Week - 3

Deterministic phenomenon : whose outcome can be predicted with a very high degree of confidence

Stochastic phenomenon : Phenomenon which can have many possible outcomes for same experimental conditions. Outcomes can be predicted with limited confidence

Sources of errors in observed outcomes -

- Model Error - Lack of Knowledge of generating process
- Measurement Error - Errors in sensors used for observing outcomes

Types of random phenomenon -

(i) Discrete : outcomes are finite
(ii) Continuous : Infinite no. of outcomes.

Sample Space - Set of all possible outcomes of a random phenomena

Event - Subset of the sample space

Probability Measure - is a function that assigns a real value to every outcomes of a random phenomena, which satiesfies following axioms -

- $0 \leq P(A) \leq 1$ (Probabilities are non-negative and less than one for any event A)

- $P(S) = 1$ (one of the outcomes should occour)

- For two mutually exclusive events A and B
  $P(A \cup B) = P(A) + P(B)$

Independent Events →
Two events are independent if occourance of one has no influence on occourance of other

Formally, A & B are independent if & only if
$P(A \cap B) = P(A) \times P(B)$

Mutually Exclusive Events →
Two events are mutually exclusive if occourance of one implies other event does not occour

Some rules of probability →

- If $B \subseteq A$, $P(B) \leq P(A)$

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Conditional Probability →
If two events A & B are not independent, then information available about the

outcome of event A can influence the predictability of event B

→ $P(B|A) = P(A \cap B) / P(A)$ if $P(A) > 0$

→ $P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$ ⇒ Bayes formula

→ $P(A) = P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c)$

Eg: In a manufacturing process, 1000 parts are produced, of which 50 are defective. We randomly take a part from the day's production.

Outcomes : {A = Defective part B = Non-defective part}

$P(A) = 50/1000$, $P(B) = 950/1000$

Suppose we draw a second part, without replacing the first part

Outcomes : {C = Defective part D = Non-defective part}

$P(C) = 50/1000$ (no information about outcome of first draw)

$P(C|A) = 49/999$ (given that first draw was of defective part)

$P(C|B) = 50/999$ (given that first draw was of non-defective part)

$P(C) = 49/999 * 50/1000 + 50/999 * 950/1000$
$= 50/1000$

$P(A|C) = P(A \cap C)/P(C) = P(C|A) \cdot P(A) / P(C) = 49/999$

Random Variable →

A random variable (RV) is a map from sample space to a real line such that there is a unique real no. corresponding to every outcome of sample space

(i) Discrete RV

Eg : throw of a dice or coin

(ii) Continuous RV

Eg : sensor readings, time interval b/w failures

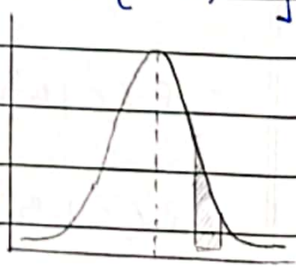
Probability Mass / Density Functions →

For a discrete RV, the probability mass function assigns a probability to every outcome in sample space

For a continuous RV, the probability density function $f(x)$ can be used to assign a probability to every interval on a real line.

Continuous RV can take any value $[-\infty, \infty]$ (area under the curve).

$$P(a < x < b) = \int_a^b f(x)\,dx$$



Cumulative density function

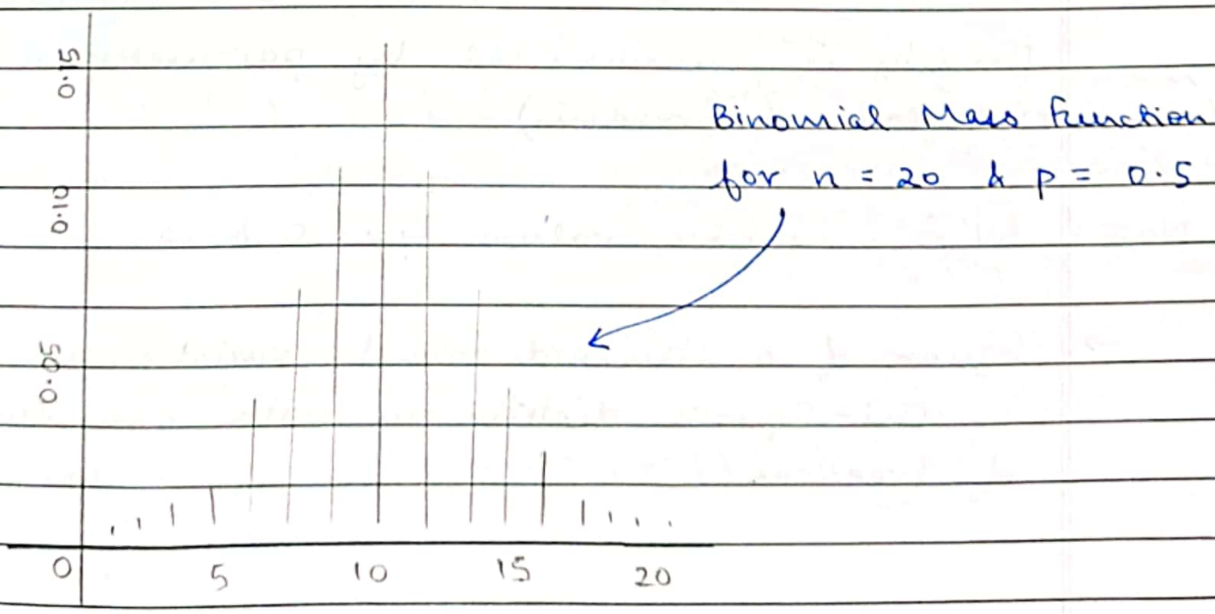
$$F(b) = P(-\infty < x < b) = \int_{-\infty}^{b} f(x)\,dx$$

