Notes on Probability and Statistics

Table of Contents

Further reading: **https://greenteapress.com/thinkstats/thinkstats.pdf**

- **Experiment** is any procedure that can be infinitely repeated and has a well-defined set of possible outcomes, known as the sample space (S).

- **Event (E)** is a subset of the sample space of an experiment. i.e., E ⊆ S

- **Probability (P)** is the likelihood of an event occurring and is given as:.

$$P(A) = \frac{no.\ of\ favourable\ outcomes\ to\ A}{Total\ no.\ of\ possible\ outcomes}$$

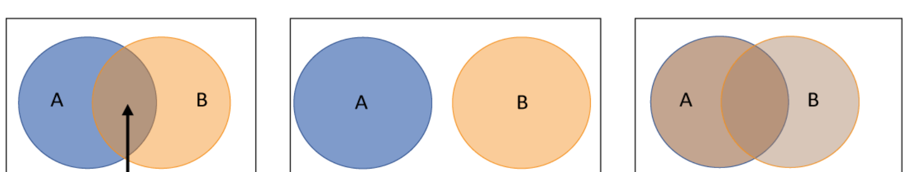
- **Mutually Exclusive Events :**
  If two events are mutually exclusive then the probability of both the events occurring at the same time is equal to zero. i.e. $P(A \cap B) = 0$.
  For example, while tossing of a coin, coming up heads or tails are two mutually Exclusive Events.

- **Mutually Exhaustive Events :** When two events 'A' and 'B' are exhaustive, it means that one of them must occur. i.e. $P(A \cup B) = 1$.
  For example, while tossing of a coin, coming up heads or tails are two mutually Exhaustive Events.

- **Independent Events :** Two events are independent if the occurrence of one does not change the probability of the other occurring.
  If two events 'A' and 'B' are independent, then : $P(A \cap B) = P(A).P(B)$

- **Conditional Probability :** is defined as the likelihood of an event or outcome occurring, based on the occurrence of a previous event or outcome.
  P(Event A will occur given that Event B has already occurred) = $P(A|B) = \frac{P(A \cap B)}{P(B)}$
  Two events A and B are mutually independent if *P(A|B) = P(A) .*
  also, *P(A|B) ≠ P(B|A) .*

- **Basic rules of probability :**
  Addition rule : $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  Multiplication rule : $P(A \cap B) = P(A|B).P(B) = P(B|A).P(A)$
  Complement rule : $P(A^c) = 1 - P(A)$



$A \cap B$ A and B — $A \cap B = \varnothing$ A and B is the empty set — $A \cup B$ A or B

- **Bayes Theorem :** Conditional probability of each of a set of possible causes (B), given an observed outcome (A) is given as:

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

E.g. A man speaks truth ¾ times. He draws a card and reports 'King'. What is the probability of it actually being a king?

P(Man speak truth)= P(T) =¾, P(drawing a king)=P(K)= 4/52 = ¼

P(drawing a king given that the man speaks truth) = P(K|T)

$$P(K|T) = \frac{P(T|K).p(K)}{P(T)} = \frac{(3/4).(1/13)}{(1/13).(3/4) + (12/13).(1/4)} = 0.2$$

- **Types of a Random variable :**
  Discrete random variable: A random variable with a finite or countable number of possible values. e.g. random variable of the outcome when a dice is thrown. It can take integer values in range 1 to 6.
  Continuous random variable: A random variable that can take an infinite number of possible values. E.g. a random variable representing the height of students in a class.
- **Probability Distribution Function (PDF)** is a function that is used to give the probability of all the possible values that a random variable can take.
  If the random variable is discrete, then it is called **Probability Mass Function**.
  For the continuous random variable, it is called **Probability Density Function**.
- **Cumulative Distribution Function (CDF):** returns the probability that a random variable will take a value less than or equal to x.
- **Expectation** of a discrete random variable(X) having a Probability mass function P(x) is the weighted average of possible values that X can take.

  i.e. $E(X) = \sum_{1}^{n} x_i.P(x_i)$

  Note :For a uniformly distributed random variable, Expectation is equal to the mean.

  Properties of Expectation :

$$E(aX) = a.E(X)$$
$$E(X + b) = E(X) + b$$
$$E(aX + b) = a.E(X) + b$$

- **Variance** is a statistical measurement that is used to determine the spread of numbers in a data set with respect to the average value or the mean. It is given as :

$$Var(X) = \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Where, σ is the **standard deviation**. Therefore,

$$Variance = (Standard\ Deviation)^2$$

For a random variable *X*, **Variance** can be given as :

$$Var(X) = E(X^2) - [E(X)]^2$$

Properties of variance:

1. $Var(k.X) = k^2.Var(X)$

2. Assuming that the samples were collected independently.

   $Var(X_1 + X_2 + X_3 + ...) = Var(X_1) + Var(X_2) + Var(X_3) + ...$

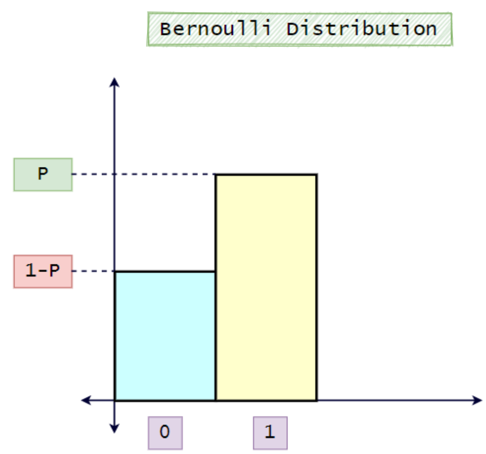
3. $Var(X + c) = Var(X)$

- **Covariance** is the variance of two quantities with respect to each other. It is a measure of how much two random variables vary together.

$$Cov(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

- **Discrete probability distributions:**

1. **Bernoulli distribution:**The distribution of a random variable which takes a binary, boolean output: 1 with probability p, and 0 with probability (1-p).
   Let X follows the Bernoulli distribution, i.e. $X \sim Bern(p)$
   then,
   $$P(X = x) = \begin{cases} 1 - p & , \quad x = 0 \\ p & , \quad x = 1 \end{cases}$$
   and
   $$E(X) = p$$
   $$Var(X) = p(1 - p)$$



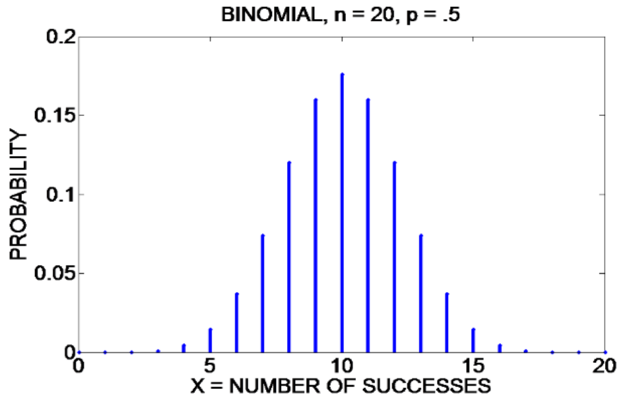
2. **Binomial distribution:** It is the probabiity distribution of getting *x* successes in *n* Bernoulli trials.
   Let X follows the Binomial distribution, i.e. $X \sim B(n, p)$

   where, n = number of trials , p = probability of success

   then,
   $$P(X = x) = {}^{n}C_{x}.p^{x}.(1 - p)^{n-x}$$

   and
   $$E(X) = n.p$$
   $$Var(X) = n.p.(1 - p)$$

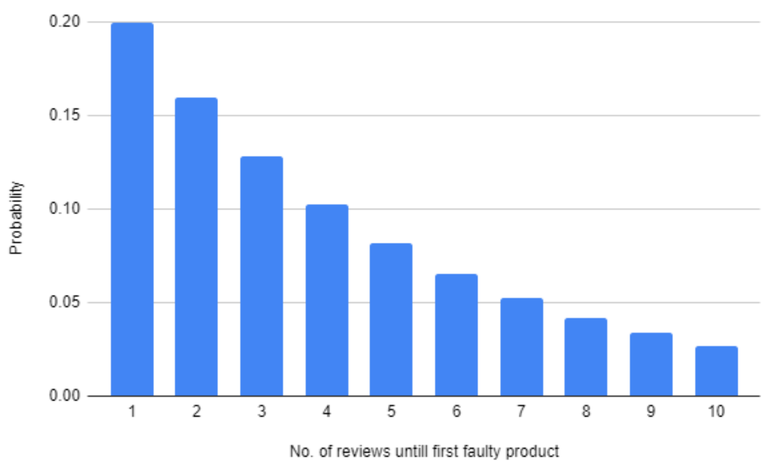
E.g. Probability Distribution of number of heads in 20 coin flips.



3. **Geometric distribution:** It is the probability distribution of number of Bernoulli trials needed to get one success.

   Its Probability mass function is given as : $P(k) = (1-p)^{k-1}.p$

   E.g. The probability of getting a faulty product after reviewing *k* non-faulty products.

