Maths final notes

Statistics: a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data.

Random variable is a variable whose value is unknown or a function that assigns values to each of an experiment's outcomes.

Joint PD: is a statistical measure that is used to calculate the probability of 2 events occurring together at the same time

Marginal PD: gives the probability of various values of the variable in the subset without reference to the value of the other variable

Probability mass function (PMF) is a function that gives the probability that a discrete random variable is exactly equal to some value. Sometimes it is also known as the discrete density function

probability density function (PDF), or density of a continuous random variable, is used to specify the probability of the random variable falling within a particular range of values, This probability is given by the integral of this variable's PDF over that range

Difference between Mutually exclusive and independent events	
Mutually exclusive events	**Independent events**
When the occurrence is not simultaneous for two events then they are termed as Mutually exclusive events.	When the occurence of one event does not control the happening of the other event then it is termed as an independent event.
The non-occurrence of an event will end up in the occurrence of an event.	There is no influence of an occurrence with another and they are independent of each other.

The mathematical formula for mutually exclusive events can be represented as P(X and Y) = 0	The mathematical formula for independent events can be defined as P(X and Y) = P(X) P(Y)
The sets will not overlap in the case of mutually exclusive events.	The sets will overlap in the case of independent events.

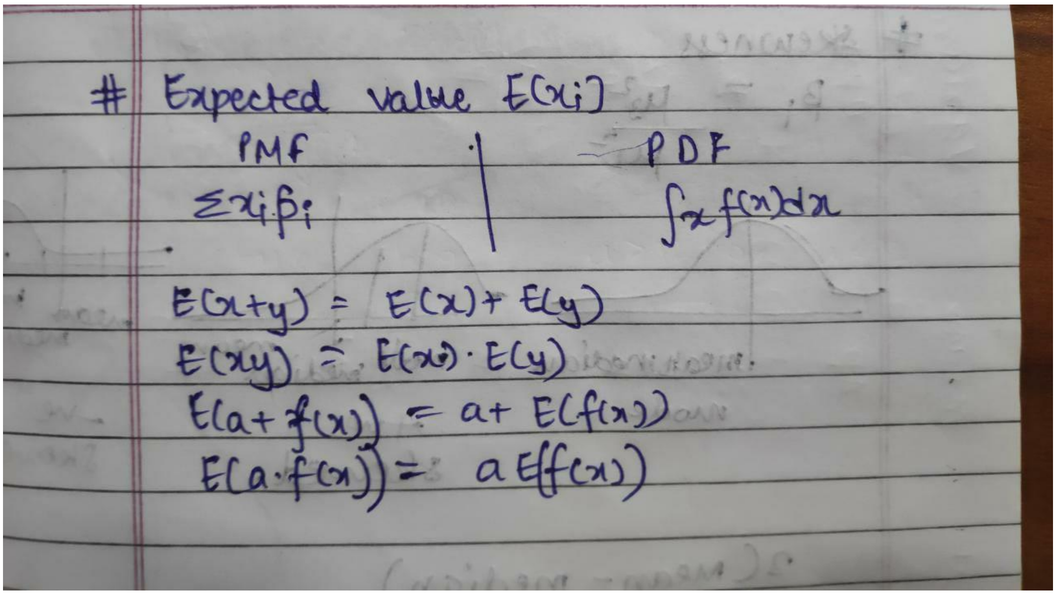
Exhaustive events : In probability, a set of events is collectively exhaustive if they cover all of the probability space:

Discrete random variable: A discrete random variable has a countable number of possible values. The probability of each value of a discrete random variable is between 0 and 1, and the sum of all the probabilities is equal to 1. It has PMF(probability mass function)

Continous random variable: Continuous random variables, on the other hand, take on values that vary continuously within one or more real intervals, and have a cumulative distribution function (CDF) that is absolutely continuous. It has pdf (probabilty distribution function)

Mathematical expectation: Mathematical expectation, also known as the expected value, which is the summation of all possible values from a random variable.

It is also known as the product of the probability of an event occurring, denoted by P(x), and the value corresponding with the actually observed occurrence of the event.



Expected value $E(x_i)$

PMF	PDF
$\Sigma x_i p_i$	$\int x f(x) dx$

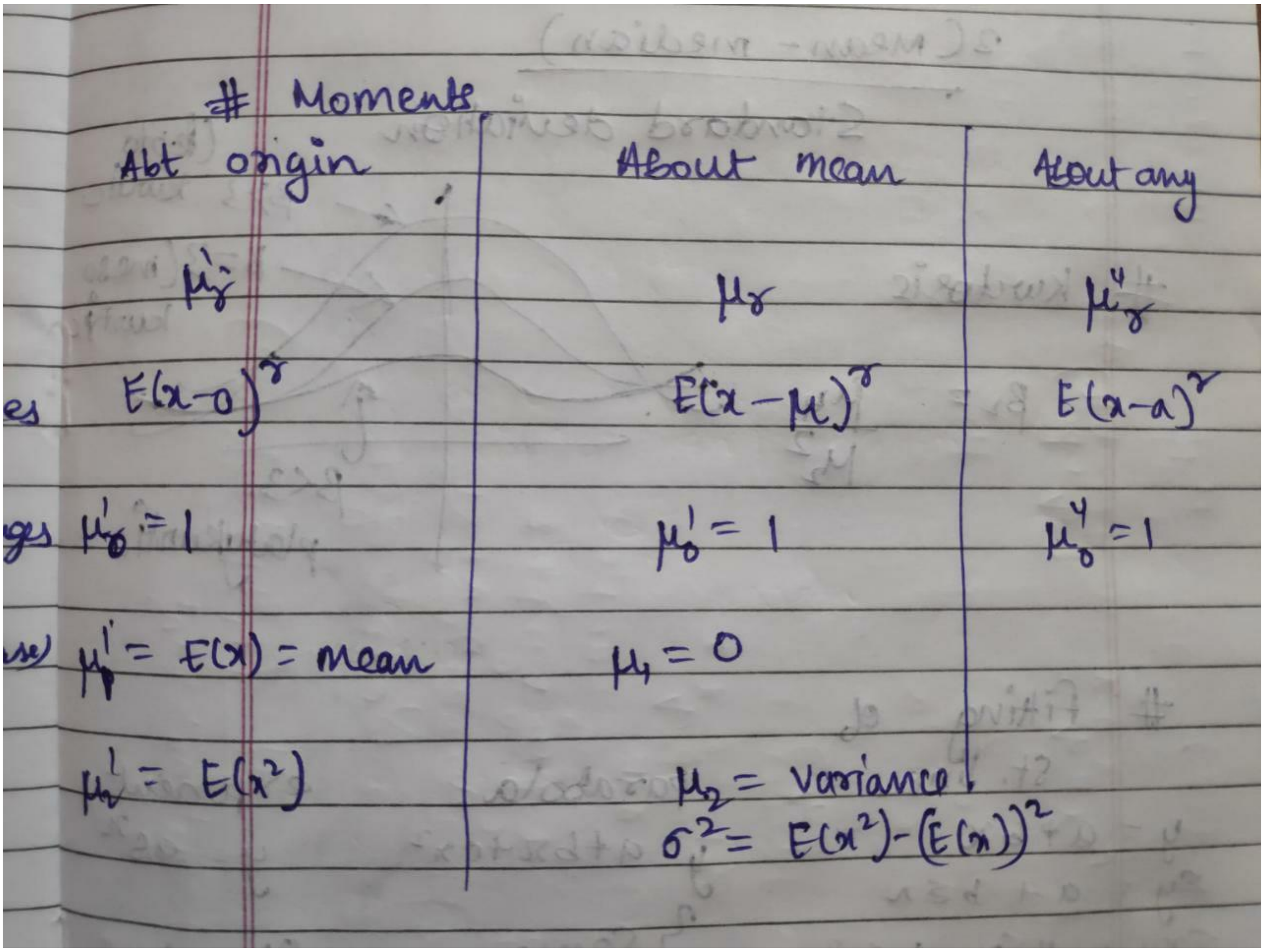
$E(x+y) = E(x) + E(y)$

$E(xy) = E(x) \cdot E(y)$

$E(a+f(x)) = a + E(f(x))$

$E(a \cdot f(x)) = a E(f(x))$

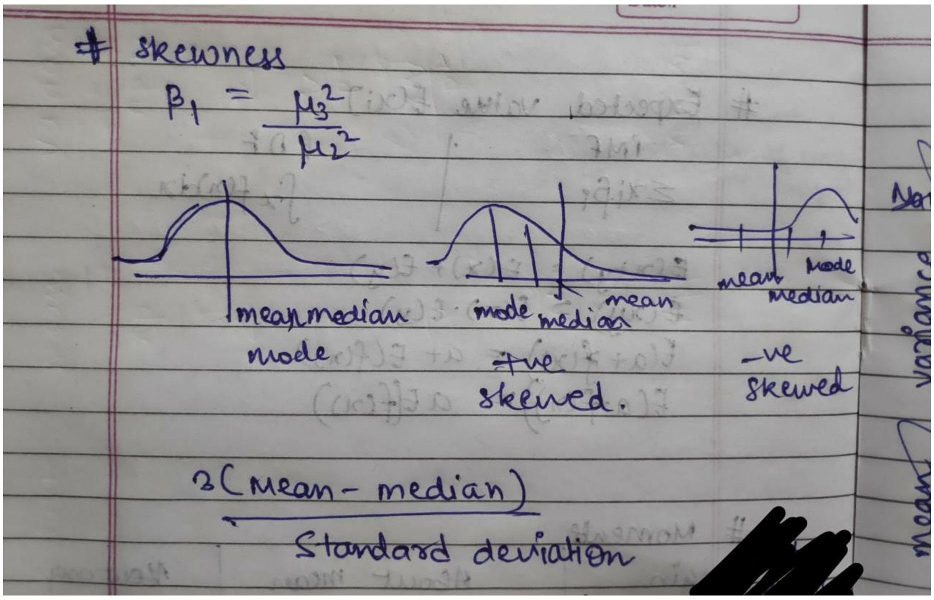
Moments



Moments

Abt origin	About mean	About any
μ_r'	μ_r	μ'^{4}_{r}
$E(x-0)^r$	$E(x-\mu)^r$	$E(x-a)^r$
$\mu_0'=1$	$\mu_0'=1$	$\mu'^{4}_{0}=1$
$\mu_1'=E(x)=$ mean	$\mu_1=0$	
$\mu_2'=E(x^2)$	$\mu_2=$ variance $\sigma^2=E(x^2)-(E(x))^2$	

Skewness: Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point.



skewness

$$\beta_1 = \frac{\mu_3^2}{\mu_2^2}$$

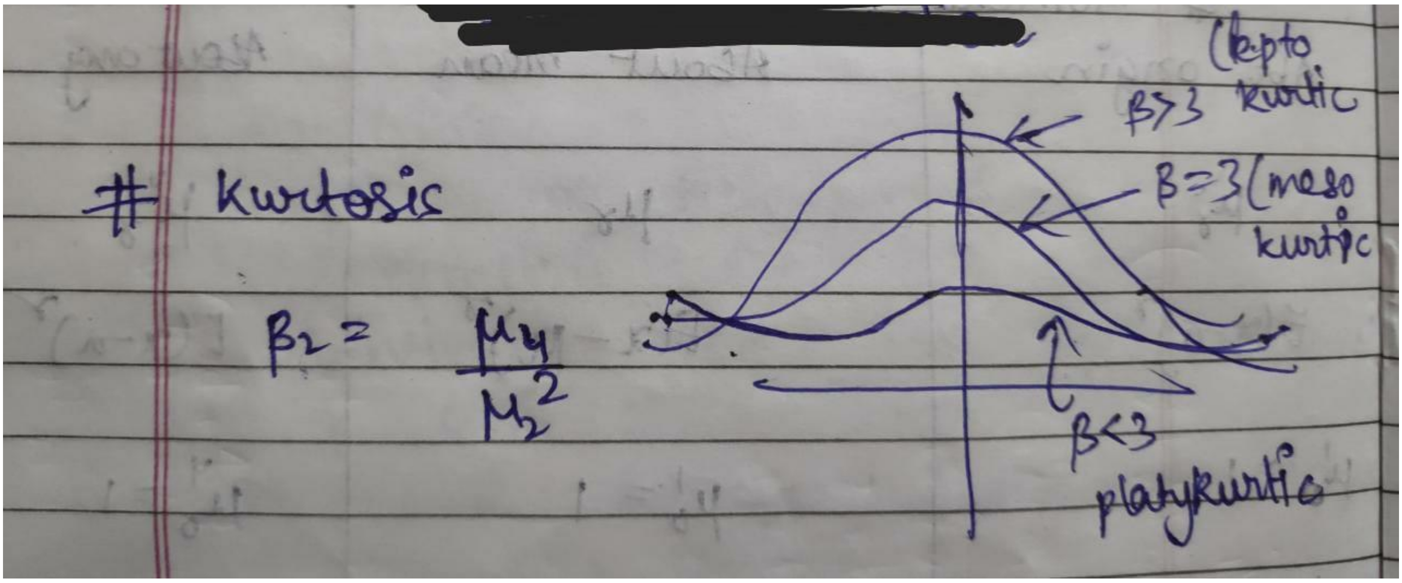
mean median
mode
mode
median
mean
+ve
skewed.
mean
mode
median
-ve
skewed
variance
mean

$$\frac{3(\text{Mean} - \text{median})}{\text{Standard deviation}}$$

[redacted]

Kurtosis:

Kurtosis is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution. That is, data sets with high kurtosis tend to have heavy tails, or outliers. Data sets with low kurtosis tend to have light tails, or lack of outliers. A uniform distribution would be the extreme case.



