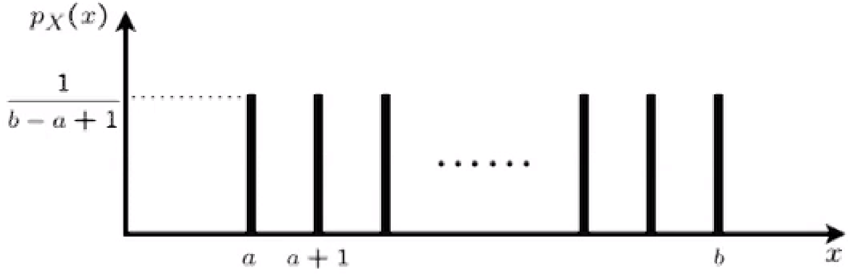
$$\mu = \frac{a + b}{2}$$

- **Variance:**

$$\sigma^2 = \frac{(b - a + 1)^2 - 1}{12}$$

- **Example:**
  Rolling a fair die:

$$X \sim \mathcal{U}(1, 6)$$



$$\mathbf{E}[X] = \frac{a + b}{2}$$

$$\text{var}(X) = \frac{1}{12}(b - a)(b - a + 2)$$

Discrete Uniform – PMF Uses

Calculating individual probabilities

- Example: For a fair die (U(1,6)), the probability of rolling a 3 is:
- $P(X = 3) = 1/(6-1+1) = 1/6 \approx 0.167$

Finding probabilities of events

- For multiple outcomes: Sum the individual probabilities
- Example: Probability of rolling an even number:
- $P(X \in \{2,4,6\}) = P(X=2) + P(X=4) + P(X=6) = 1/6 + 1/6 + 1/6 = 3/6 = 0.5$

Range probabilities

- Example: Probability of rolling between 2 and 5 inclusive:
- $P(2 \leq X \leq 5) = (5-2+1)/(6-1+1) = 4/6 = 0.667$

Discrete Probability Distributions
The Binomial Distribution

- Models the number of successes in a fixed number of independent Bernoulli trials.
- Bernoulli trial is the experiment that has only two outcomes (Yes and No).
- Parameters: n (number of trials) and p (probability of success on each trial).

Example: The defect ratio in a car factory is 2% and the factory produces 80 cars in a month,. We can use the binomial distribution to answer questions similar to:

- what is the probability that exactly 2 cars are defective?
- what is the probability that less than 2 cars are defective?
- what is the probability that more than 2 cars are defective?
- What is the probability that all cars produced in a month will have no defects?

- **Probability Mass Function (PMF):**

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

where:

- n = number of trials
- x = number of successes
- p = probability of success
- $\binom{n}{x}$ = combinations of x successes from n trials
- **Mean (Expected Value):**

The mean answers the question, what should we expected on average?