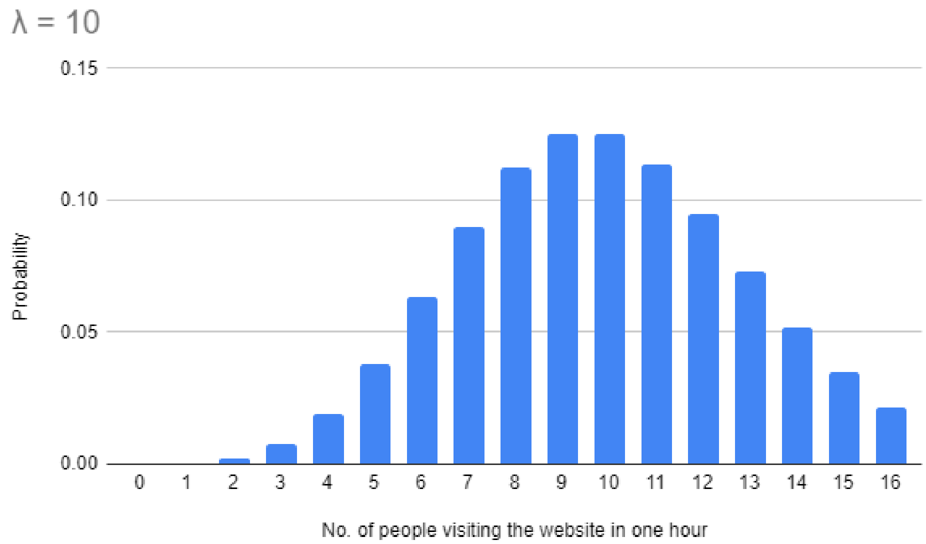


4. **Poisson distribution:** It gives the probability of an event happening a certain number of times (x) within a given interval.

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

   Expectation, mean and variance are all equal to λ.

For example, the Probability distribution of no. of people visiting a website in one hour.



- **Continuous Probability distributions:**

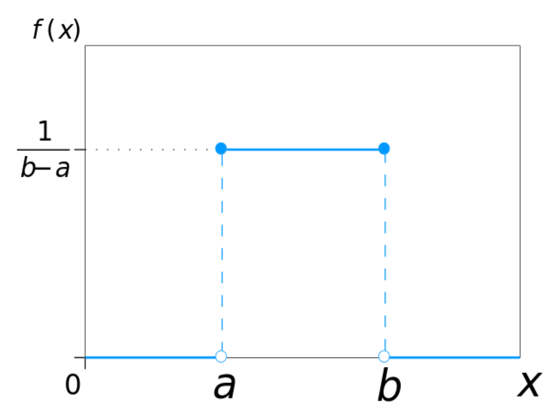
1. **Uniform distribution :** uniform distribution refers to a type of probability distribution in which all outcomes are equally likely.

   The PDF of a continuous uniform distribution is given as:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b, \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$$

Mean = $E(X) = \dfrac{a+b}{2}$

Variance $(\sigma^2) = \dfrac{(b-a)^2}{12}$



2. **Normal distribution:** This distribution is very common in nature and has a symmetric bell shaped curve. It is also called as Gaussian distribution.

   PDF of Normal distribution is given as

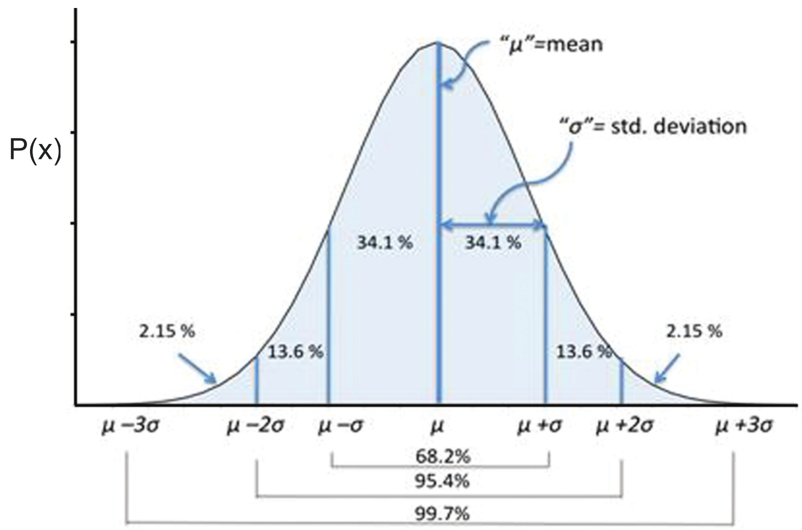
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Where μ is mean and σ is standard deviation.

Properties of Normal distribution :

- ➢ Symmetric about mean and has a bell shaped distribution.
- ➢ Mean (μ) = mode = median
- ➢ **Empirical rule :**

  Around 68% of values are within 1 standard deviation from the mean.
  Around 95% of values are within 2 standard deviations from the mean.
  Around 99.7% of values are within 3 standard deviations from the mean.