[redacted]

kurtosis

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

(lepto
β>3 kurtic
β=3 (meso
kurtic
β<3
platykurtic

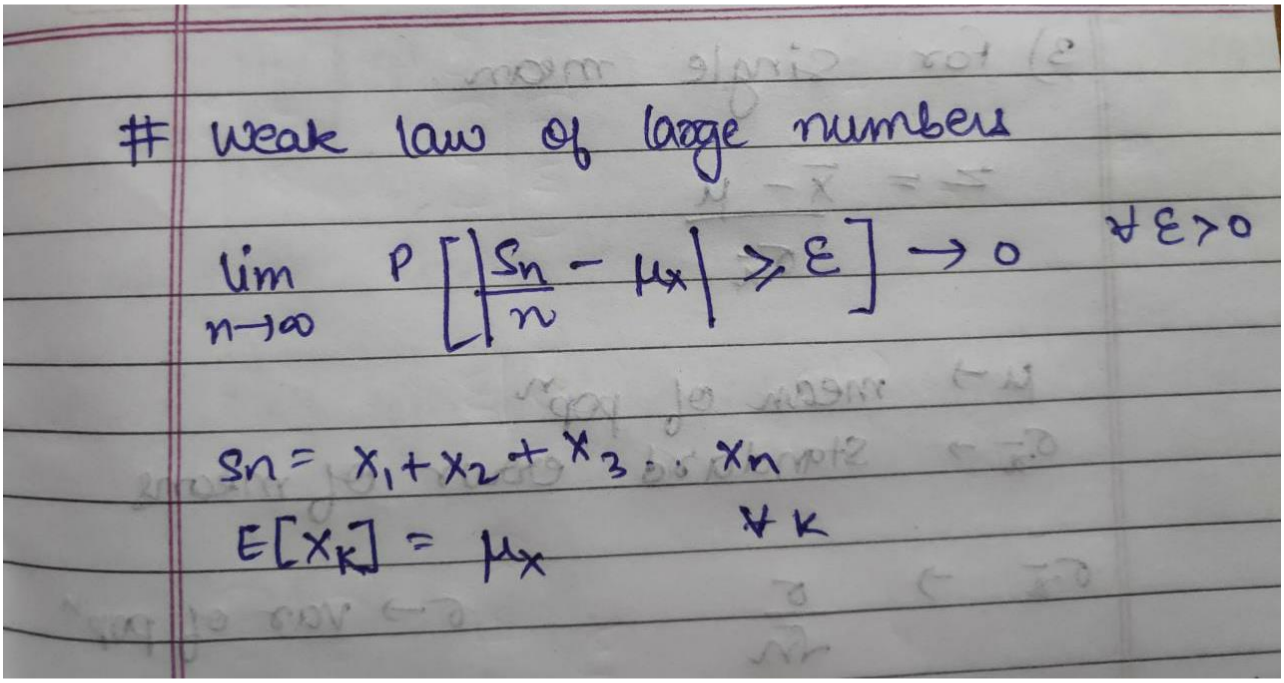
Change of origin in mgf

Changes the mean but variance remains same

Change in scale in mgf

Changes the mean and variance

Weak law of large numbers:



weak law of large numbers

$$\lim_{n \to \infty} P\left[\left|\frac{S_n}{n} - \mu_x\right| \geq \varepsilon\right] \to 0 \quad \forall \varepsilon > 0$$

$$S_n = X_1 + X_2 + X_3 \ldots X_n$$

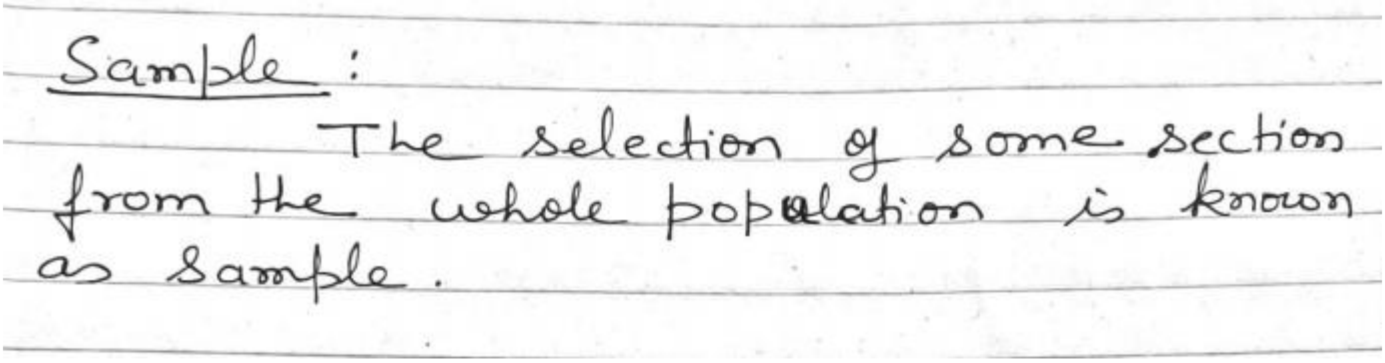
$$E[X_k] = \mu_x \quad \forall k$$

Central limit theorem:

The central limit theorem states that if you have a population with mean μ and standard deviation σ and take sufficiently large random samples from the population **with replacement**, then the distribution of the sample means will be approximately normally distributed.

Population: it is the collection of all items of interest to our study and is usually denoted by an uppercase N, the numbers we have obtained when using a population are called parameters (u and sigma)

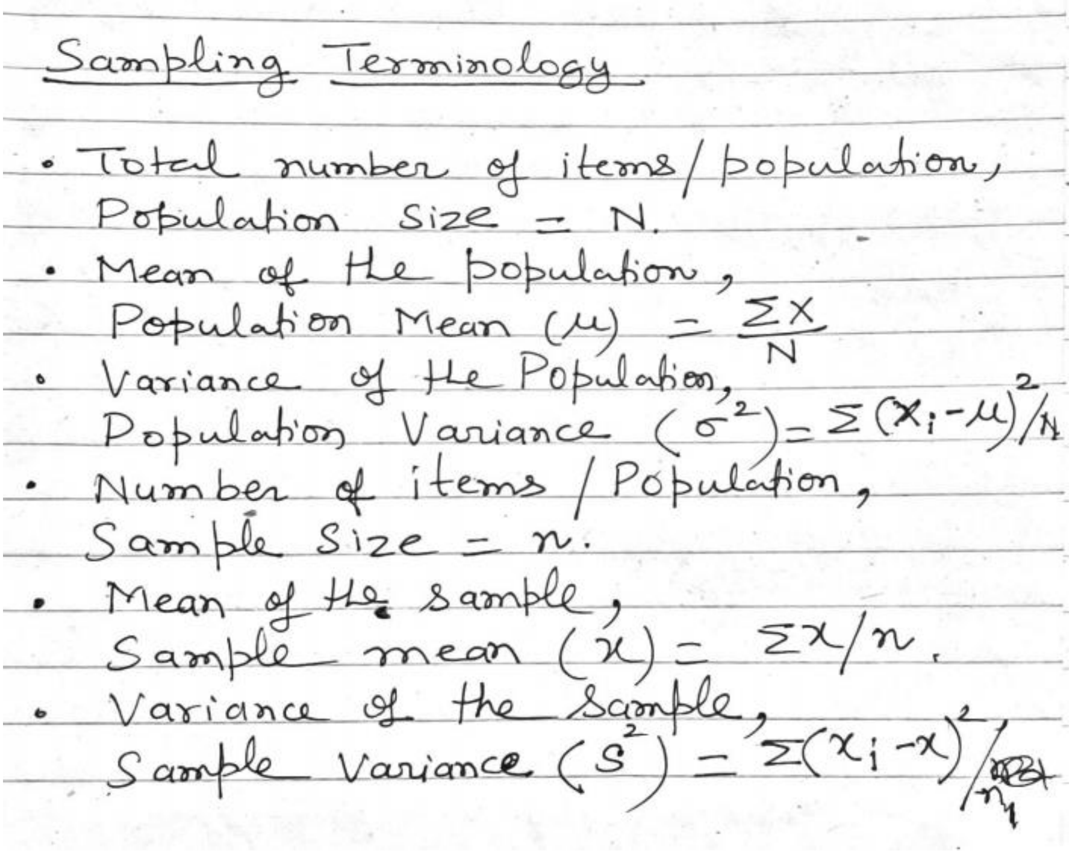
Sample:



Sample :

The selection of some section from the whole population is known as sample.

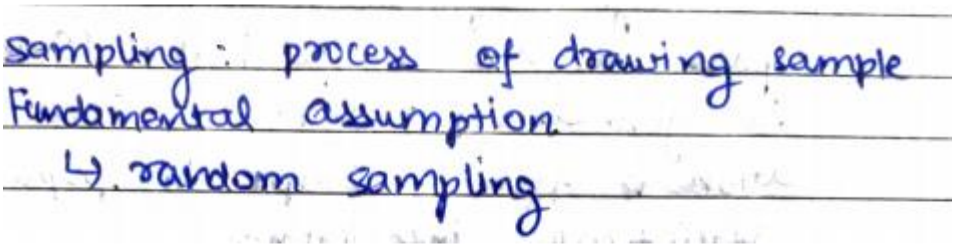
Sampling terminology



Sampling Terminology

- Total number of items / population, Population Size = N.
- Mean of the population, Population Mean $(\mu) = \frac{\Sigma X}{N}$
- Variance of the Population, Population Variance $(\sigma^2) = \Sigma(x_i - \mu)^2 / N$
- Number of items / Population, Sample Size = n.
- Mean of the sample, Sample mean $(x) = \Sigma x / n$.
- Variance of the Sample, Sample Variance $(S^2) = \Sigma(x_i - x)^2 / n_1$

Sampling:

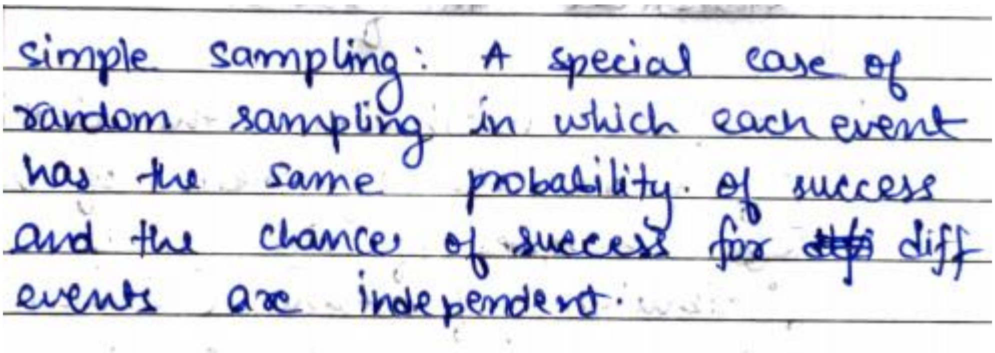


sampling : process of drawing sample

Fundamental assumption

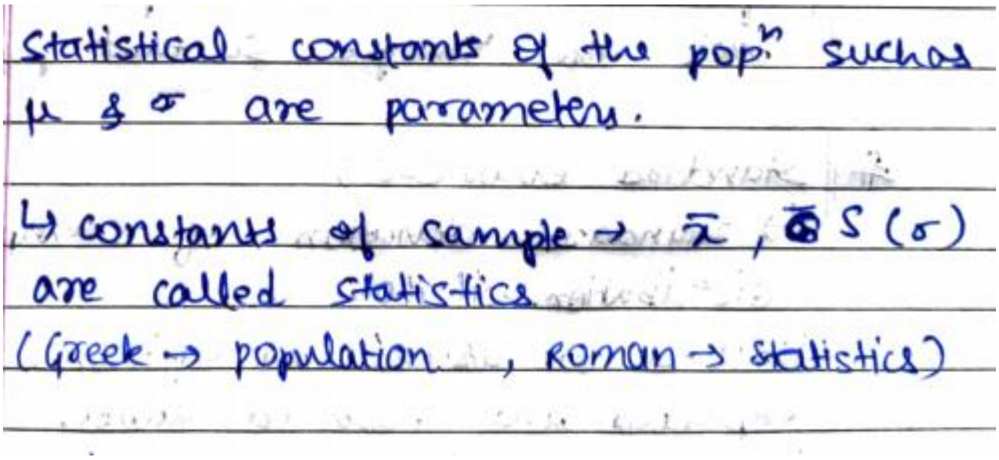
↳ random sampling

Simple sampling:



simple sampling: A special case of random sampling in which each event has the same probability of success and the chances of success for ~~diff~~ diff events are independent.