$$\mu = E(X) = np$$

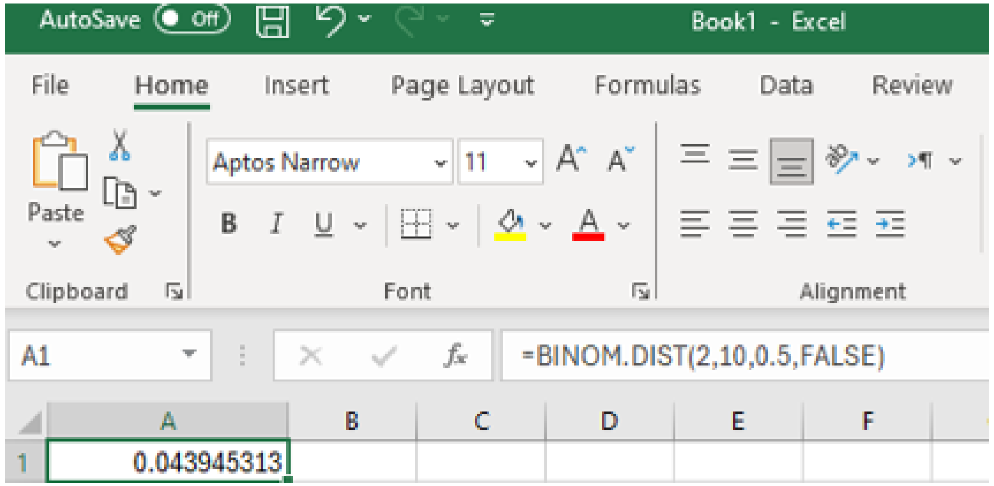
- **Variance:**

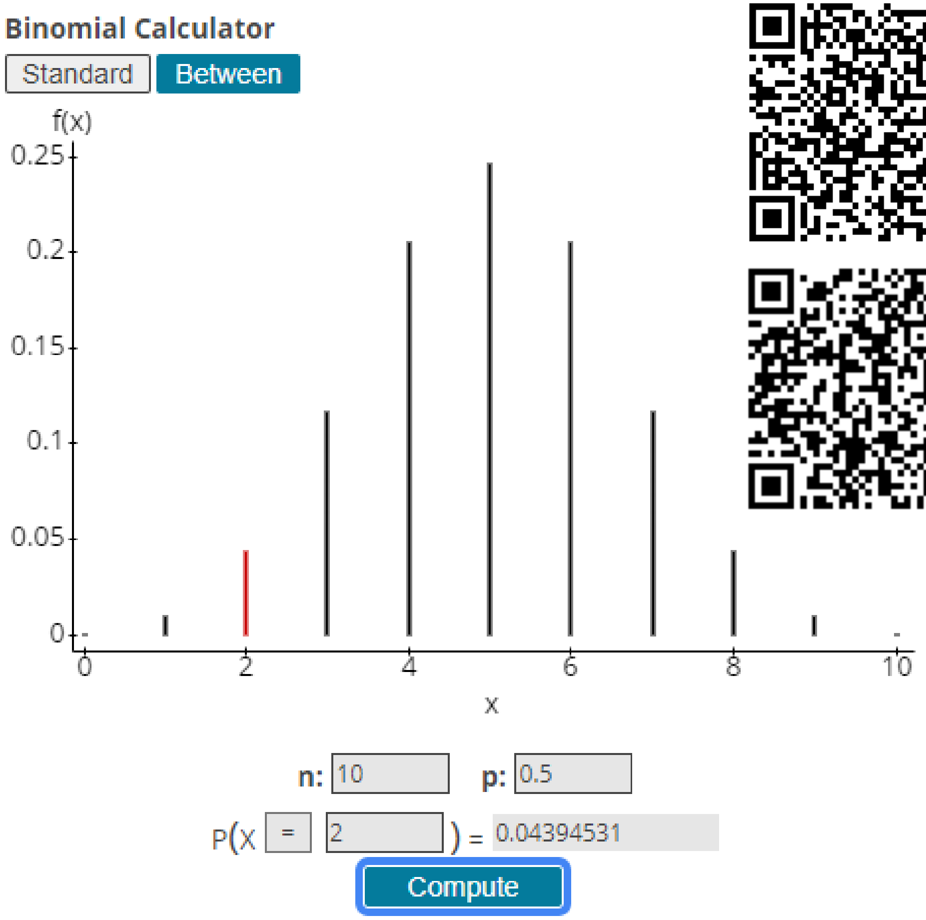
$$\sigma^2 = Var(X) = np(1-p)$$

Measures the spread of the data or how much the outcomes fluctuate around the mean. A small variance means most values are close to the mean; a large variance means more spread.

The Binomial Distribution

What is the probability of having 2 Heads ONLY out of 10 trials, knowing that the probability of Head = 50%?





Excel Function
=BINOM.DIST(2,10,0.5,FALSE)

The larger the number of trials (e.g. ask 100 students) the more the distribution will look like normal distribution.

Discrete Probability Distributions
The Poisson Distribution

- Poisson Distribution can be used to model the number of times an event occurs within a specified time period or in a given space.
- It's particularly useful for estimating the likelihood of rare events over short intervals.
- The key feature of the Poisson distribution is that it is defined by just one parameter, **λ** (lambda), which represents the average rate at which events occur.

Example: Suppose that Zein call center receives 1000 calls/hour on average, and that data center has a capacity of receiving 1100 calls/hours.

Use the Poisson Distribution to find the probability that the number of calls will exceed the call center's capacity, which will require additional temporary staff?

$$P(X=x)=\frac{\lambda^x e^{-\lambda}}{x!}$$

Where:

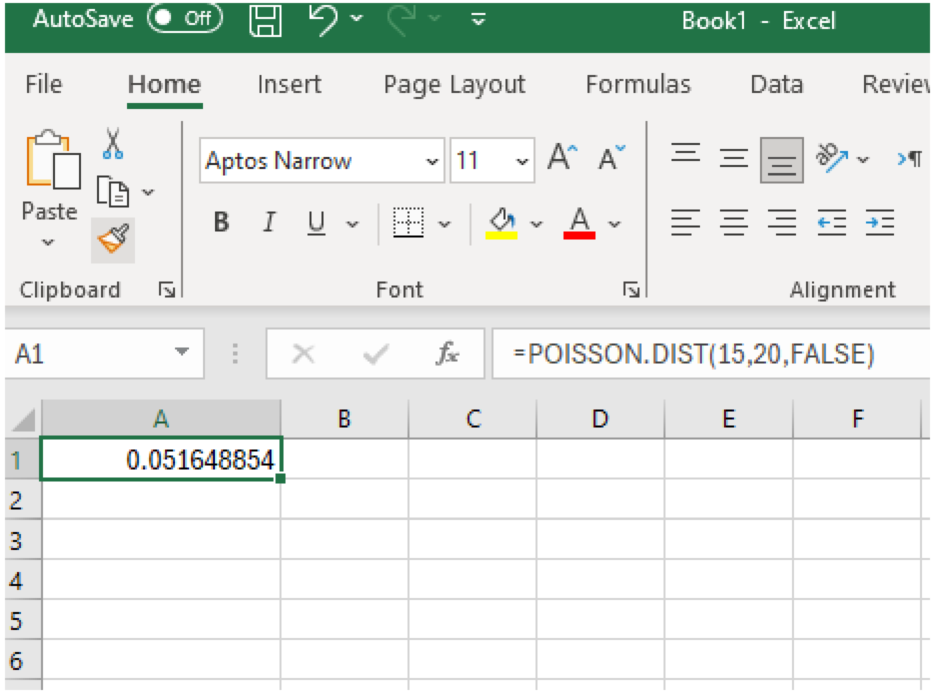
- $x = 0, 1, 2, 3, \ldots$ (number of occurrences)
- λ = average (mean) number of occurrences in the interval
- $e \approx$ 2.71828 (Euler's number)