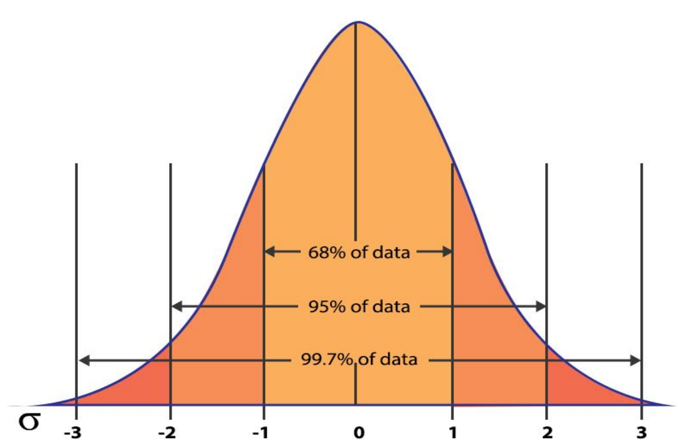
3. **Standard Normal distribution:** It is a special case of Normal distribution when the mean is 0 and the standard deviation is 1.

   Any normal distribution can be standardized by converting its values into z-scores.

   z- score of a value *x* in normal distribution is given by :

$$z = \frac{x - \mu}{\sigma}$$

   z-score tells us about the distance from mean in terms of standard deviation.

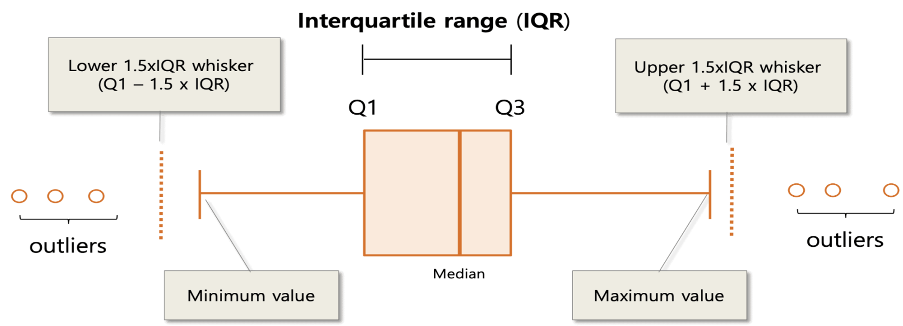


- **Outlier :** Any data point which is far away from the rest of the data points.
- **Inter Quartile Range (IQR)** is a measure of the middle 50% of a data set.

$$IQR = Q3 - Q1$$

  where, Q3 is the third quantile value and Q1 is the first quantile value of the data.
- IQR is used for the purpose of **detecting outliers** in the data**.**
  All the points having value greater than ($Q3 + 1.5 \times IQR$) **or** less than ($Q1 - 1.5 \times IQR$) are considered to be the outliers.

- **Boxplot** is a standardized way of displaying the distribution of data based on a five-number summary. These are "minimum", first quartile [Q1], median, third quartile [Q3], and "maximum".



- A **Sample** is an analytic subset of a larger population. If the sample is well selected, the sample will be generalizable to the population.
- **Sample vs population :**

Population mean = $\mu = \frac{1}{N}\sum_{1}^{N} x_i$, *Population Variance* = $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$

Sample mean = $\overline{x} = \frac{1}{n}\sum_{1}^{n} x_i$, *Sample Variance* = $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$

The reason we use *(n-1)* rather than *n* in the denominator is so that the sample variance will be an unbiased estimator of the population variance.
As the sample size is increased, the sample mean gets closer to the population mean and the variance in the means of samples decreases.

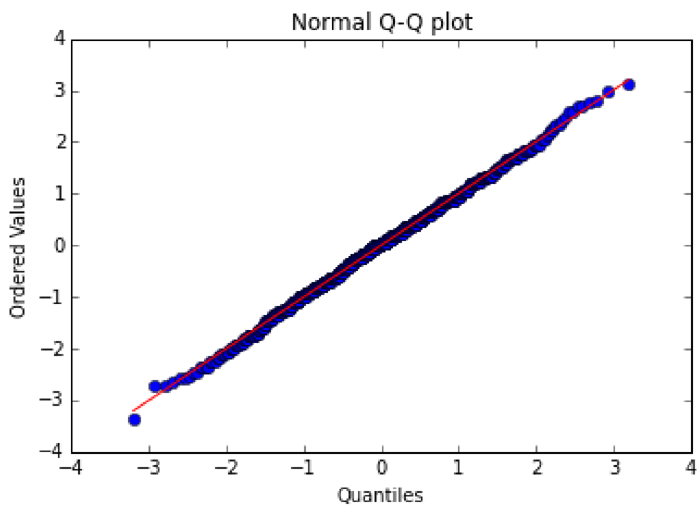
- **Standard error** is the spread of sample means around the population mean.

$$Standard\ error = \frac{\sigma}{\sqrt{n}}$$

As the sample size increases, Standard error decreases.

- **Central Limit Theorem (CLT)** states that 'the distribution of sample means is Gaussian, no matter what the shape of the original distribution is.
  The assumption of CLT is that the population mean and population standard deviation should be finite and sample size should be >=30 (commonly accepted number is 30).

- **Quantile-Quantile (Q-Q) plot:** is a graphical way for comparing two probability distributions by plotting their quantiles against each other.
  If the two distributions being compared are similar, the points in the Q–Q plot will approximately lie on the line *y* = *x*. Given below is the Q-Q plot of two similar distributions.