Parameters and staistics:

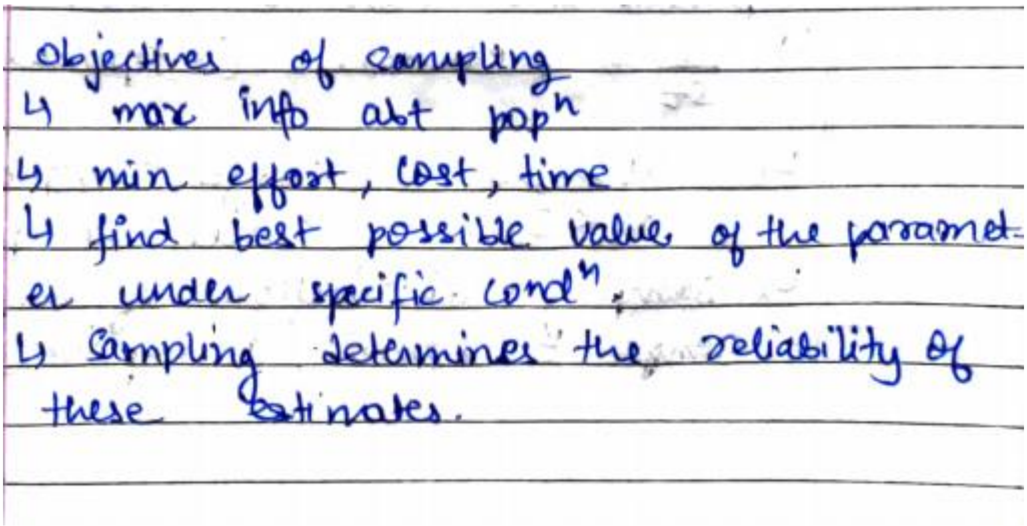


Statistical constants of the pop^n such as μ & σ are parameters.

↳ constants of sample → $\bar{x}$, S (σ) are called statistics

(Greek → population, Roman → statistics)

Objective of sampling:



objectives of sampling

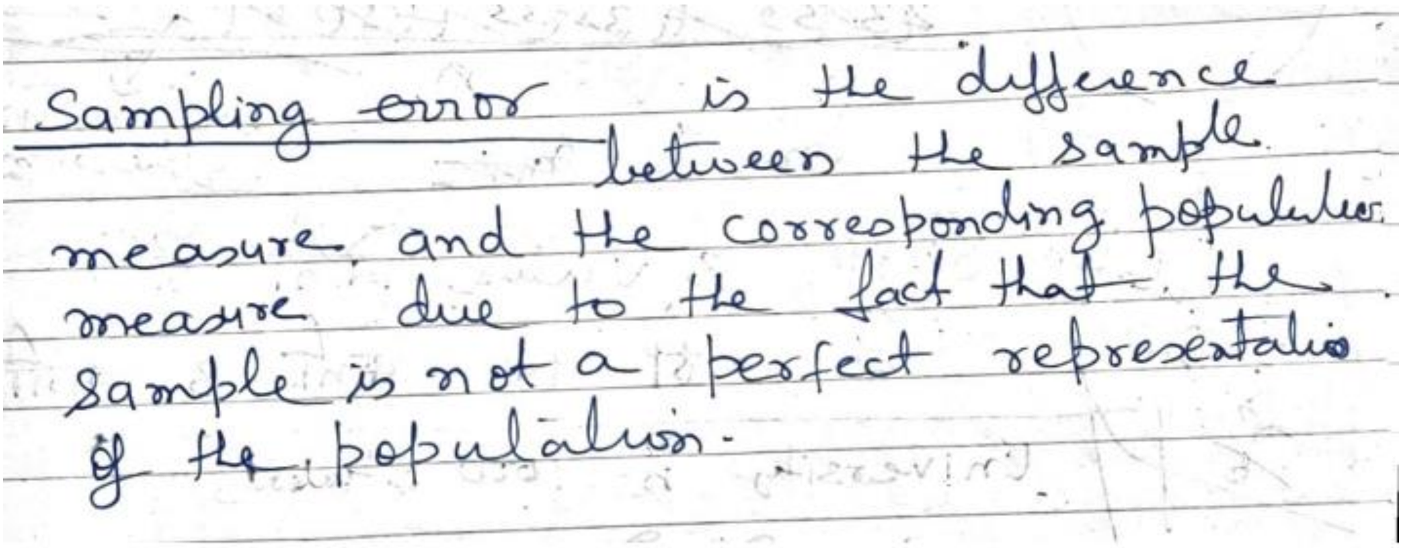
↳ max info abt pop^n

↳ min effort, cost, time

↳ find best possible value of the parameter under specific cond^n.

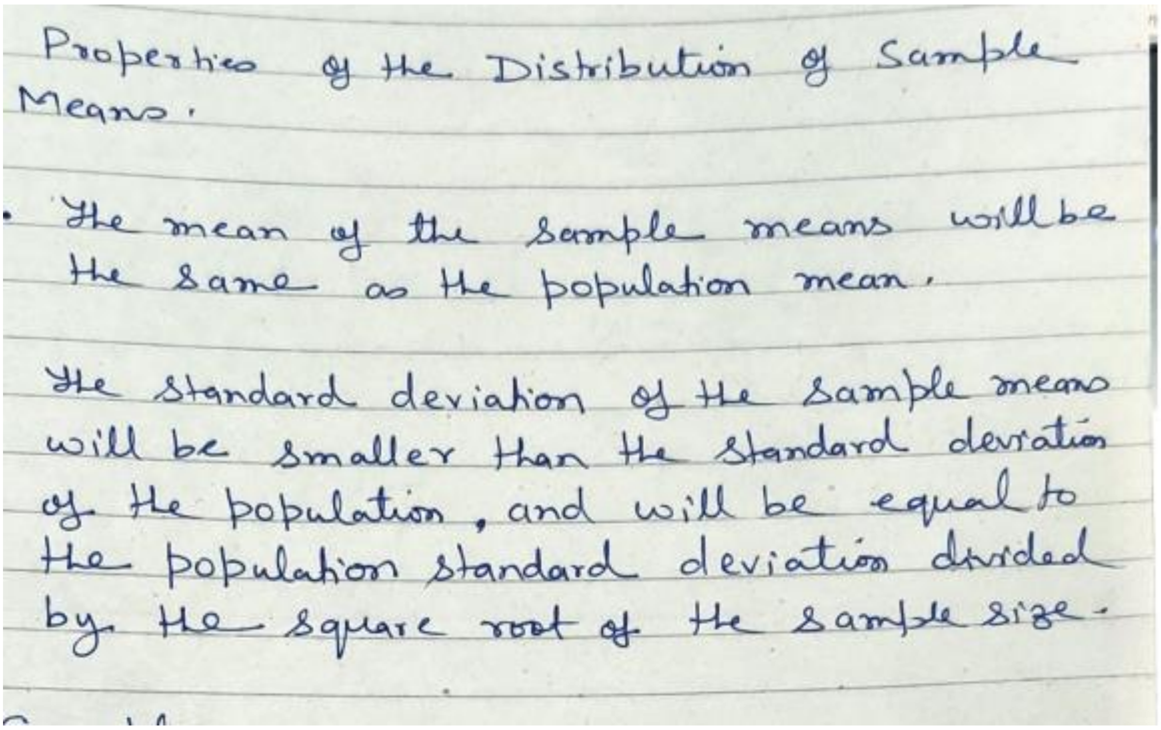
↳ sampling determines the reliability of these estimates.

Sampling error:



Sampling error is the difference between the sample measure and the corresponding population measure due to the fact that the sample is not a perfect representation of the population.

Properties os sampling means and variance

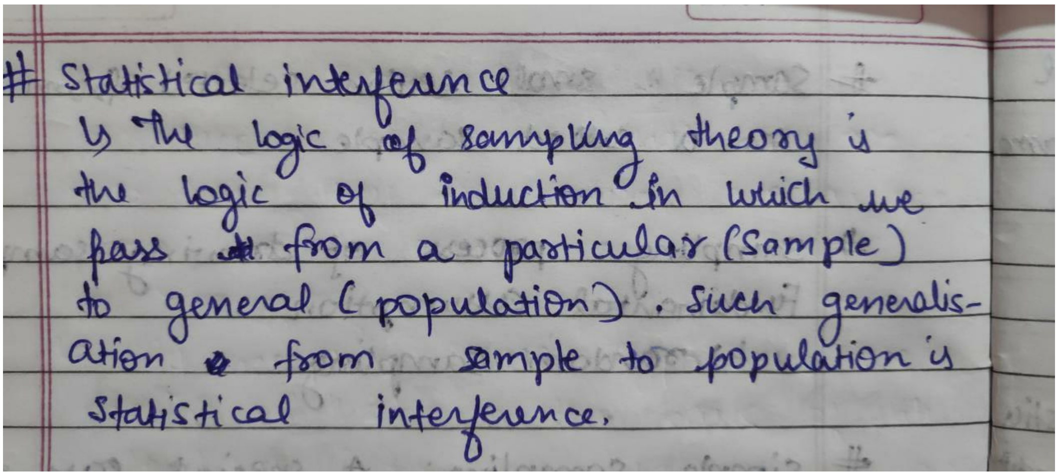


Properties of the Distribution of Sample Means.

- The mean of the sample means will be the same as the population mean.

The standard deviation of the sample means will be smaller than the standard deviation of the population, and will be equal to the population standard deviation divided by the square root of the sample size.