Key Properties:

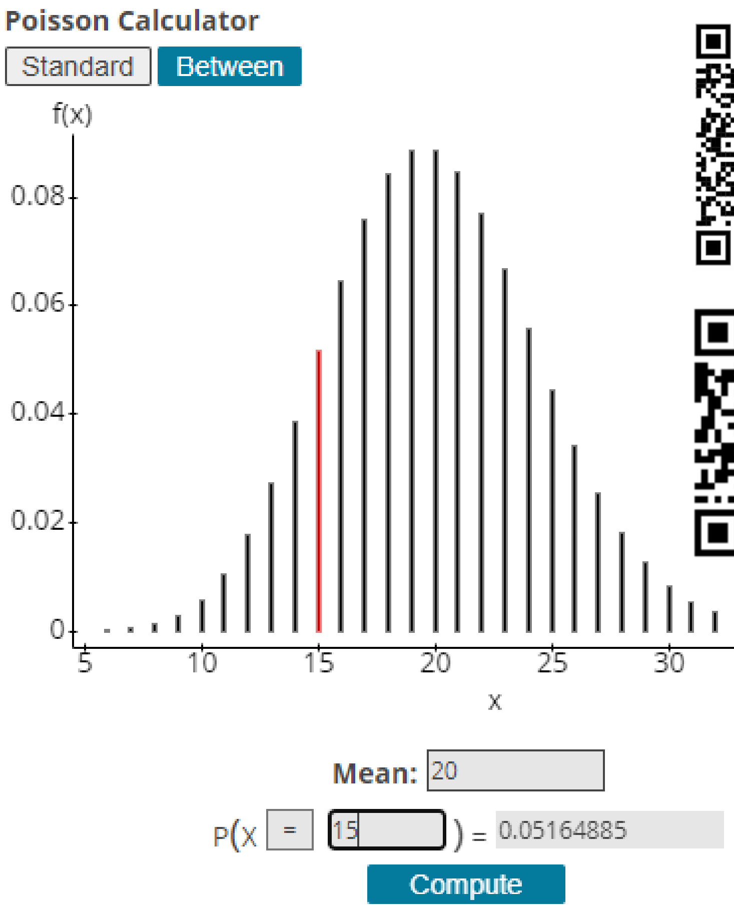
- Mean: $\mu = \lambda$ The mean answers the question, what should we expected on average?
- Variance: $\sigma^2 = \lambda$

Measures the spread of the data or how much the outcomes fluctuate around the mean. A small variance means most values are close to the mean; a large variance means more spread.

Poisson Distribution



Excel Function
=POISSON.DIST(15,20,FALSE)



The larger the average rate **λ,** the more the distribution will look like normal distribution.

Continuous Probabilities Distributions

The Continuous Uniform Distribution

Definition: A continuous uniform distribution models a scenario where **any value within a contin**
interval [a, b] is equally likely.

Probability Density Function (PDF):

$$f(x) = \frac{1}{b-a}, \quad \text{for } a \leq x \leq b$$

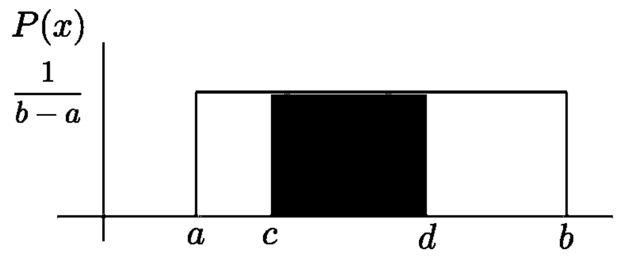
Properties:

- Mean: $\mu = \frac{a+b}{2}$
- Variance: $\sigma^2 = \frac{(b-a)^2}{12}$
- The graph is a flat line from a to b.

Example:

- Delivery time takes between 2 and 5 days with equal likelihood.
  $a = 2, b = 5$

Uniform Distribution



$$Mean: \quad \mu = \frac{a+b}{2}$$

$$S.D.: \quad \sigma = \sqrt{\frac{(b-a)^2}{12}}$$

<u>*Probability*</u>

$$P(c \leq X \leq d) = \frac{d-c}{b-a}$$

The Continuous Uniform Distribution
Business Cases Studies

Inventory Arrival

A retailer's deliveries arrive uniformly between 8 AM and 12 PM.

$X \sim \mathcal{U}(8, 12)$, where X is arrival time (in hours).

Mean arrival time:

$\mu = \frac{8+12}{2} = 10$ AM

Probability of arrival before 9 AM:

$P(X < 9) = \frac{9-8}{12-8} = 0.25 = 25\%$

Staffing Decision:

Extra staff likely needed around mean time (10 AM).

Resource Planning:

Loading dock should be ready by 8 AM.

Manufacturing

A machine produces widgets with diameters uniformly distributed between 9.8mm and 10.2mm.

Widget diameters are uniformly distributed between 9.8 mm and 10.2 mm.

$X \sim \mathcal{U}(9.8, 10.2)$, where X is diameter (in mm).