Binomial Mass Function →

Probability of obtaining K heads in n coin tosses with p being the probability of obtaining a head in any toss

RV 'x' represents no. of heads obtained

- Sample space : [0, 1, 2 ... n]
- $f(x) = \dfrac{n!}{K!\,(n-K)!}\, p^{K} (1-p)^{n-K}$
- One outcome : $\underbrace{HH\ldots H}_{K \text{ times}}\ \underbrace{TT\ldots T}_{n-K \text{ times}}$
- Probability Mass function characterised by one parameter p.
- For large n, it tends to a Gaussian distribution.



Gaussian or Normal Density Function

Distribution used to characterise random errors in data.

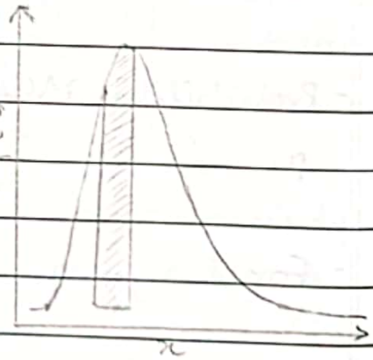
$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Probability density function is characterised by two parameters ~~##~~ μ & σ

Density function is symmetric

In case of standard normal distribution, $\mu = 0$ & $\sigma = 1$

Chi-square density function →



$$f(x) = \frac{1}{2^{n/2}\,\Gamma(n/2)} x^{n/2-1} e^{-x/2}$$

Density is characterised by parameter n (degrees of freedom)

NOTE: RV 'x' only takes values b/w 0 & ∞.

→ Square of a standard normal variable will be a Chi-Square distribution with one degree of freedom.

E stands for Expectation

classmate

Date

Page

Moments of a Probability Density Function →

Similar to describing a function using derivatives, a PDF can be described by its moments.

For continuous distributions →

$$E[x^k] = \int_{-\infty}^{\infty} x^k f(x)\, dx$$

For discrete distributions →

$$E[x^k] = \sum_{i=1}^{N} x_i^k\, p(x_i)$$

Mean (or first moment) : $\mu = E[x]$

Variance : $\sigma^2 = E[(x-\mu)^2] = E[x^2] - \mu^2$

Standard deviation = Square root of variance = σ

Properties of Gaussian RVs →

centrality parameter

Mean : $E[x] = \mu$ (value at which the density function attains the highest probability)

Variance : $E[(x-\mu)^2] = \sigma^2$ (how wide the distribution will be)

→ width of the distribution

Symbolically, $x \sim N(\mu, \sigma^2)$

→ Standard Gaussian RV $z \sim N(0,1)$

Scanned with CamScanner

→ If $x \sim N(\mu, \sigma^2)$ and $y = ax + b$
Then $y \sim N(a\mu + b, a^2\sigma^2)$

→ Standardisation
If $x \sim N(\mu, \sigma^2)$,
Then $z = \frac{(x-\mu)}{\sigma} \sim N(0,1)$

Computation of Probability using R

- Function to compute probability given a value X
- Lower tail probability $= P(-\infty < x < X)$

$$= \int_{-\infty}^{X} f(x)\,dx$$

- Functions pnorm (X, mean, std, 'lower.tail' = TRUE/FALSE)

If lower.tail = TRUE, it will give area b/w $-\infty$ and X

If lower.tail = ~~TRUE~~ FALSE, it will give area b/w X and ∞

Default value is true.