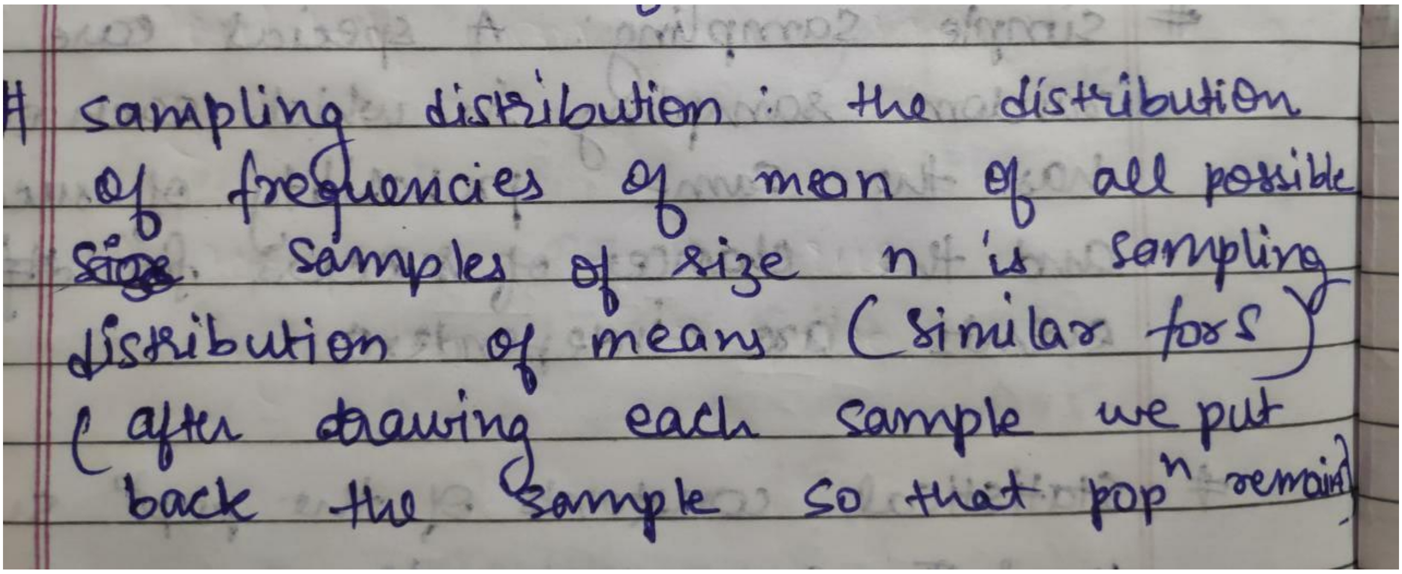
Satistical interference



Statistical interference

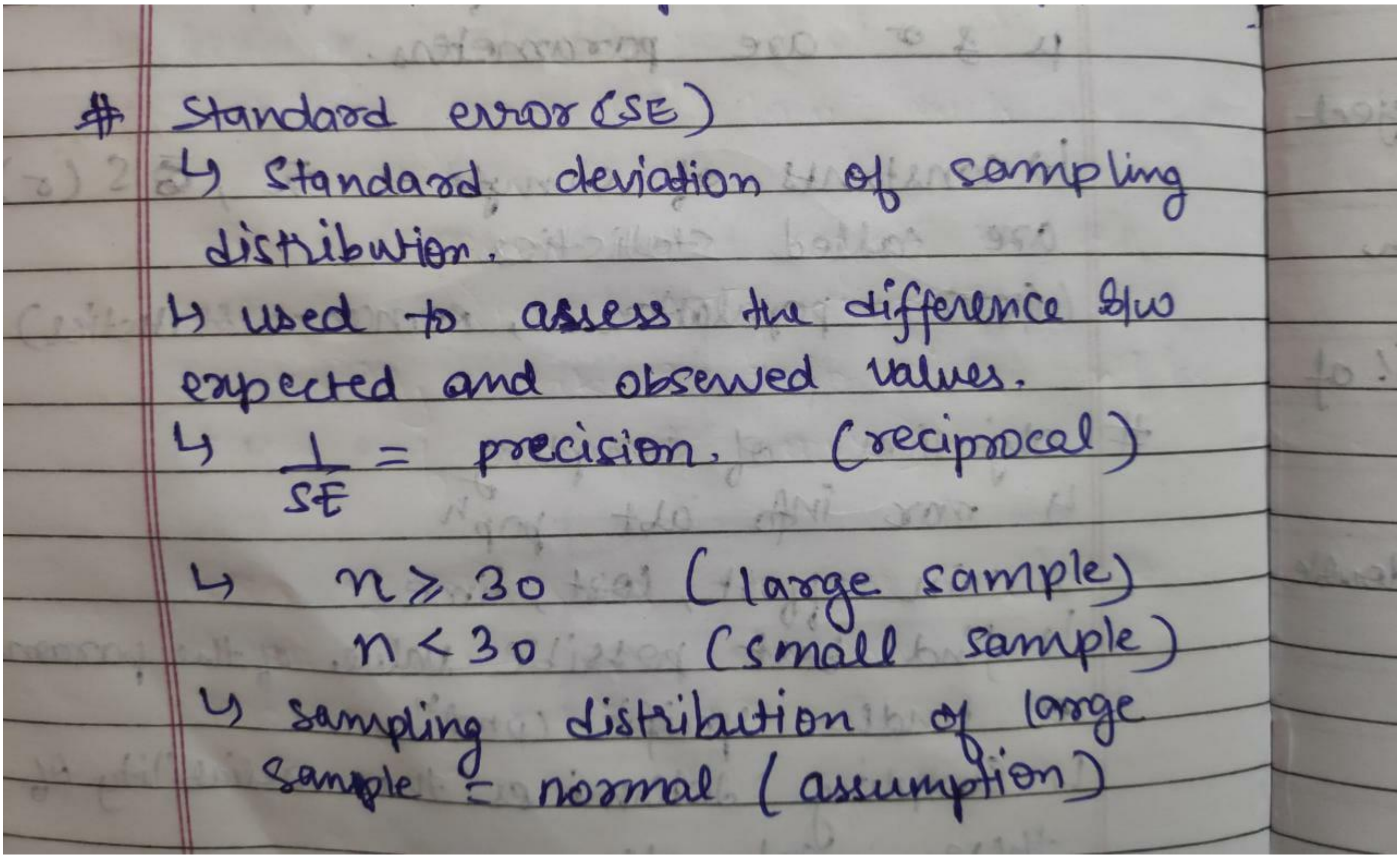
is the logic of sampling theory is the logic of induction in which we pass from a particular (Sample) to general (population). Such generalisation from sample to population is statistical interference.

sampling distribution



sampling distribution : the distribution of frequencies of mean of all possible ~~Size~~ samples of size n is sampling distribution of means (similar for S) (after drawing each sample we put back the sample so that pop^n remain

Standard error



Standard error (SE)

↳ Standard deviation of sampling distribution.

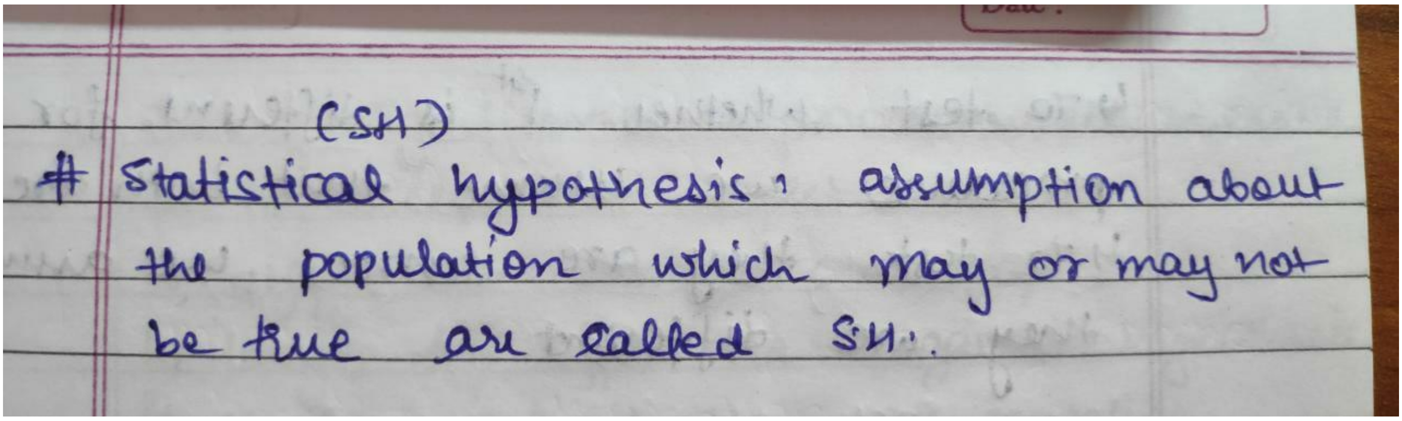
↳ used to assess the difference b/w expected and observed values.

↳ $\frac{1}{SE}$ = precision. (reciprocal)

↳ $n \geq 30$ (large sample)
$n < 30$ (small sample)

↳ sampling distribution of large sample ≃ normal (assumption)

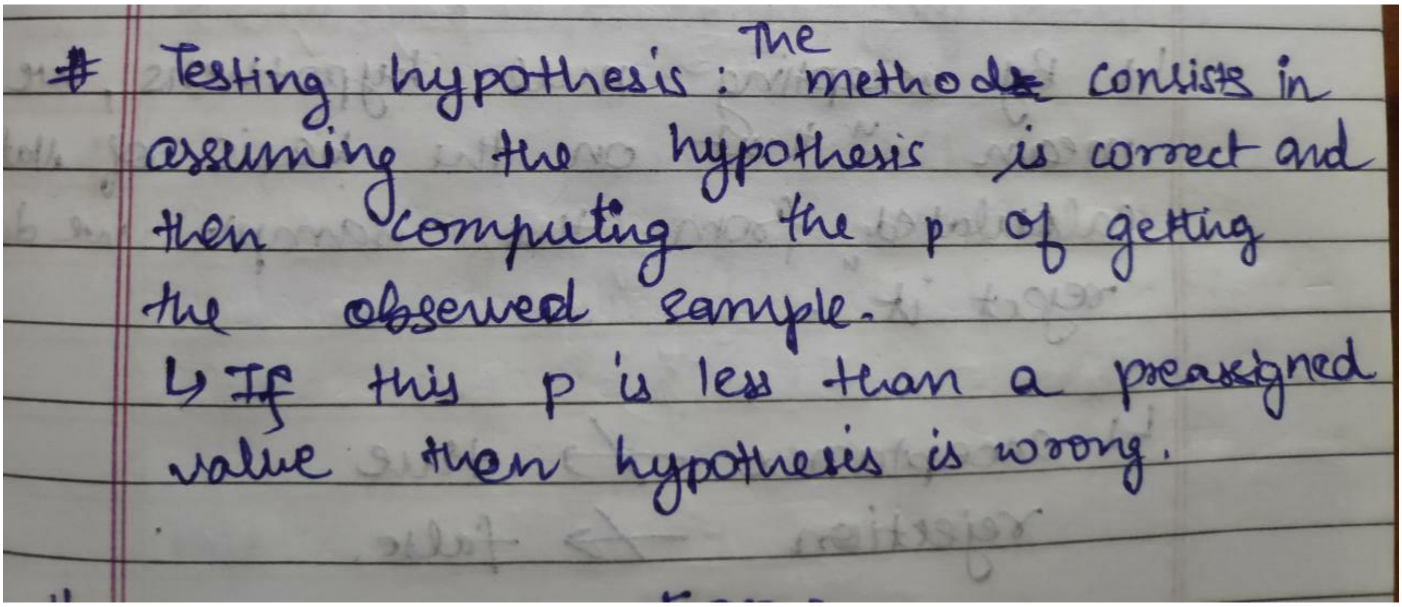
Statistical hypothesis



(SH)

# Statistical hypothesis : assumption about the population which may or may not be true are called SH.

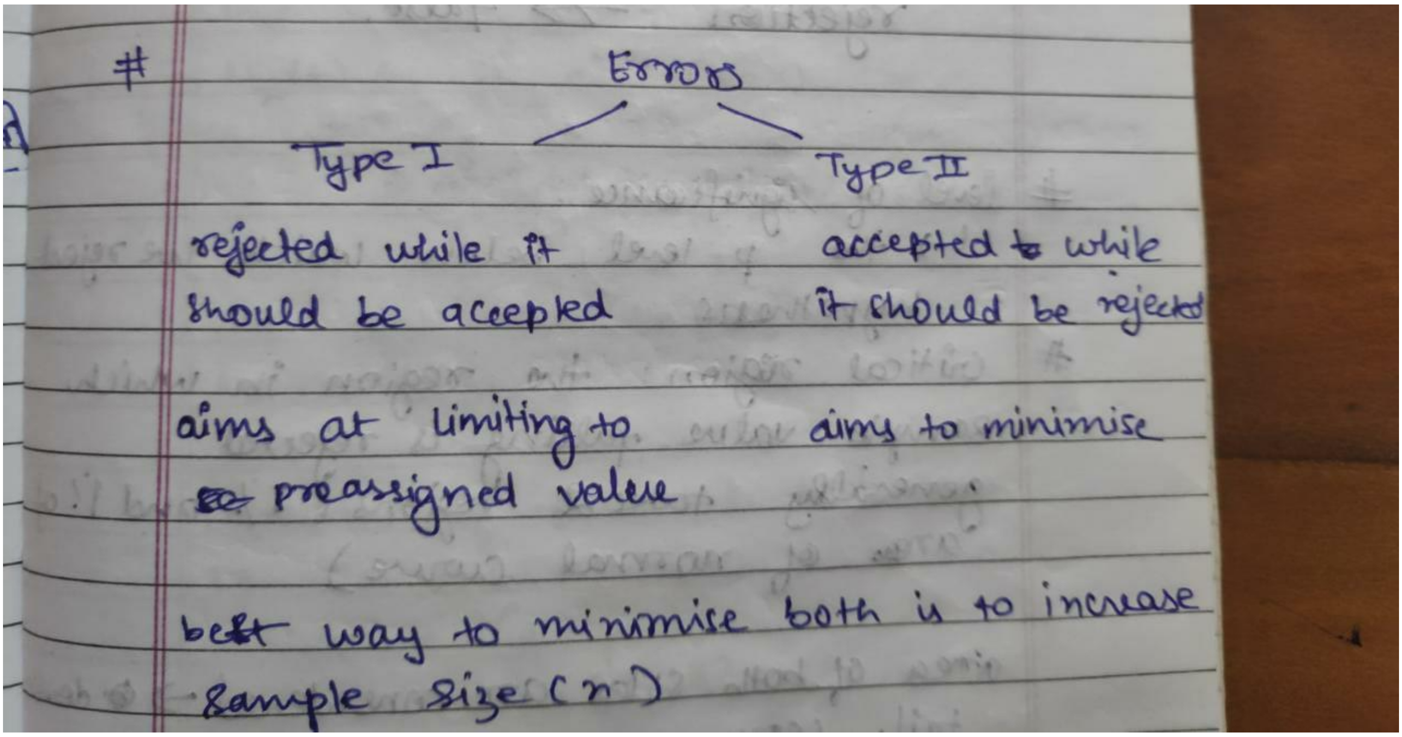
Testing hypothesis



# Testing hypothesis : The method~~s~~ consists in assuming the hypothesis is correct and then computing the p of getting the observed sample.

↳ If this p is less than a preassigned value then hypothesis is wrong.