Mean diameter:

$\mu = \frac{9.8+10.2}{2} = 10.0$ mm

Percentage exceeding 10.1 mm:

$P(X > 10.1) = \frac{10.2-10.1}{10.2-9.8} = 0.25 = 25\%$

Quality Control:

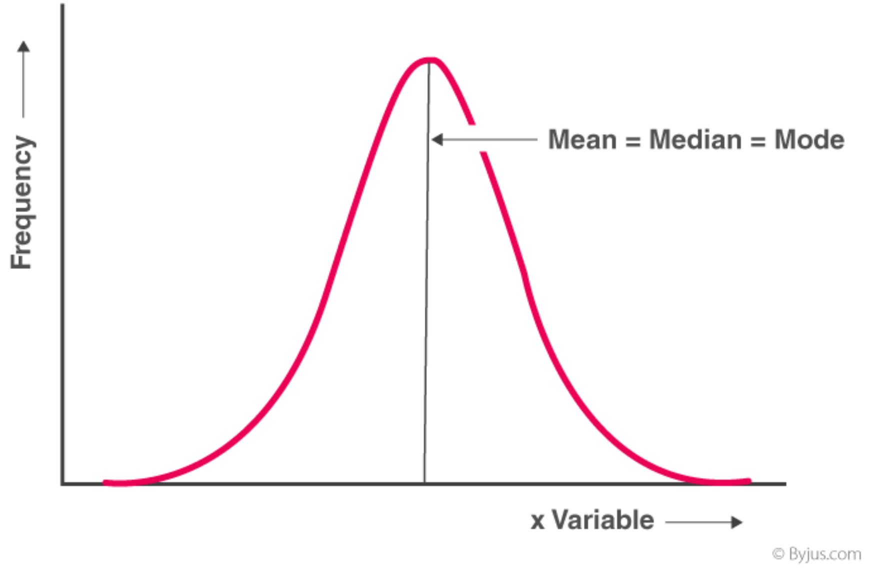
25% of widgets exceed 10.1 mm.

Variance:

$\sigma^2 = \frac{(10.2-9.8)^2}{12} = \frac{0.16}{12} = 0.0133 \text{ mm}^2$

The Normal Distribution

- The normal distribution is the most extensively used probability distribution in statistical work.
- It is also known as the Gaussian distribution and the Bell Curve.
- The x-axis represents the outcomes of the normal process (e.g. students scores, heights, salaries…etc.).
- The AREA UNDER THE CURVE represents the probability of the outcome.
- The normal distribution has a symmetric distribution where most of the observations cluster around the central peak.
- Theoretically, a normal random variable can assume any value between minus infinity and plus infinity.
- Normal Distribution Main Features:
  - Symmetrically distributed
  - Long Tails / Bell Shaped
  - Mean/ Mode and Median are the same



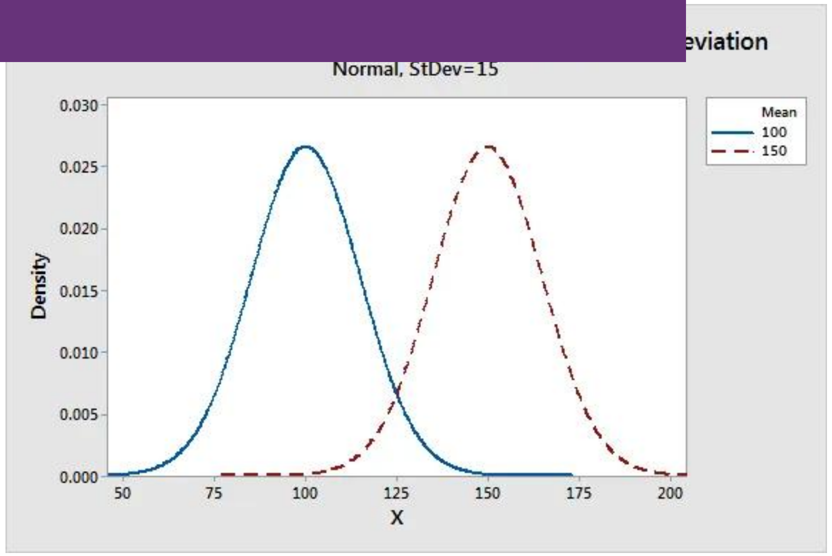
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

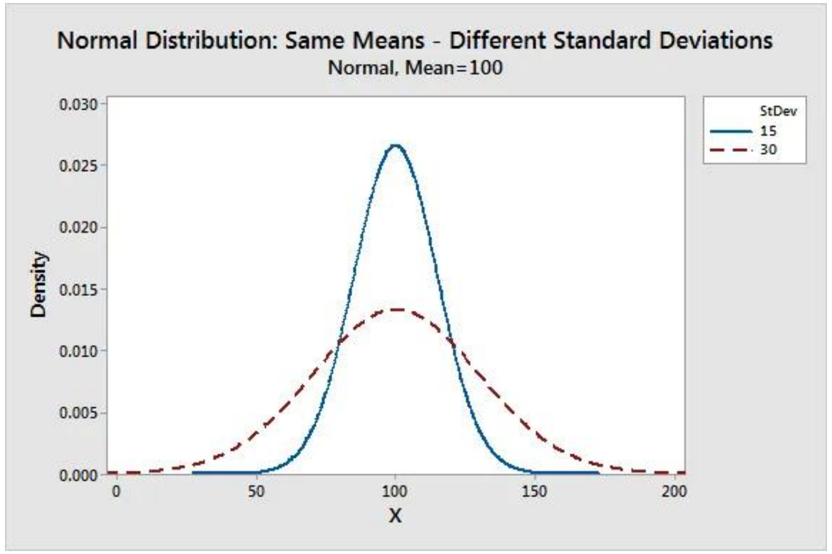
Parameters of the Normal Distribution

- The normal distribution is completely described by two parameters, the Mean and the Standard Deviation.