Other functions in R →

→ Function to compute X given probability p classmate

Date

Page

- Function qnorm(p, mean, std, 'lower.tail' = TRUE/FALSE)

- lower tail probability $= P(-\infty < x < X)$

$$= \int_{-\infty}^{X} f(x)\, dx = p$$

(Inverse probability function)

→ Function dnorm to compute density function value

→ Function rnorm to generate random numbers from the distribution

Joint Probability Distribution Function of two RVs →

Joint PDF of two RVs x & y : $f(x,y)$

- $P(x \leq a, y \leq b) = \int_{-\infty}^{b} \int_{-\infty}^{a} f(x,y)\, dx \cdot dy$

- Covariance b/w x & y : $\sigma_{xy} = E[(x-\mu_x)(y-\mu_y)]$

- Corelation b/w x & y : $\rho_{xy} = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y}$

- Two RVs x & y are uncorrelated if $\sigma_{xy} = 0$

- Two RVs x & y are independent if ~~$f(x,y) =$~~ $f(x,y) = f(x) f(y)$

Multivariate Normal Distribution →

→ A vector of RVs $x = [x_1 \; x_2 \; \cdots \; x_n]^T$

→ Multivariate Gaussian distribution : $x \sim N(\mu, \Sigma)$

- $E[x] = \mu$: Mean vector
- $E[(x-\mu)(x-\mu)^T] = \Sigma$: Variance - covariance matrix
- $f(x) = \dfrac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-1/2 (x-\mu)^T \Sigma^{-1} (x-\mu)}$: PDF

→ Structure of Σ

covariance b/w x_1 & x_2

variance of x_1 variance of x_2 covariance b/w x_1 & x_n

$$\Sigma = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1 x_2} & \cdots & \sigma_{x_1 x_n} \\ \sigma_{x_2 x_1} & \sigma_{x_2}^2 & \cdots & \cdots \\ \vdots & \vdots & & \vdots \\ \sigma_{x_n x_1} & \cdots & \cdots & \sigma_{x_n}^2 \end{bmatrix}$$

variance of x_n

Population : Set of all possible outcomes of a random experiment characterised by $f(x)$.

Sample set (realization) : Finite set of observations obtained through an experiment.

Inference : Conclusions derived regarding the population (PDF, parameters) from the sample set.

Descriptive Statistics (Analysis) -

(i) Graphical
Eg : box plots, probability plots

(ii) Numerical
Eg : Mean, mode, range, variance, moments

Inferential -

(i) Estimation

(ii) Hypotheses testing

Measures of Central Tendancy →

- Mean →

Best estimate in least squares criterion
Unbiased estimate of population mean : $E[\bar{x}] = \mu$
Affected by outliers

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

- Median →

Represent sample set by a single value
Value of x_i such that 50% of the values are less than x_i and 50% of the observations are greater than x_i
Robust wrt outliers in data
Best estimate in least absolute deviation sense

Date
Page

Mode →

Represents sample set by a single value
Value that occurs most often

Measures of spread -

Represents spread of sample set.

Sample variance : $s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$

Susceptible to outliers

- Unbiased estimate of population variance:
  $E[s^2] = \sigma^2$

- Standard deviation is square root of variance

Mean absolute deviation (MAD) : $\bar{d} = \frac{1}{N} \sum_{i=1}^{N} |x_i - \bar{x}|$

More robust measure (less susceptible to outliers)

Range : $R = x_{max} - x_{min}$

Sample Mean →
For any distribution, sample mean is an unbiased estimate of population mean.

If $x_i \sim N(\mu, \sigma^2)$ & all observations are mutually independent, then $\bar{x} \sim N\left(\mu, \frac{\sigma^2}{N}\right)$

Scanned with CamScanner

Sample Variance →

For any distribution, sample variance is an unbiased estimate of the population variance

If $x_i \sim N(\mu, \sigma^2)$ and all observations are mutually independent,

Then $\frac{(N-1)s^2}{\sigma^2} \sim \chi^2_{N-1}$

s^2 → sample variance

σ^2 → population variance

↳ chi-square distribution with N-1 degrees of freedom.

Graphical Analysis →

(i) Histogram

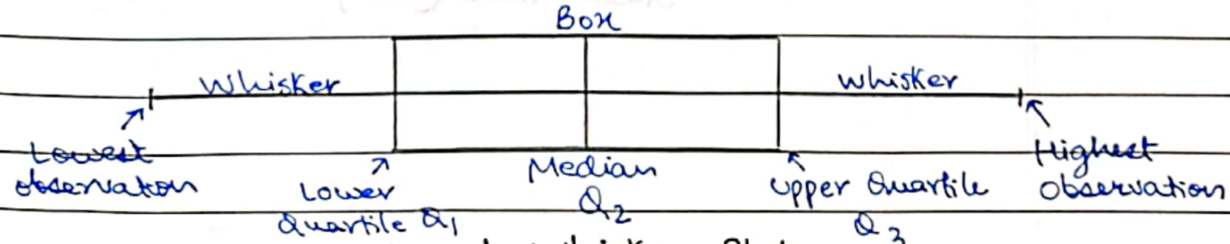
~~* include upper limit of interval in that interval~~

~~* don't include the value in next interval where it is lower limit~~

(ii) Box plot →

Find quartiles (Q_1, Q_2 and Q_3), minimum and maximum.