Errors type1 and type2

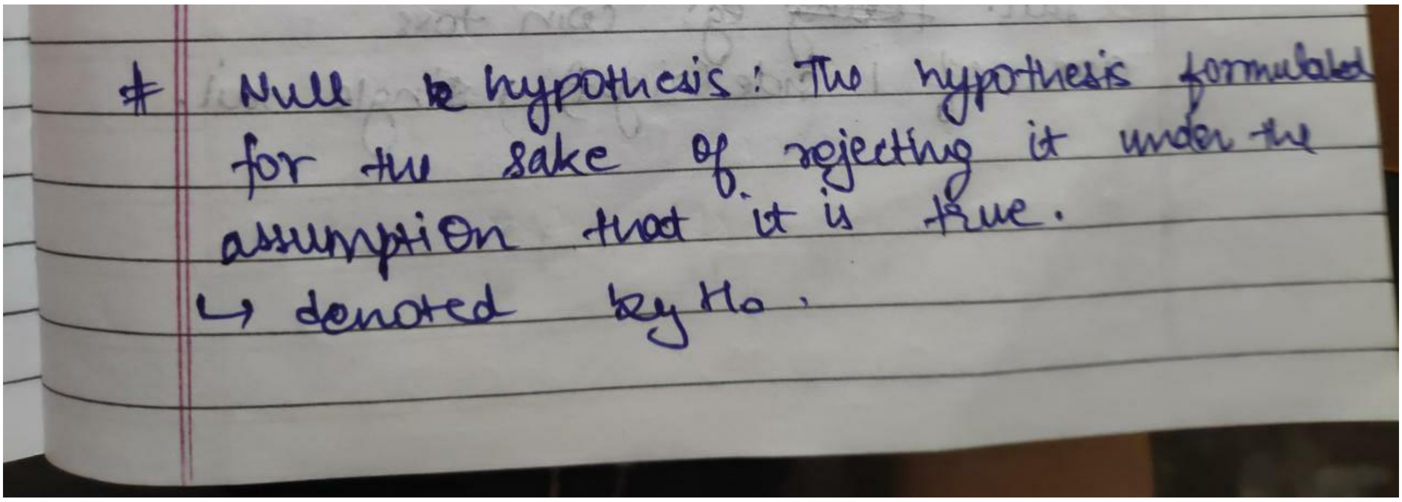


# Errors

Type I	Type II
rejected while it should be accepted	accepted ~~to~~ while it should be rejected
aims at limiting to ~~so~~ preassigned value	aims to minimise

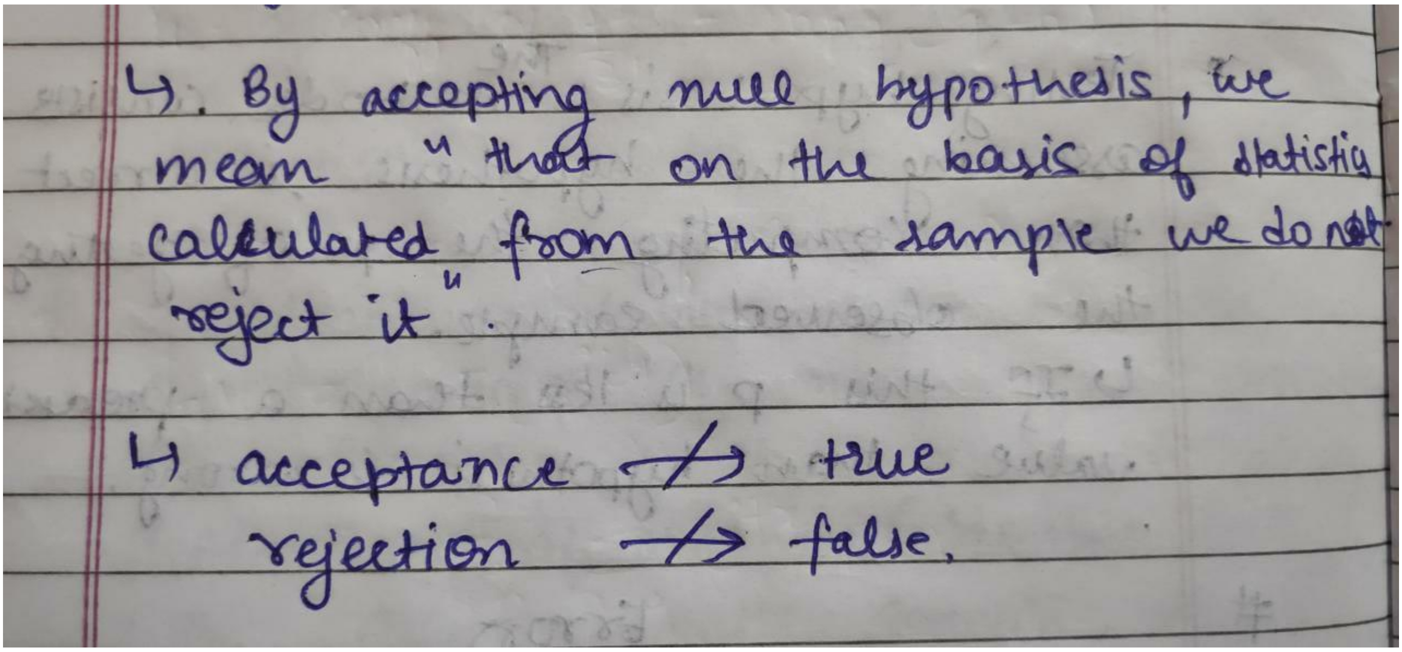
best way to minimise both is to increase sample size (n)

Null hypothesis



Null ~~b~~ hypothesis: The hypothesis formulated for the sake of rejecting it under the assumption that it is true.

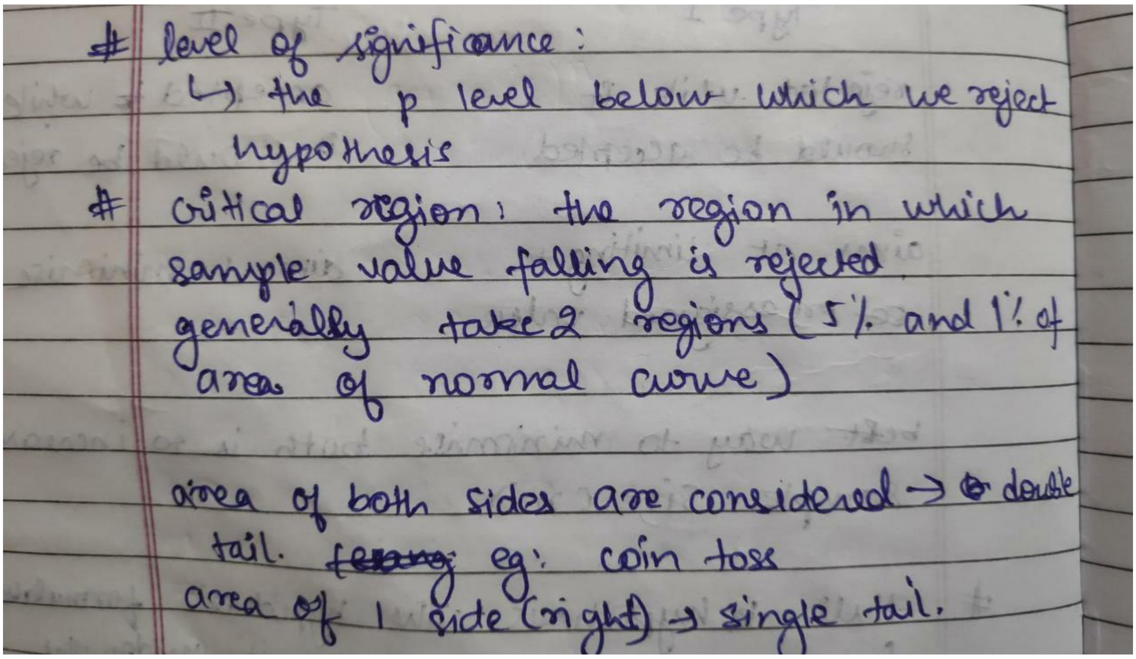
↳ denoted by H_0.



↳ By accepting null hypothesis, we mean "that on the basis of statistics calculated from the sample we do not reject it".

↳ acceptance ↛ true
rejection ↛ false.

Level of significance and critical region



level of significance:

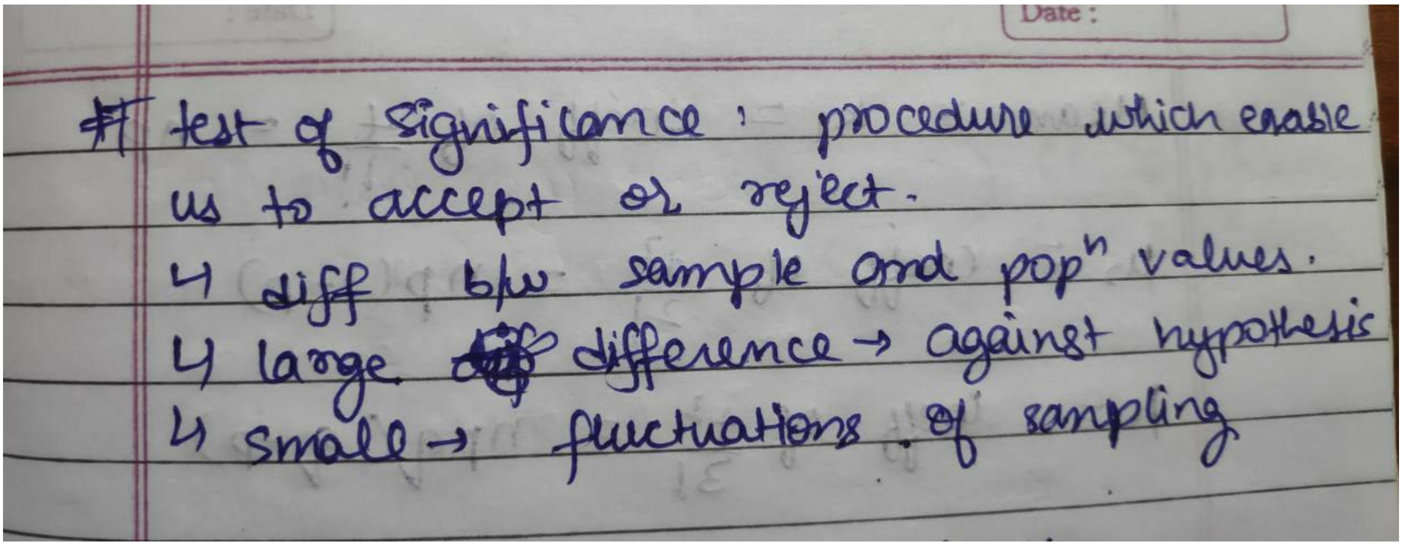
↳ the p level below which we reject hypothesis

critical region: the region in which sample value falling is rejected.

generally take 2 regions (5% and 1% of area of normal curve)

area of both sides are considered → double tail. eg: coin toss
area of 1 side (right) → single tail.

Test of significance



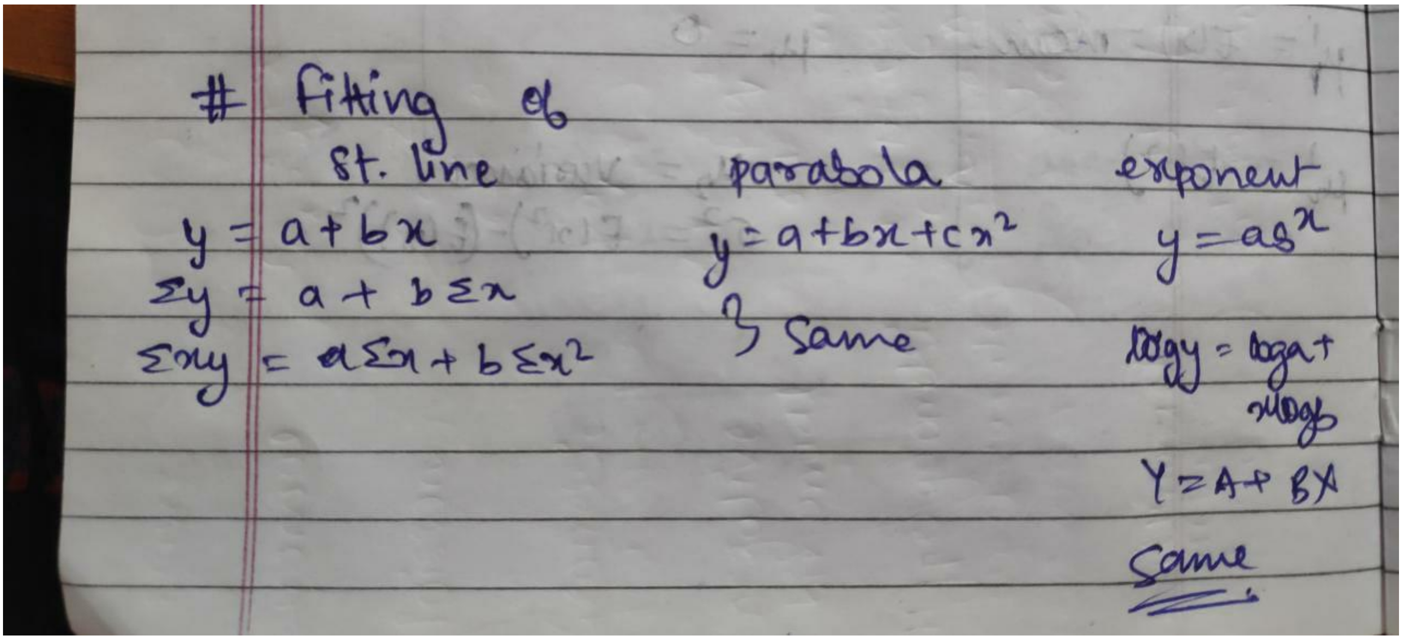
test of significance : procedure which enable us to accept or reject.