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

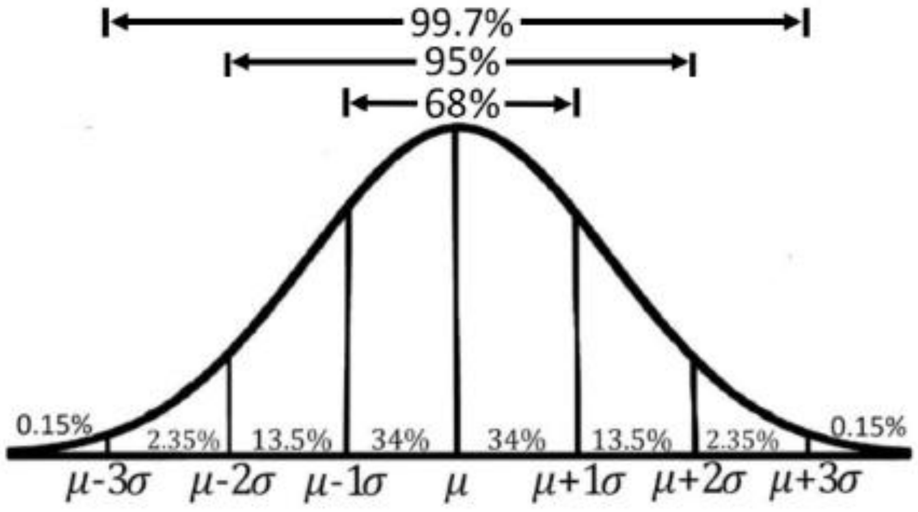
- Where X is a continuous random variable (the expected outcome of a continuous random process).
- μ = The Mean of the Outcomes of the Random Process.
- σ = The Standard Deviation of the Outcomes of the Random Process.
- π is approximately 3.14159.
- e is approximately 2.71828.





The Empirical Rule for the Normal Distribution

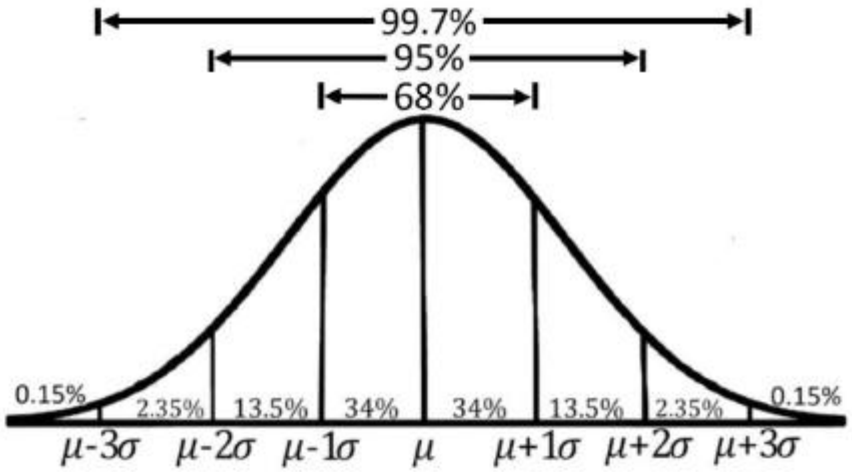
- The Empirical Rule describes how data (continuous outcomes) are spread in a normal distribution.
- The proportion of the area that falls under the curve between two points on a probability distribution indicates the probability that a value will fall within that interval.
- If we know that the data are drawn from a relatively symmetric and bell-shaped distribution - perhaps by a visual inspection of its histogram, then we can make statements about the percentage of observations that fall within certain intervals.



- Approximately 68% of data fall within one standard deviation (σ) of the mean (μ).
- Around 95% are within two standard deviations.
- About 99.7% lie within three standard deviations.