- **Hypothesis testing** is a method of statistical inference used to decide whether the data at hand sufficiently support a particular hypothesis.
  A hypothesis test is the means by which we generate a **test statistic** that directs us to either reject or not reject the null hypothesis.
  If **p value** is lower than **significance level**; then we reject the **null hypothesis**, else we fail to reject the null hypothesis.

  For e.g., Testing whether a coin is fair or not. Say we conducted our experiment and got 65 heads out of 100 tosses.

  The **Null hypothesis** **(**H_0**)** represents the assumption that is made about the data sample whereas the **alternative hypothesis** **(**H_a**)** represents a counterpoint .

  Null hypothesis **(**H_0**) :** the coin is fair (p=0.5)

  Alternative hypothesis **(**H_a**) :** the coin is not fair (p>0.5)

  By observing the alternative hypothesis we can say that it is a right tailed test.

Test statistic(T) is a sensible way to measure the outcome of the experiment which can help us answer our question. i.e. no. of heads in 100 tosses (experiment) which follows a binomial distribution.

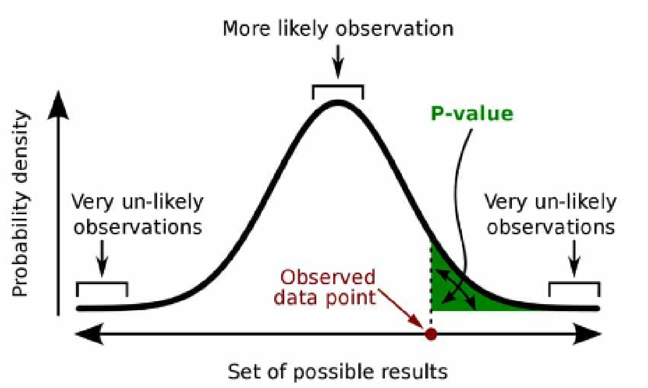
p value : P(observing Test statistic as extreme or more than T_{obs}= 65 given that null hypothesis is true)

$= P(T \geq (T_{obs} = 65)|H_0)$. $= 1 - P(T \leq (T_{obs} = 65)|H_0)$

$= 1 - CDF(T = 64)$ where, $CDF(T = 64) = P(T \leq 64)$

$= 0.0017$

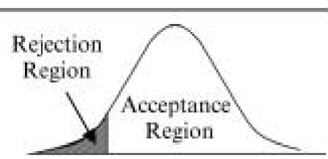
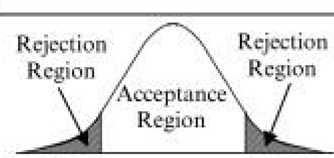
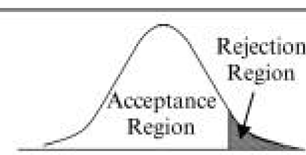
Since p-value is less than 5% (Significance level), we reject the null hypothesis and accept the alternative hypothesis.



Critical value is a cut-off value that is used to mark the start of a region where the test statistic, obtained in hypothesis testing, is unlikely to fall in.

- Types of Hypothesis testing :
  One-tailed test : the critical distribution area is one-sided, meaning the test sample is either greater or lesser than a specific value. E.g The above example of testing whether a coin is fair or not is a one tailed test. To be specific, a right tailed test.

  Two-tailed test : the critical distribution area is two-sided.

One-Tailed Test (Left Tail)	Two-Tailed Test	One-Tailed Test (Right Tail)
$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X < \mu_0$	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X \neq \mu_0$	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X > \mu_0$
 Rejection Region; Acceptance Region	 Rejection Region; Acceptance Region; Rejection Region	 Acceptance Region; Rejection Region

- **Types of errors:**
  1. **Type 1 error (α):** occurs during Hypothesis testing when a null hypothesis is rejected, even when it is accurate.It means concluding that results are statistically significant when, in reality, they came about purely by chance. Probability of Type 1 error is equal to the significance level.

  2. **Type 2 error (β):** occurs when one fails to reject a null hypothesis which is really false.It is also called as **Power of test**.

- **Framework for Hypothesis testing :**
  1. Define the experiment and a sensible test statistic variable.
  2. Define the null hypothesis and alternate hypothesis.
  3. Decide a test statistic and a corresponding distribution.
  4. According to alternate hypothesis, determine whether the test should be left tailed, right tailed or two tailed.
  5. In the next step, the p - value is determined for the test statistic being as extreme as the observed test statistic.
  6. In the next step, we choose the significance level, mostly the default significance level of 5 % is chosen.
  7. Finally, we compare the obtained p - value with the chosen significance level and decide whether to accept the null hypothesis or reject it.

- **Z-test :** is a hypothesis test which uses Standard normal distribution as baseline.