↳ diff b/w sample and popⁿ values.

↳ large ~~diff~~ difference → against hypothesis

↳ small → fluctuations of sampling

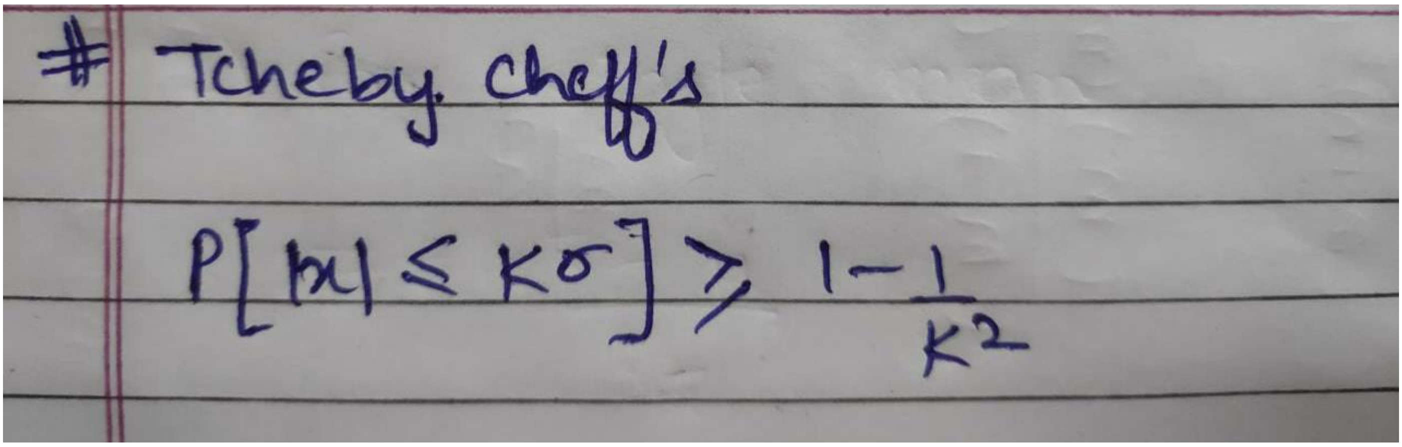
Fitting to straight line, parabola, exponential



fitting of

st. line	parabola	exponent
$y = a+bx$	$y = a+bx+cx^2$	$y = ab^x$
$\Sigma y = a + b\Sigma x$	} Same	$\log y = \log a + x\log b$
$\Sigma xy = a\Sigma x + b\Sigma x^2$		$Y = A + BX$
		Same

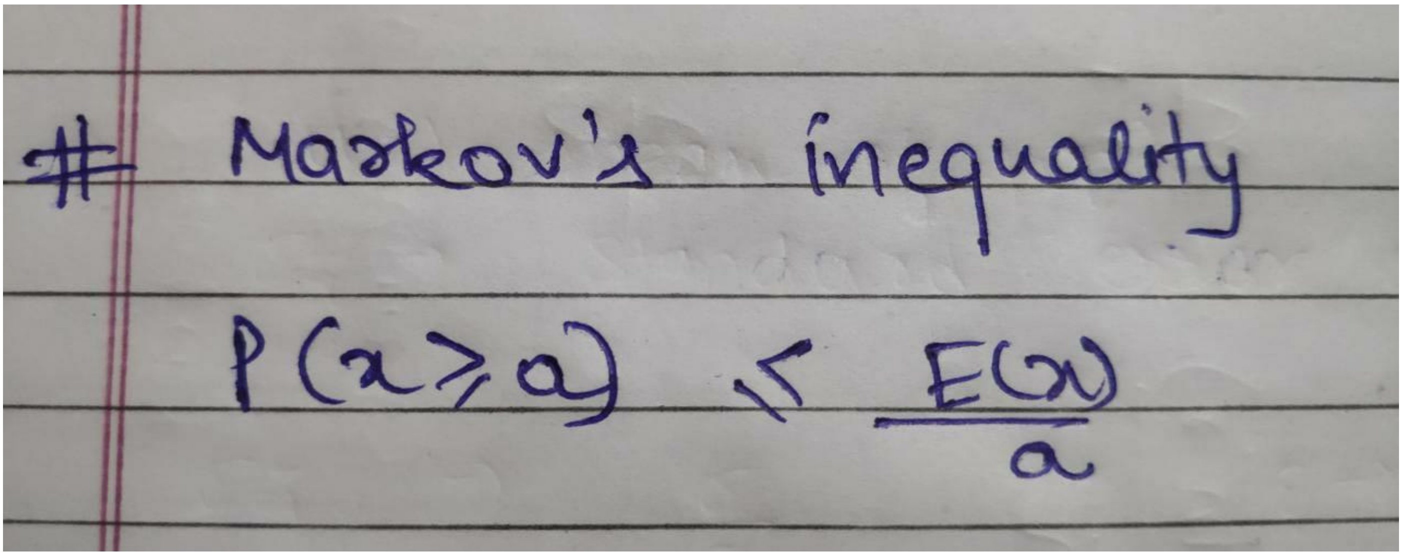
Tchebycheffs



Tcheby cheff's

$$P[|x| \leq K\sigma] \geq 1 - \frac{1}{K^2}$$

Markov



Markov's inequality

$P(x \geq a) \leq \frac{E(x)}{a}$

Attribute sampling refers to a statistical sampling tool used by the auditors to analyze the features of a particular population

Confidence limits for the mean are an interval estimate for the mean. Confidence limits tell you how accurate your estimate of the mean is likely to be.

The chi-square compares the size any discrepancies between the expected results and the actual results, given the size of the sample and the number of variables in the relationship.

Degrees of Freedom refers to the maximum number of logically independent values, which are values that have the freedom to vary, in the data sample.

Goodness of fit test helps you see if your sample data is accurate or somehow skewed

The significance level, also denoted as alpha, is a measure of the strength of the evidence that must be present in your sample before rejecting the null.

In a **test of independence,** we state the null and alternative hypotheses in words. Since the contingency table consists of two factors, the null hypothesis states that the factors are independent and the alternative hypothesis states that they are not independent (dependent).

The **contingency coefficient** is a coefficient of association that tells whether two variables or data sets are independent or dependent of each other

Yates' correction is to prevent the overestimation of statistical significance for small data when 'zero cells' are present in a 2 × 2 contingency table.

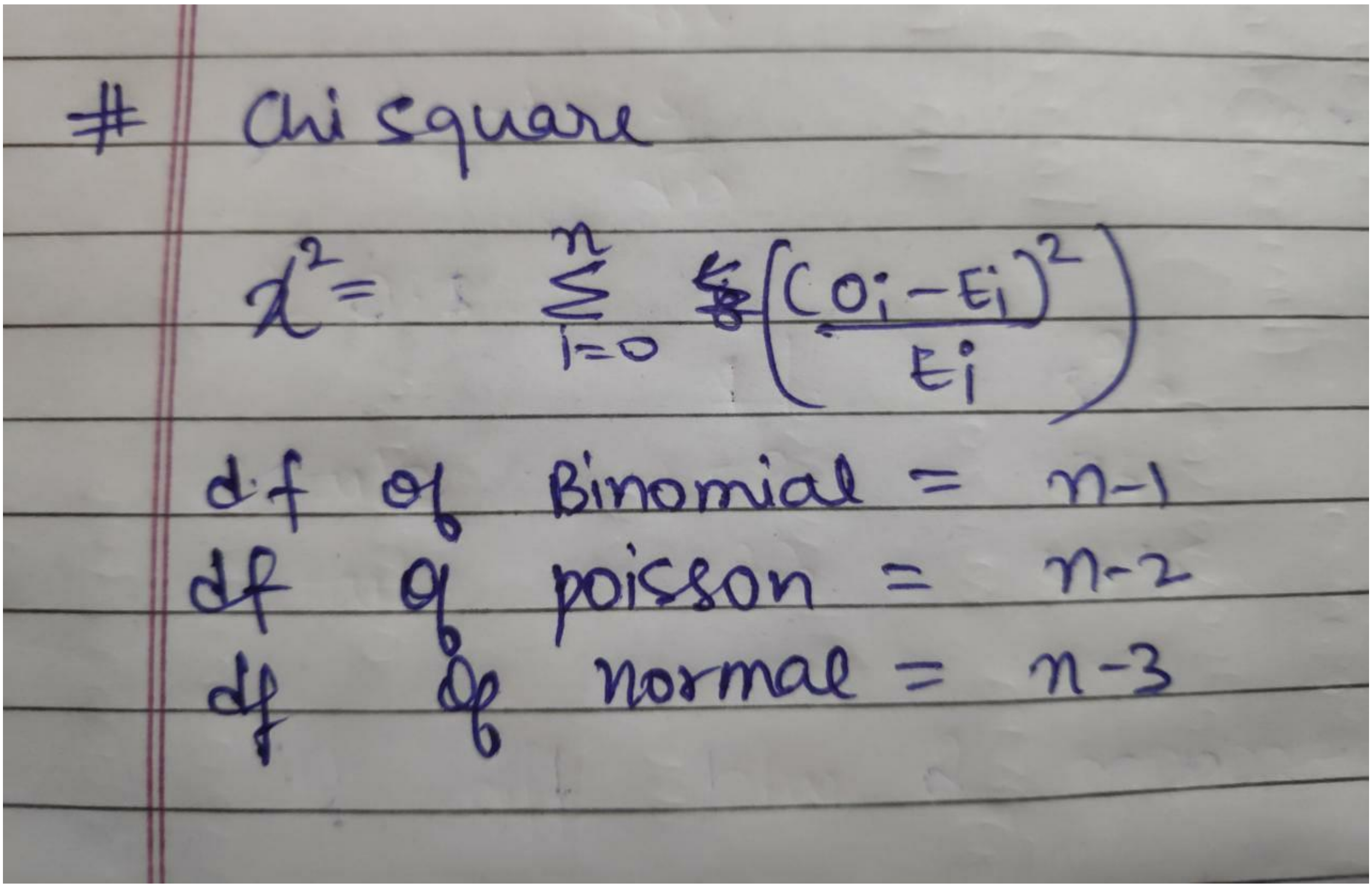
Point estimation, in statistics , the process of finding an approximate value of some parameter—such as the mean (average)—of a population from random samples of the population

interval estimation is the use of sample data to calculate an interval of possible values of an unknown population parameter;

Maximum likelihood estimation is a method that determines values for the parameters of a model. The parameter values are found such that they maximise the likelihood that the process described by the model produced the data that were actually observed.

Test statistic	**Null and alternative hypotheses**	**Statistical tests that use it**
t-value	Null: The means of two groups are equal Alternative: The means of two groups are not equal	T-test Regression tests
z-value	Null: The means of two groups are equal Alternative:The means of two groups are not equal	Z-test
F-value	Null: The variation among two or more groups is greater than or equal to the variation between the groups Alternative: The variation among two or more groups is smaller than the variation between the groups	ANOVA ANCOVA MANOVA
X2-value	Null: Two samples are independent Alternative: Two samples are not independent (i.e. they are correlated)	Chi-squared test Non-parametric correlation tests