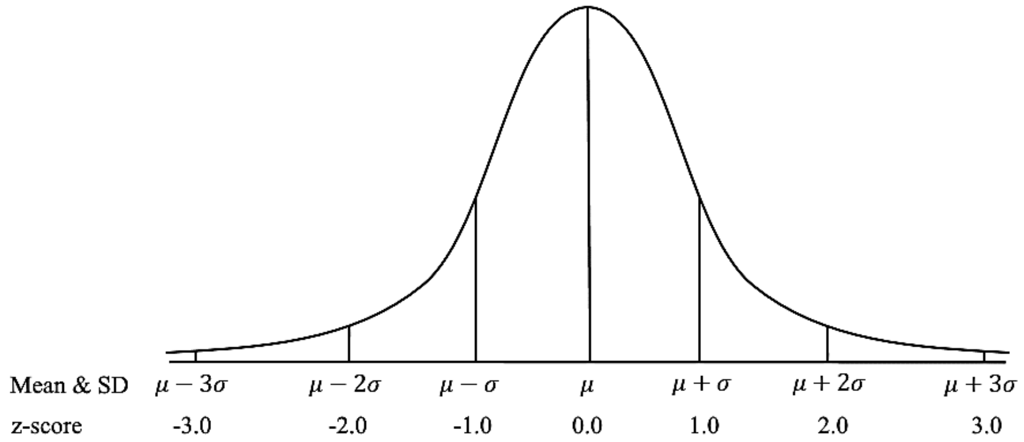
Empirical Rule Example

- A large lecture class has 280 students. The professor has announced that the mean score in an exam is 74 with a standard deviation of 8.
- Use the Empirical Rule to answer the following:
  a. Approximately how many students within 58 and 90?
  b. Approximately how many students scored higher than 90?



The Z - Score

- Definition: Z-Score represents the number of standard deviations a data point is from the mean of its distribution.
- Formula: $Z = \frac{x - \mu}{\sigma}$



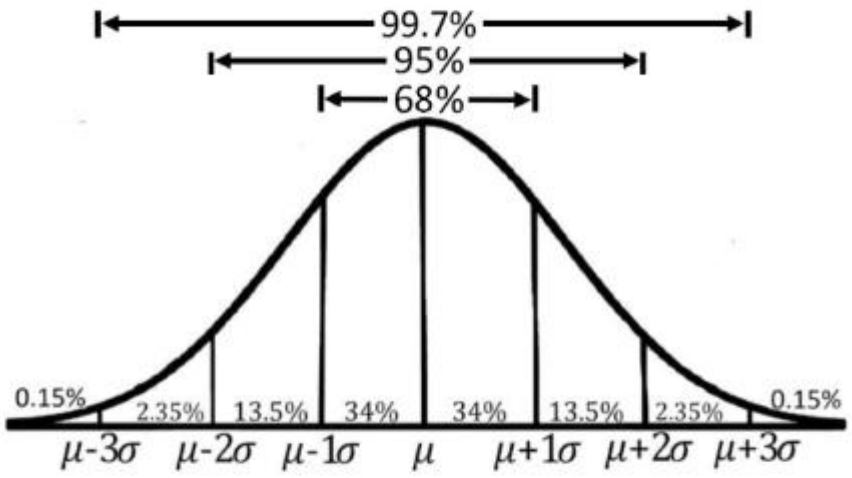
Where X is a data point, μ is the mean, and σ is the standard deviation.

Interpretation:

- A z-score of 0 corresponds to a data point that is exactly at the mean.
- Positive z-scores indicate values above the mean.
- Negative z-scores indicate values below the mean.

Back to our Empirical Rule Example

- A large lecture class has 280 students. The professor has announced that the mean score in an exam is 74 with a standard deviation of 8.



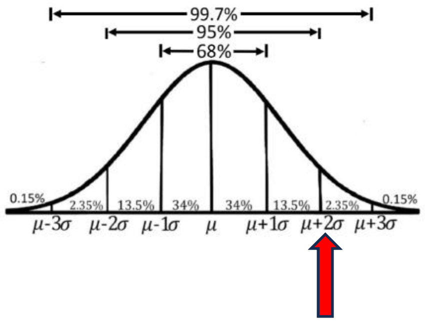
- Use the Empirical Rule to answer the following:
  a. Approximately how many students within 58 and 90?
  b. Approximately how many students scored higher than 90?

Solution for part (a)

- First, we find the z-scores corresponding to 58 and 90.
- Z for 58 = (58 – 74) / 8 = -16/8 = -2, Z for 90 = (58 – 74) / 8 = 16/8 = +2
- Using the empirical rule, we know that 95% of observations falls within 2 standard deviations (each side) from the mean, therefore, the number of students under this criteria = 280 x .95 = 266 students.

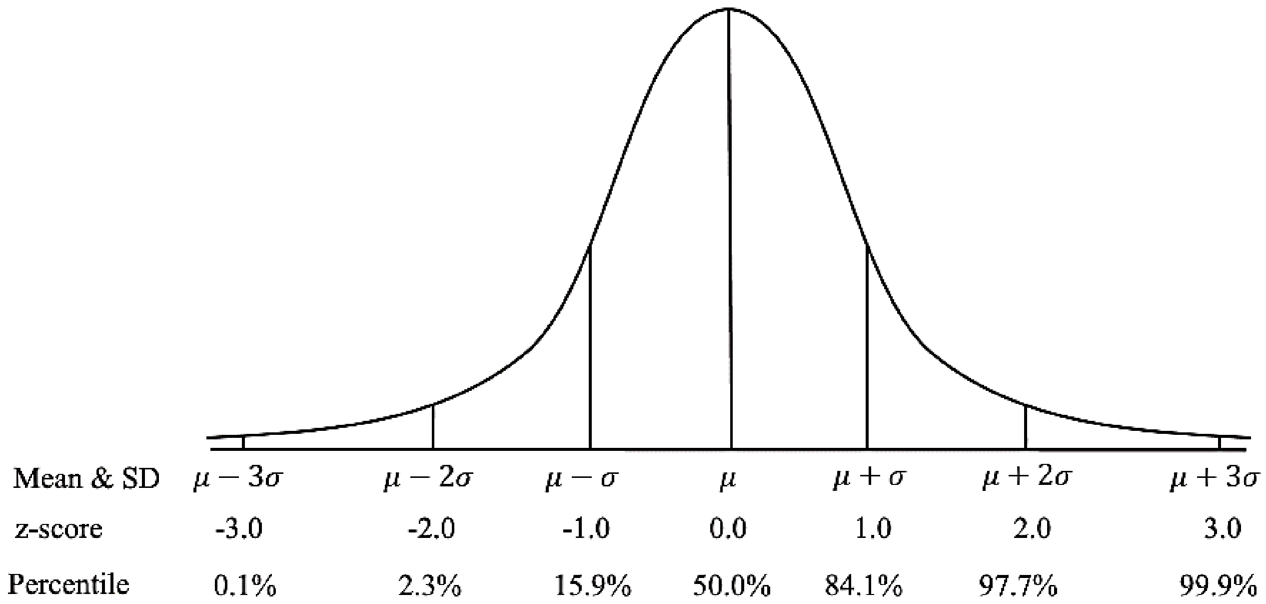
Solution for part (b)