  Conditions of Z-test:
  - ➢ we assume that the population has finite mean and standard deviation.(Parametric method)
  - ➢ The second condition is that either we should be aware of the standard deviation of the population of the given samples or we should estimate them well when the size of the samples are not too small.

  One sample Z-test : One sample z-test is used to determine whether a particular population parameter, which is mostly mean, significantly different from an assumed value.

  Framework for one sample z-test :
  1. Define Null and alternative hypothesis (H_0 and H_a) and collect sample data.
  2. Get the baseline gaussian distribution under null hypothesis.

$$N\left(\mu, \frac{s}{\sqrt{n}}\right)$$

  3. Determine the type of test. (Left tailed/ right tailed/ two tailed)
  4. Decide the significance level (α) .
  5. From z-table, calculate critical value in terms of z values. i.e. z_c .
  6. Convert critical values (z) in terms of test statistic.

$$c = \mu \pm z_c\left(\frac{s}{\sqrt{n}}\right)$$

7. Compare observed mean $\overline{x}$ with critical value c.
8. Accept or reject the null hypothesis.

Example :

30 years ago, the avg, weight of elephants was 4,000kg. We sampled 25 asian elephants and found the average weight to be 3850kg. Can you conclude if the elephants size has shrunk with time?
Take α (significance level)= 0.05 and standard deviation of sample (s) =500 kg.

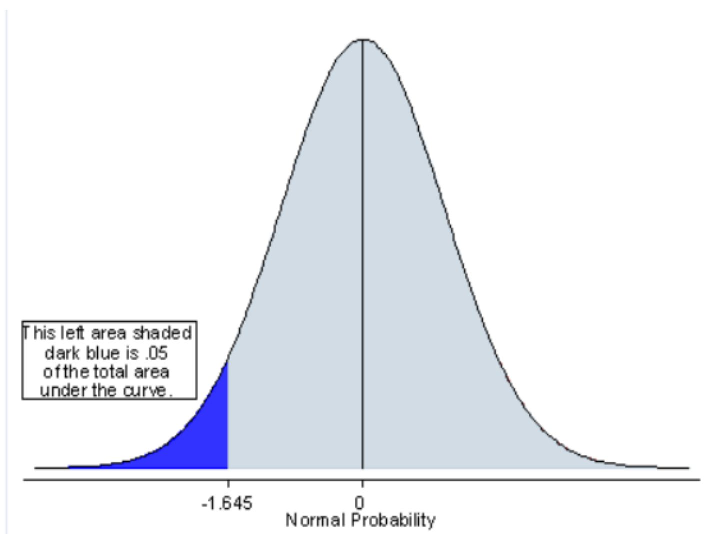
Given : $\overline{x}$ = 3850, s = 500 and n=25

H_0 : μ = 4000 , H_a: μ < 4000

The gaussian distribution under null hypotheses is :

$$N\left(\mu, \frac{s}{\sqrt{n}}\right) = N\left(4000, \frac{500}{\sqrt{25}}\right) = N(4000\ ,\ 100)$$

The test to be performed is a left tailed test.



Now, determining the critical value,

$$x_c = c = \mu + z_c \cdot \left(\frac{s}{\sqrt{n}}\right)$$

= 4000 +(-1.65) (100) = 3835

Now, since $x_c > \overline{x}$, therefore, p-value < α

Thus, we reject the null hypothesis.

Two sample Z-test : A two sample z-test is used to compare the means of two populations.

Example :
Lets take the example of mean recovery time of two different populations who were given two different drugs.Let there be a sample of 100 observations of this variable subjected to the medicine M_1 and a sample of 90 observations of the same variable subjected to the medicine M_2.
Let μ_1 and μ_2 be the population mean recovery times for both the experiments respectively.
We require to know if $\mu_1 = \mu_2$.

Lets first define the null and alternative hypothesis:

$H_0 : \mu_1 = \mu_2$,$H_a : \mu_1 \neq \mu_2$

The test statistic in z-test is defined as:

$$T = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

where, $n_1 = 100$ and $n_2 = 90$
the population standard deviations σ_1 and σ_2 can be considered as the respective sample standard deviations, s_1 and s_2 as the samples are large enough (>30).

Therefore, test statistic can be written as:

$$T = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

It is proved that this test statistic variable (T) follows the normal distribution with mean and standard deviation as 0 and 1 respectively, i.e, T ~ Z(0,1).

for the next step let's say that the observed test statistic value (T_{obs}) is 3.2 using the values of the observed $\overline{x}_1$, $\overline{x}_2$ and σ_1 , σ_2 from the set of observations of both the samples.
Here, the alternate hypothesis to be true, either the test statistic value should be a large positive number or large negative number and thus, two tailed tests will be required for this problem.