Chi sqaure variate



# Chi square

$$\chi^2 = \sum_{i=0}^{n} \left(\frac{(O_i - E_i)^2}{E_i}\right)$$

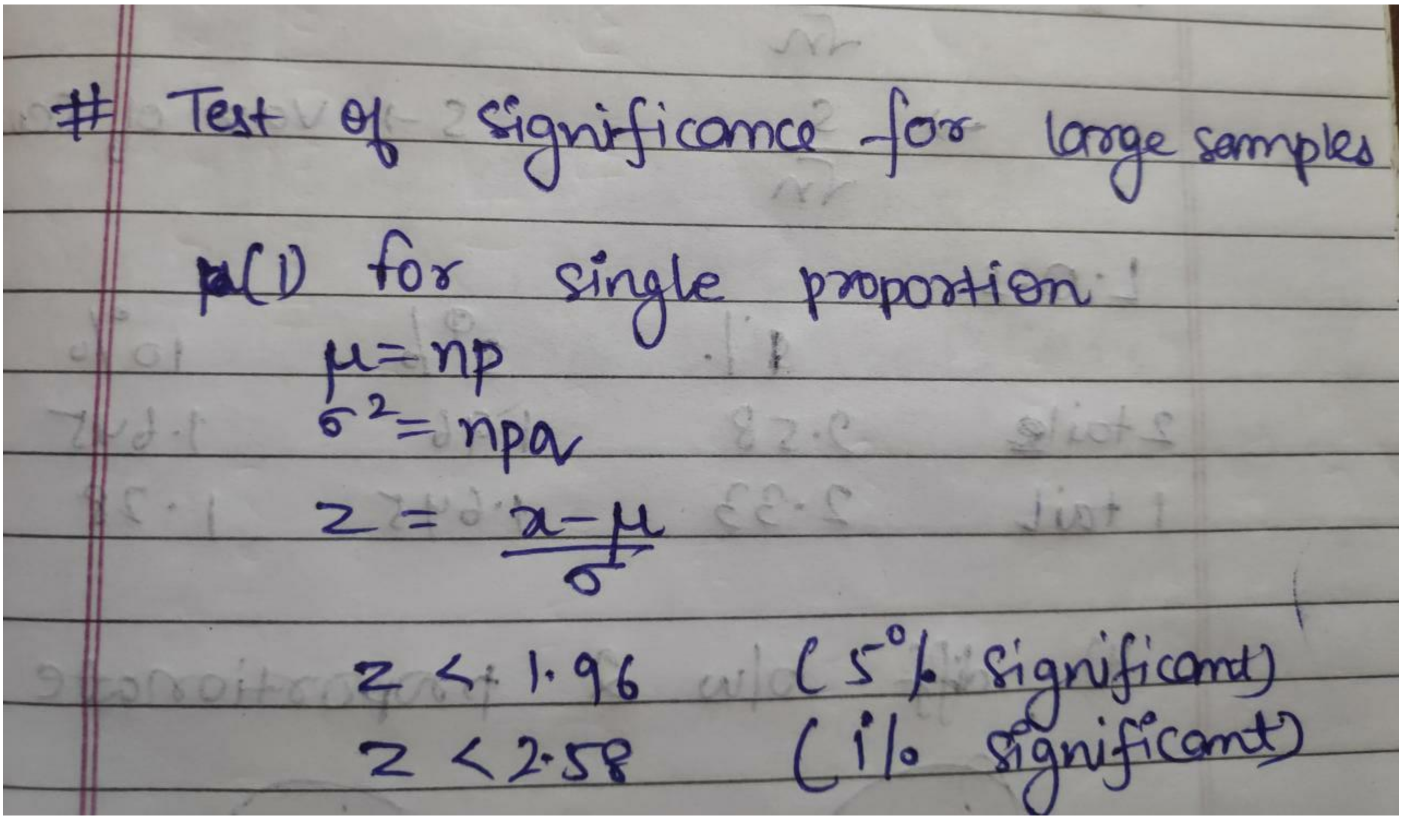
d.f of Binomial = $n-1$

df q poisson = $n-2$

df of normal = $n-3$

Test of significance for large samples:

1) **For a signle proportion**



# Test of significance for large samples

(1) for single proportion

$\mu = np$

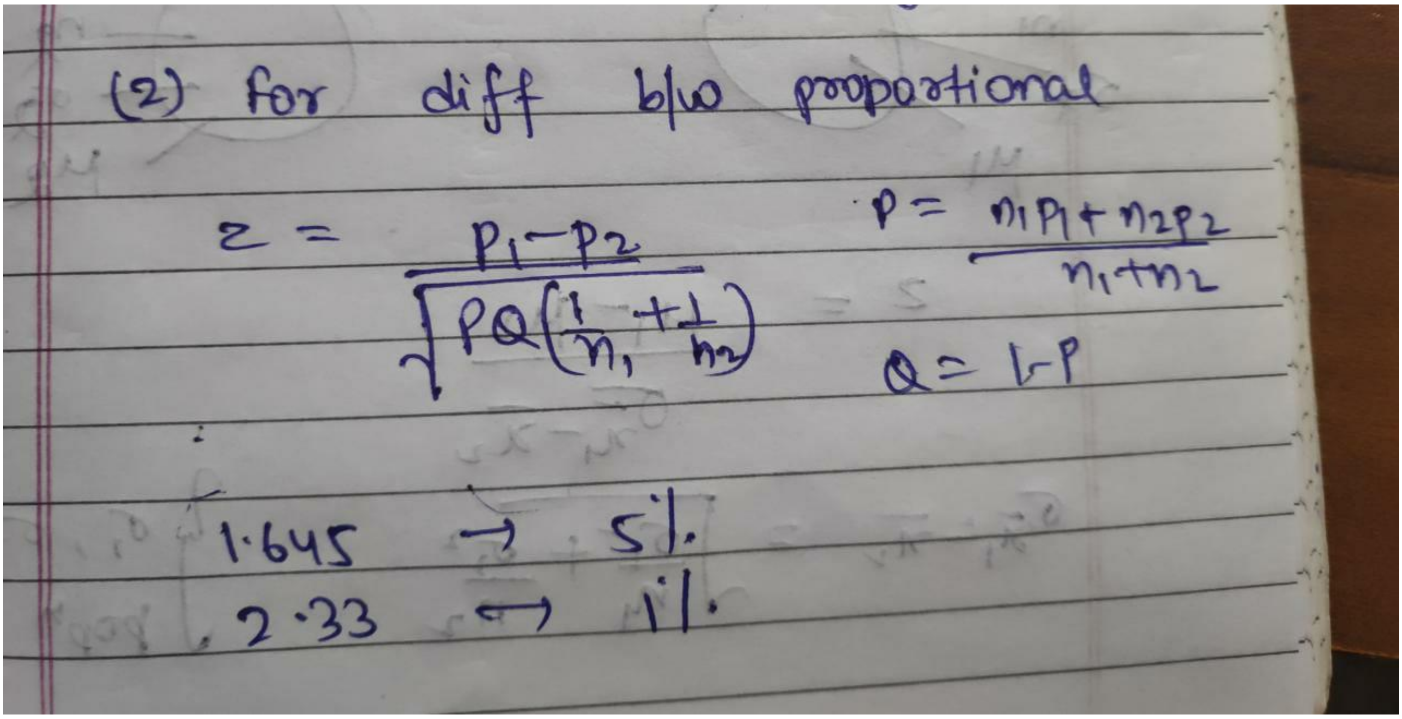
$\sigma^2 = npq$

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

$z < 1.96$ (5% significant)

$z < 2.58$ (1% significant)

2) **For difference between proportion**



(2) for diff b/w proportional

$$z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

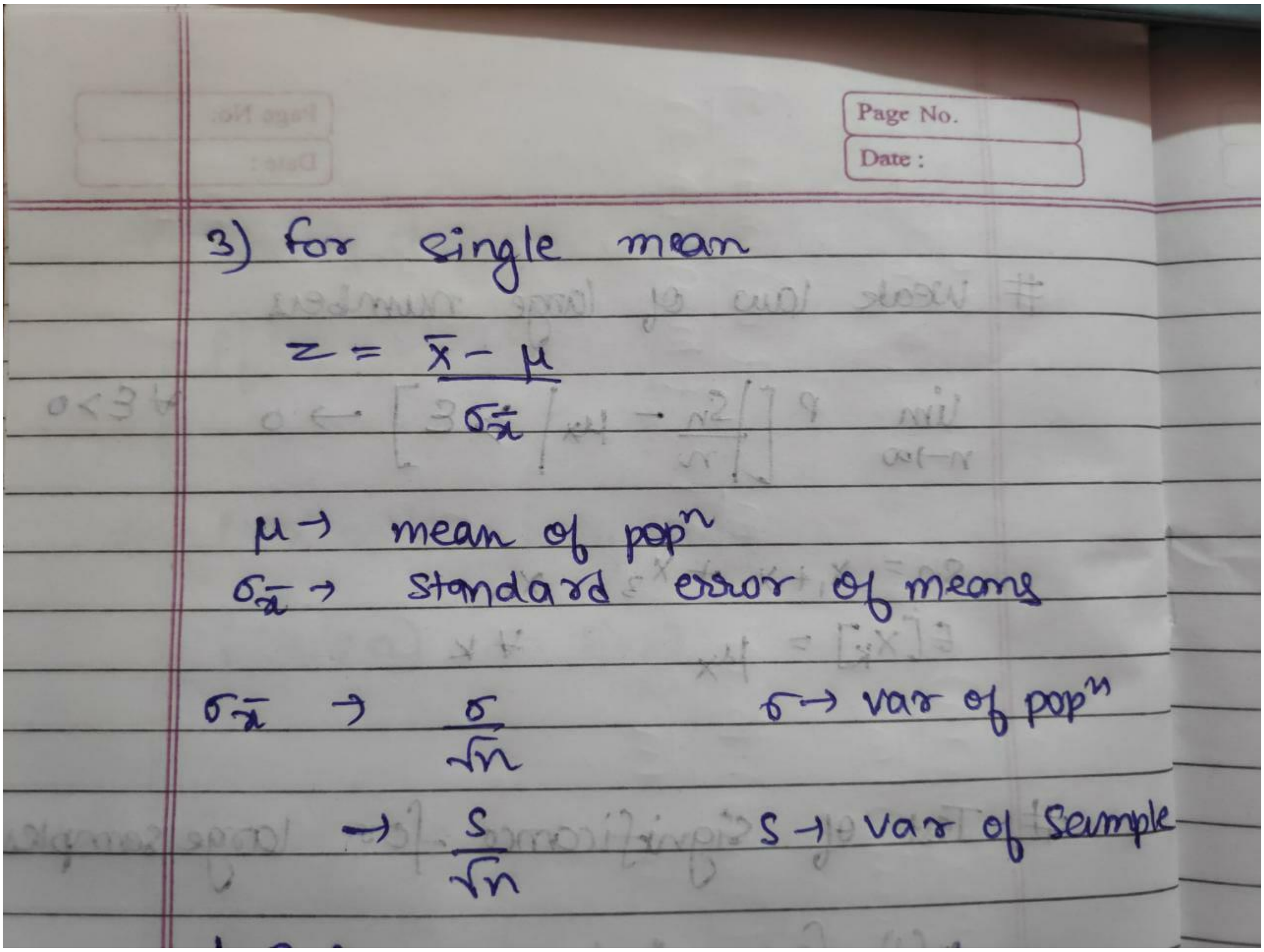
$$P = \frac{n_1p_1 + n_2p_2}{n_1+n_2}$$

$$Q = 1-P$$

1.645 → 5%

2.33 → 1%

3) For single mean



Page No.

Date :

3) for single mean

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

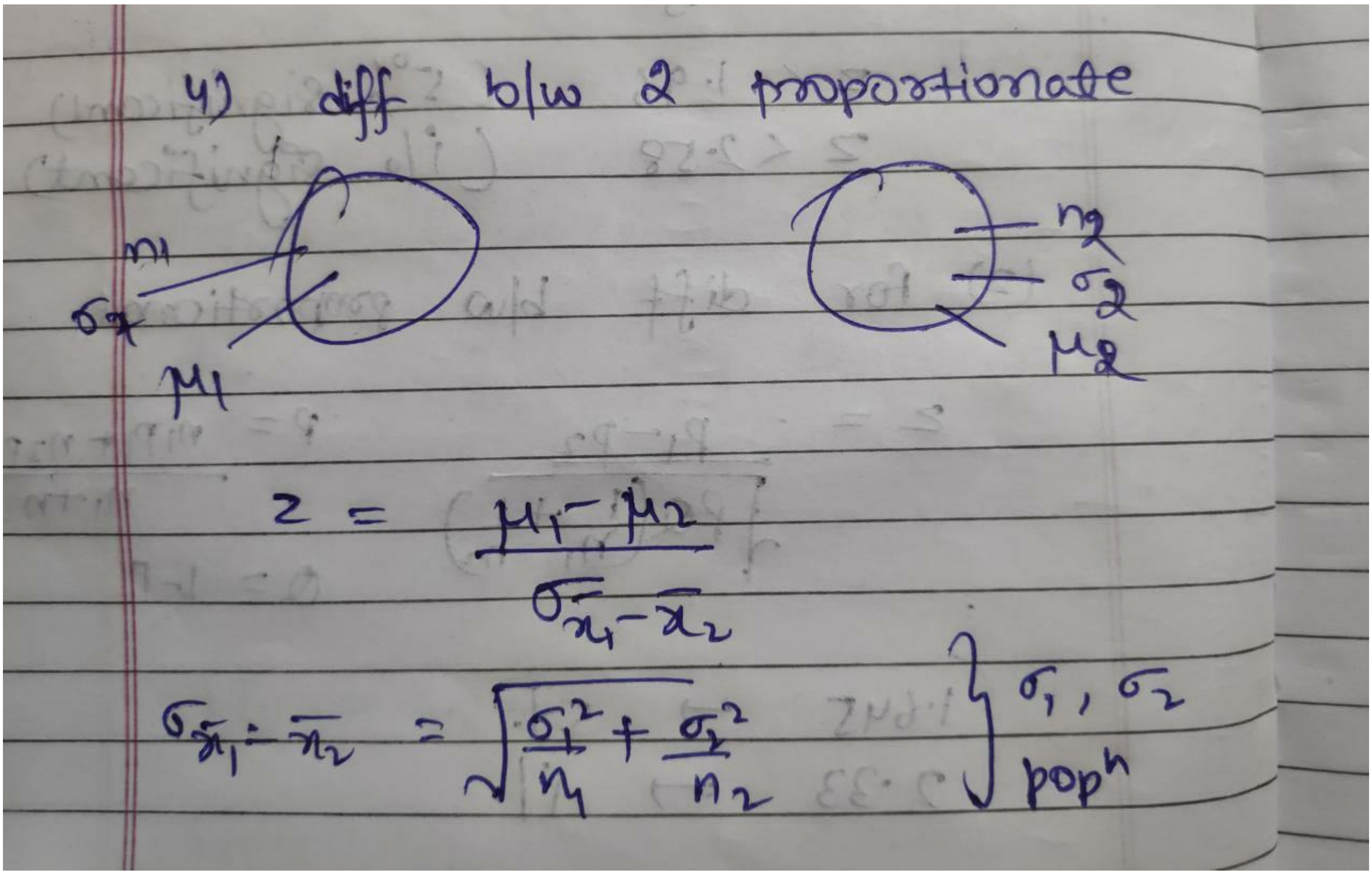
$\mu \rightarrow$ mean of pop^n

$\sigma_{\bar{x}} \rightarrow$ standard error of means

$\sigma_{\bar{x}} \rightarrow \frac{\sigma}{\sqrt{n}}$ $\sigma \rightarrow$ var of pop^n