- The score 90 is 2 standard deviations above the mean.
- Therefore, 97.5% of students are below that value and 2.5% are above it.
- Consequently, 2.5/100 * 280 = 7 students are above 90.



Z-Scores Uses in Statistics

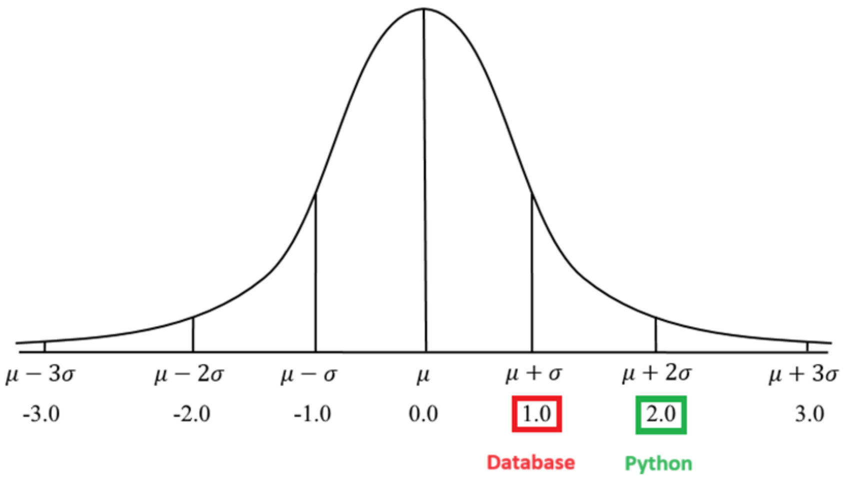
- **Standardization**: Z-scores standardize different distributions to be comparable by converting them to a common scale.
- **Identification of Outliers**: Data points with z-scores less than -2 or greater than 2 are often considered outliers.
- **Compare values** from different populations (e.g. exam results between courses).
- **Probabilities and Percentiles**: Z-scores link data points to their probabilities and percentiles in a normal distribution. E.g., a z-score of 1.96 corresponds to the 97.5th percentile.



Z-Scores Example – Comparing Values

- Khalid achieved 90 in both Python & Database exams. The mean μ for the Python exam was 70 and σ = 10, The μ for the Database exam was 80 and σ = 10.
- Use standard scores to determine where did Khalid achieve a better score.

- Standard score for the Python Exam: (90−70)/10 =2
- Standard score for the Database Exam: (90−80)/10 =1
- Khalid's Z-score for the Python exam is 2, which means his score was 2 standard deviations above the mean.
- His Z-score for the Database exam is 1, indicating his score was 1 standard deviation above the mean.
- Khalid performed better on the Python exam than Database exam since his Z-score is higher for that exam.



How to Find the Right Distribution

- Distfit library tests around 90 different distributions automatically.
- It's like having a statistician who's willing to try every possible option and tell you which one works best.

```
from distfit import distfit
import numpy as np

# Let's say you have some data
my_data = np.random.normal(25, 8, 2000) # 2000 data points

# Set up the distribution fitter
fitter = distfit(method='parametric')

# Let it try different distributions and find the best fit
fitter.fit_transform(my_data)

# See what it found
print("Best fit:", fitter.model['name'])
print("Parameters:", fitter.model['params'])
```

https://medium.com/data-science-collective/how-to-find-the-right-distribution-for-your-data-a-practical-guide-for-non-statistician-with-two-dc2aa0ed707f

Z-Tables

The Z-Tables

- The area under the curve in continuous probability distributions represents the probability of outcome occurrence.
- The Empirical Rule has pre-defined areas for certain Z-Scores.
- What can we do if we need the area between other Z-Scores e.g. +1.25, -2.33?
- It will be hard to compute the integral for different intervals:

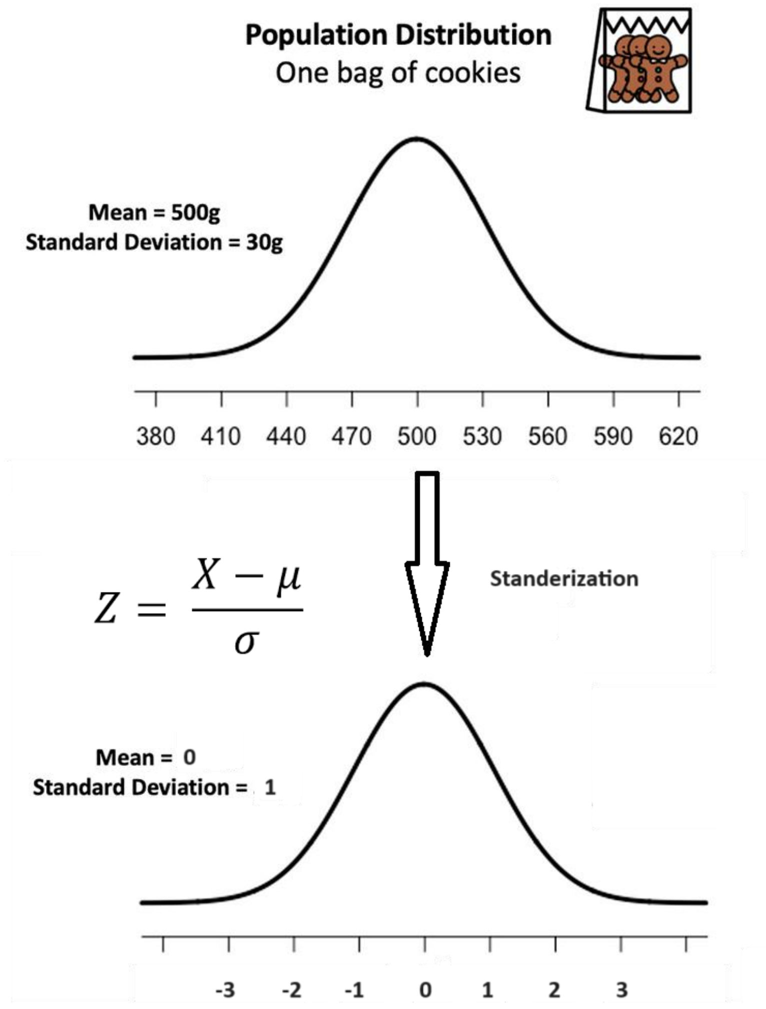
$$P(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

- A simpler solution would be to use a table of pre-computed areas (The Z-Table).
- But we need to compute that area ONCE. Therefore, we need to create a Standard Distribution.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Standard Normal Distribution and Standard Z Scores

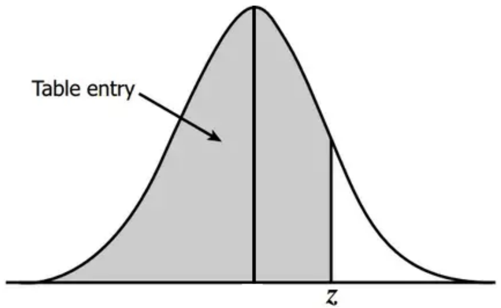
- The standard normal distribution is a special case of the normal distribution where the mean is zero and the standard deviation is 1.
- This distribution is also known as the Z-distribution.
- A value on the standard normal distribution is known as a standard score or a Z-score.
- A standard score represents the number of standard deviations above or below the mean that a specific observation falls.
- For example, a standard score of 1.5 indicates that the observation is 1.5 standard deviations above the mean.
- On the other hand, a negative score represents a value below the average.
- The mean has a Z-score of 0.



https://medium.com/towards-data-science/what-is-p-value-370056b8244d

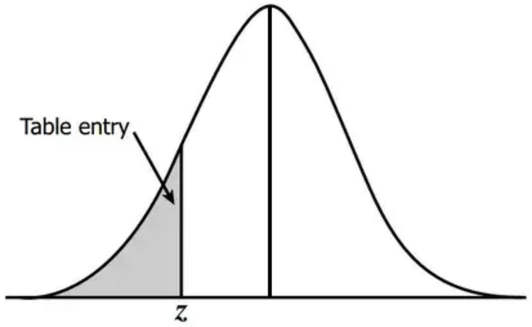
Z-Table for Z Values above 0

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998



Z-Table for Z Values below 0

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641



Z-Tables Example

- The average score for the MID exam in Descriptive Statistics course was 70. The standard Deviation for that exam was 6.
- Knowing that the exam scores are normally distributed, what is the probability of having a score higher than 75 in that exam?

First, we need to find the standard position (z-score) for the 75 score.
Z score for (75) = (75 – 70) / 6 = 5/6 = .833
Using the z table, the closest value for .833 is: 0.7967 which corresponds to 79.67% of scores are below 75.
To find the probability of having a score over 75, we subtract 79.67% out of the total probability of 100%:
100 – 79.67 = 20.33%

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	$\boxed{.7967}$	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Z-Tables using Excel Functions

- The average score for the MID exam in Descriptive Statistics course was 70. The standard Deviation for that exam was 6.
- What is the probability of having a score higher than 90 in that exam?
- We can use Excel NORM.DIST function. This function expects 4 values, the value we want to find its probability, the mean, the standard deviation and we set cumulative = TRUE.
- NORM.DIST(75,70,6,True) = 0.797672
- Therefore, the probability of having a score > 75 = 1 - .797672 = .2023% ~ 20%