For the next step, p - value will be estimated. As per 68 - 95 - 99 rule for the Normal Distribution, we know that 1 % of the values of the test statistic will lie for the value of the test statistic greater than 3 times the standard deviation of the distribution, i.e, (3*1) = 3.

since 3.2 $>$ 3, hence, less than 1 % of the values of the test statistic will lie after 3.2.
Thus, the p - value will be less than the significance level, i.e, half of 5 % as it is a two tailed test and hence, the null hypothesis will be rejected.

- **T-test** is a type of hypothesis testing that is used to compare the means of two groups.
  The test statistic follows a **T-distribution** in this test.
  It is used when the sample size is too small.(n< 30).
  For e.g. It is used In extreme medical experiments where the risk is minimized using the less number of patients or subjects.

z-test cannot be used when the sample size is too small as we can't estimate the population standard deviations for both these samples.

The test statistic (T_t) for the T - test is given as:

$$T_t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where $\bar{x}_1$ and $\bar{x}_2$ are the observed means of the two samples,
s_1 and s_2 are the respective standard deviations and n_1 and n_2 are the respective sample sizes.

The T - distribution under the null hypothesis is defined using a parameter *v* which is $(n_1 + n_2 - 2)$. This parameter v is also called **degrees of freedom**. As the value of the parameter *v* decreases the tails of the curves become larger.

Assumptions of t-test :

1. The population means and standard deviations for both samples have to be finite.
2. The observations in the samples should be random and independent.
3. If the population standard deviations of the samples are unknown, they can't be estimated as the sizes of the samples are small.

- **ANOVA (Analysis of variance):** is a type of statistical test used to determine if there is a statistically significant difference between **two or more** categorical groups by testing for differences of means using variance

  The test statistic for ANOVA is denoted by **f** which is given as :

$$f = \frac{MSB}{MSW}$$

where, MSB = mean of the squared distances between the groups
and MSW = the mean of the squared distances within the groups.

The test statistic f follows a distribution referred to as the F distribution which is represented by two parameters (k-1) and (n-k).
where k is the number of groups and n is the total sample size.

Assumptions of ANOVA: There are certain assumptions made for ANOVA to be valid as it is a parametric test. These are :

1. The distributions of data of each group should follow the Gaussian Distribution.
2. The variance of each group should be the same or close to each other.
3. The total n observations present should be independent of each other.

Example of ANOVA test :

Lets take the example of recovery times of covid patients.

Suppose we have an observed data where we are inspecting *k* drugs.

For each drug, a group of unique *m* patients are tested.

Let's say *n* is the total number of patients which will be *(m*k)*.

Now, a matrix $X_{i\,j}$ is given in which i denotes the group whose range will be from 1 to k and j denotes the column of patients whose range will be from 1 to m.

The element $X_{i\,j}$ of the matrix $X_{i\,j}$ in the matrix denotes the recovery time of the j^{th} patient in the i^{th} group.

The task of the problem is whether all these medicines or drugs have the same average recovery times.

Framework for ANOVA:

1. Define null and alternative hypothesis. Here,

   H_0 : There is no difference between group means.

   H_a : There is some difference b/w group means.

2. Let $\overline{X}_1$, $\overline{X}_2$, …… $\overline{X}_k$ be the mean recovery times of the m patients of the groups 1,2 ….., k respectively and let $\overline{X}$ be the overall mean recovery times of all the n patients.
   Caculate MSB and MSW.

   $$MSB = \frac{\sum_{i=1}^{k} n_i(\overline{X_i} - \overline{X})^2}{k-1} \qquad MSW = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m}(X_{ij} - \overline{X_i})^2}{n-k}$$

3. Calculate test statistic **f** which is defined as the ratio of MSB and MSW.

4. The p - value corresponding to f is estimated using F-distribution and compared with the significance level to determine whether the null hypothesis has to be accepted or rejected.

- **KS (Kolmogorov - Smirnov) test:** is used for determining whether the distributions of two samples are same or not. Here, the test statistic T_{KS} follows a distribution which is called the **Kolmogorov Distribution.**

  For example, suppose we have a sample X with *n* number of observations having mean and standard deviation equal to $\overline{x}$ and s respectively.

  Now let's consider another sample Y with *m* number of observation and follows a Gaussian distribution.

  We need to know whether the sample X follows the same distribution as sample Y (Gaussian).

  Framework for KS Test :

1. Define Null and alternate hypothesis.

   H_0 : X ≈ Y (X and Y have same distribution) ,

   H_a : X ≠ Y (X and Y have different distributions)

2. Define Test statistics : The KS test uses the principle of comparing the CDFs of the two samples for defining the test statistic.Therefore, the test statistic **(T_{KS})** is the maximum absolute value of the difference or gap in the CDFs of the two samples X and Y, which is also referred to as the **supremum** of the same absolute difference.

   The test statistic T_{KS} follows a distribution which is called the Kolmogorov Distribution

3. Determine the type of test: For the alternate hypothesis to be true, T_{KS} has to be a large positive number. Thus, it is a right-tailed test.