$\rightarrow \frac{s}{\sqrt{n}}$ $s \rightarrow$ var of semple

4) Difference between 2 proportionate



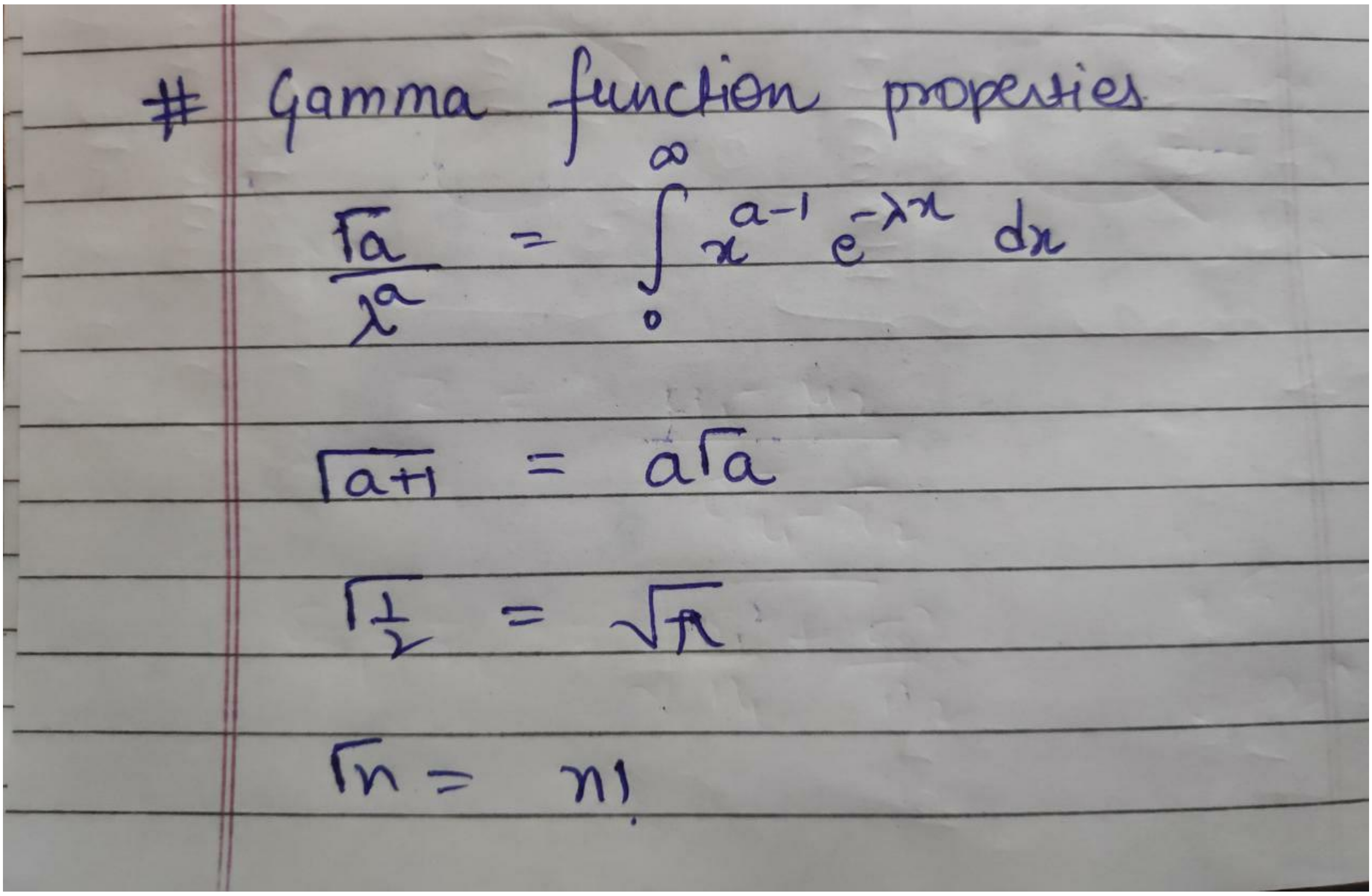
4) diff b/w 2 proportionate

n1
σ1
μ1
n2
σ2
μ2

$$z = \frac{\mu_1 - \mu_2}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad \Big\} \; \sigma_1, \sigma_2 \text{ pop}^n$$

Gamma function properties



Gamma function properties

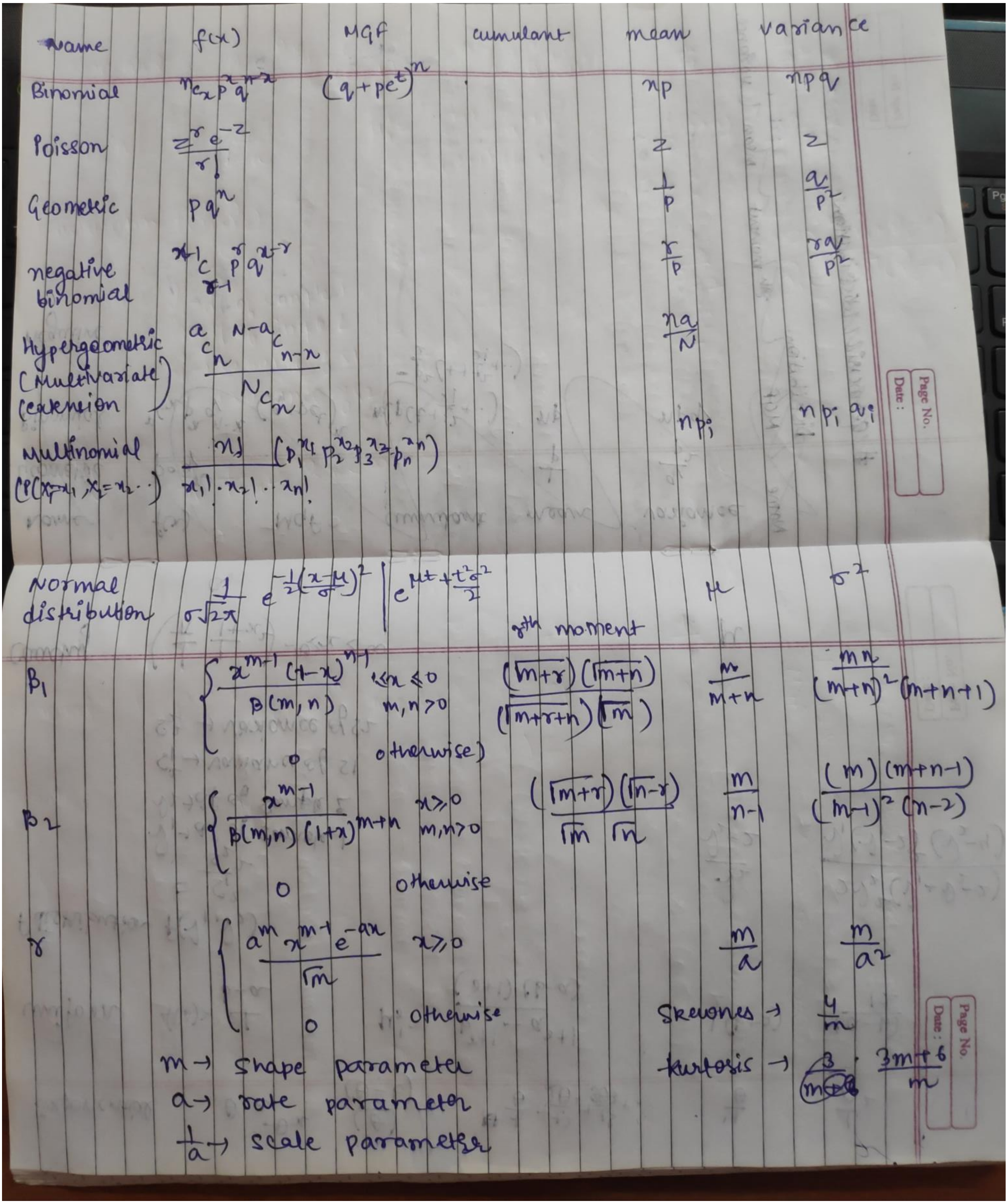
$$\frac{\Gamma a}{\lambda^a} = \int_0^\infty x^{a-1} e^{-\lambda x} \, dx$$

$$\Gamma_{a+1} = a \Gamma a$$

$$\Gamma_{\frac{1}{2}} = \sqrt{\pi}$$

$$\Gamma n = n!$$

Distributions



Name	f(x)	MGF	cumulant	mean	variance
Binomial	${}^{n}C_{x}\, p^{x} q^{n-x}$	$(q+pe^{t})^{n}$		np	npq
Poisson	$\frac{z^{r} e^{-z}}{r!}$			z	z
Geometric	pq^{n}			$\frac{1}{p}$	$\frac{q}{p^{2}}$
negative binomial	${}^{x-1}C_{r-1}\, p^{r} q^{x-r}$			$\frac{r}{p}$	$\frac{rq}{p^{2}}$
Hypergeometric (Multivariate) (extension)	$\frac{{}^{a}C_{n}\; {}^{N-a}C_{n-x}}{{}^{N}C_{n}}$			$\frac{na}{N}$	
Multinomial ($P(X_1=x_1, X_2=x_2 \cdots)$)	$\frac{n!}{x_1! \cdot x_2! \cdots x_n!}(p_1^{x_1} p_2^{x_2} p_3^{x_3} \cdots p_n^{x_n})$			np_i	$n p_i q_i$
Normal distribution	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^{2}}$	$e^{\mu t + \frac{t^{2}\sigma^{2}}{2}}$		μ	σ^{2}
			rth moment		
β_1	$\begin{cases} \frac{x^{m-1}(1-x)^{n-1}}{B(m,n)} & 1\le x \le 0 \\ & m,n>0 \\ 0 & \text{otherwise} \end{cases}$		$\frac{(\Gamma m+r)(\Gamma m+n)}{(\Gamma m+r+n)(\Gamma m)}$	$\frac{m}{m+n}$	$\frac{mn}{(m+n)^{2}(m+n+1)}$
β_2	$\begin{cases} \frac{x^{m-1}}{B(m,n)(1+x)^{m+n}} & x\geq 0 \\ & m,n>0 \\ 0 & \text{otherwise} \end{cases}$		$\frac{(\Gamma m+r)(\Gamma n-r)}{\Gamma m\; \Gamma n}$	$\frac{m}{n-1}$	$\frac{(m)(m+n-1)}{(m-1)^{2}(n-2)}$
γ	$\begin{cases} \frac{a^{m} x^{m-1} e^{-ax}}{\Gamma m} & x\geq 0 \\ 0 & \text{otherwise} \end{cases}$			$\frac{m}{a}$	$\frac{m}{a^{2}}$
				Skewnes →	$\frac{4}{m}$
				kurtosis →	$\frac{3}{m+6}$ $\frac{3m+6}{m}$

m → shape parameter

a → rate parameter

$\frac{1}{a}$ → scale parameter

Exponential	$\theta \cdot e^{-\theta x}$	$\frac{(\theta)}{(\theta - t)}$	~~θ~~ $\frac{t}{\theta} + \frac{t^2}{2\theta^2} + \frac{t^3}{3\theta^3} + \cdots$	$\frac{1}{\theta}$	$\frac{1}{\theta^2}$
uniforn	$f(x) = \frac{1}{b-a}$	$\mu_r' = \frac{b^{r+1} - a^{r+1}}{(r+1)(b-a)}$		$\frac{1}{2}(b+a)$	$\frac{(a-b)^2}{12}$
f distribution	$f(\nu_1, \nu_2)$ $= \frac{S_1^2}{S_2^2}$ $\nu_1 \to$ df of sample 1 $\nu_2 \to$ df of sample 2 $S_1^2 \to$ variance of S1 $S_2^2 \to$ variance of S2			$\frac{\nu_2}{\nu_2 - 2}$	$\frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_1 - 2)^2(\nu_2 - 4)}$
cauchy	$\left(\frac{1}{\pi} \frac{1}{1+x^2}\right) -\infty < x < \infty$ 0			μ	

Replacee $x \to \frac{x-\mu}{\lambda}$