4. As the final step, we determine the p-value (i.e. the area under the curve of the PDF of this distribution plot after the value T_{obs}.) for the observed test statistic under the null hypothesis and compare it with the significance level after which we conclude whether the null hypothesis should be accepted or rejected.

   The KS test is a non - parametric test as there are no assumptions made for the population.

- **Correlation** is the degree of mutual relationship between two variables.

  There are two main types of correlation coefficients:

  1. Pearson correlation coefficient

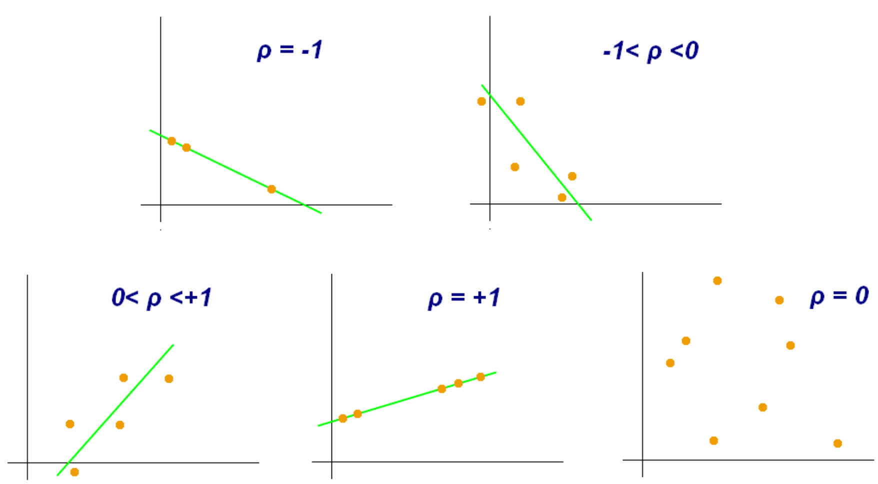
  2. Spearman rank Correlation Coefficient

- **Pearson correlation coefficient (PCC)** for two variables X & Y is given as the covariance divided by the product of the standard deviations of the variables.

  i.e.

$$\rho_{xy} = \frac{Cov(X,Y)}{\sigma_x . \sigma_y}$$

It takes into account the standard deviations.



Limitation of PCC is that it only captures the linear relationship between the variables. It fails to capture the non-linear patterns.

- **Spearman Rank Correlation Coefficient (SRCC) :** It is a statistical measure of the strength of a monotonic relationship between paired data.SRCC captures the monotonicity of the variables rather than the linearity.

  Calculation of SRCC :

  1. Each variable (X and Y) is sorted to assign the ranks for the values of each variable.
  2. Based on these rank values derived, we form an another matrix, let's say D' for R_X and R_Y which represents the ranks of the values of the variables X and Y respectively.
  3. Calculate the difference between the ranks (d) and the square value of d.
  4. Add all your d square values and insert into the formula given below :

$$\rho = 1 - \frac{6\Sigma\, d_i^2}{n(n^2 - 1)}$$

For example, Consider the score of 5 students in Maths and Science that are mentioned in the table.

Students	Maths	Science
A	35	24
B	20	35
C	49	39
D	44	48
E	30	45

After sorting and ranking the data we get the following table :

	X	R_x	Y	R_y		
Students	Maths	Rank	Science	Rank	d	d square
A	35	3	24	5	2	4
B	20	5	35	4	1	1
C	49	1	39	3	2	4
D	44	2	48	1	1	1
E	30	4	45	2	2	4
						14

Therefore,

$$\rho = 1 - (6 * 14)/(5)(24) = 0.3$$

<u>**Combinatorics**</u>

- **Number of ways to arrange '*n'* elements** is given by :

  Factorial(n) = n! = (n).(n-1).(n-2).(n-3) 3.2.1

- A **permutation** is the choice of '*k'* objects from a set of '*n'* objects without replacement and where the order matters.

  For example, the number of ways to arrange '*n*' distinct objects in '*k*' places

$$^{n}\mathrm{P}_{k} = \frac{n!}{(n-k)!}$$

- A **combination** is the choice of '*k*' things from a set of '*n'* things without replacement, where order does not matter.

  For example, the number of ways to select *k* students from a class of *n* students is :

$$^{n}\mathrm{C}_{k} = \frac{n!}{(n-k)!(k)!}$$

- **Some Properties :**

$$^{n}\mathrm{C}_{k} = {}^{n}\mathrm{C}_{(n-k)}$$

$$^{n}\mathrm{C}_{k} = {}^{n-1}\mathrm{C}_{k-1} + {}^{n-1}\mathrm{C}_{k}$$

$$^{n}\mathrm{C}_{0} + {}^{n}\mathrm{C}_{1} + {}^{n}\mathrm{C}_{2} + \ldots\ldots + {}^{n}\mathrm{C}_{n} = 2^{n}$$

$$0! = 1$$

$$1! = 1$$

$$n! = (n-1)!$$