Box is b/w Q_1 and Q_3, and whiskers is b/w min & max values.

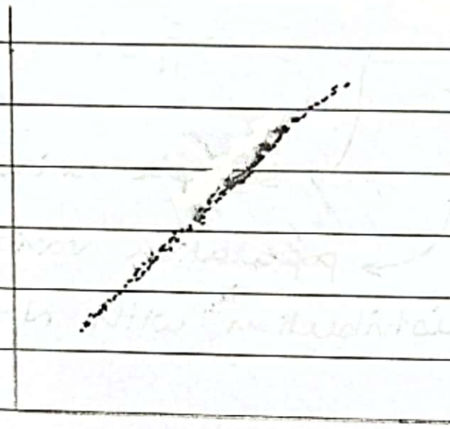


Box & whisker Plot

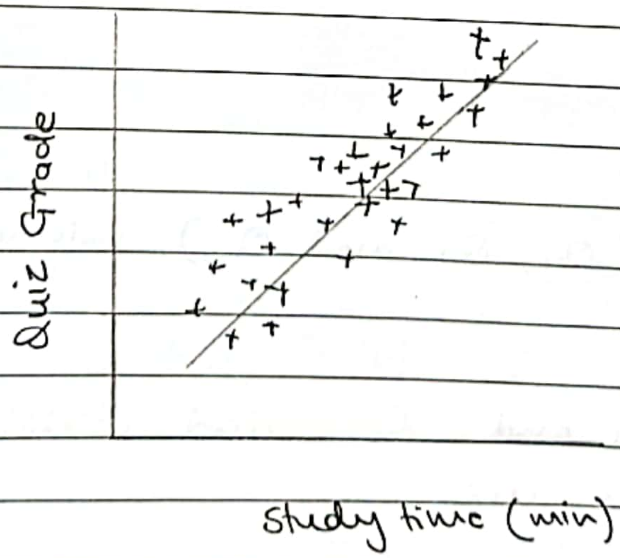
Date
Page

(iii) Probability plot →
(p-p or q-q plot)

If the points fall on the 45° line, they come from the appropriate distribution



(iv) Scatter Plot →
It plots one random variable against another to examine if there is any dependance



Hypothesis Testing →

The hypothesis is generally converted to a test of the mean or variance parameter of a population (or differences in means or variances of populations).

A hypothesis is a statement or postulate about the parameters of a distribution (or model).

Null Hypothesis H_0 : The default or states quo postulate that we wish to reject if the sample set provides different sufficient evidence (eg: $\eta = \eta_0$).

Alternative Hypothesis H_1 : The alternative postulate that is accepted if the null hypothesis is rejected (eg : $\eta < \eta_0$)

No hypothesis test is perfect. There are inherent errors since it is based on observations which are random

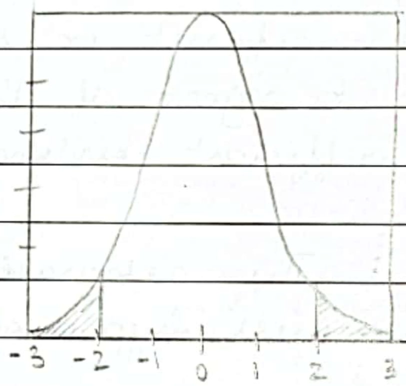
The performance of a hypothesis test depends on

- Extent of variability in data
- No. of observations (Sample size)
- Test statistic (function of observations)
- Test criterion (threshold)

Two sided test

$H_0 : \mu = 0$
$H_1 : \mu \neq 0$

Test statistic standard normal RV z.

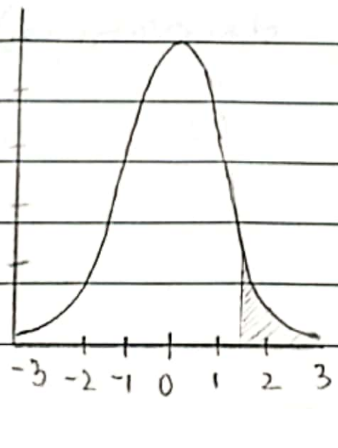


Reject H_0 if $z \leq -2$ or $z \geq 2$

One sided test

$H_0 : \mu = 0$
$H_1 : \mu > 0$

Test statistic standard normal RV z.



~~Reject H_0 if $z \geq 1.5$~~

Statistical Power = 1 - Type-II error probability

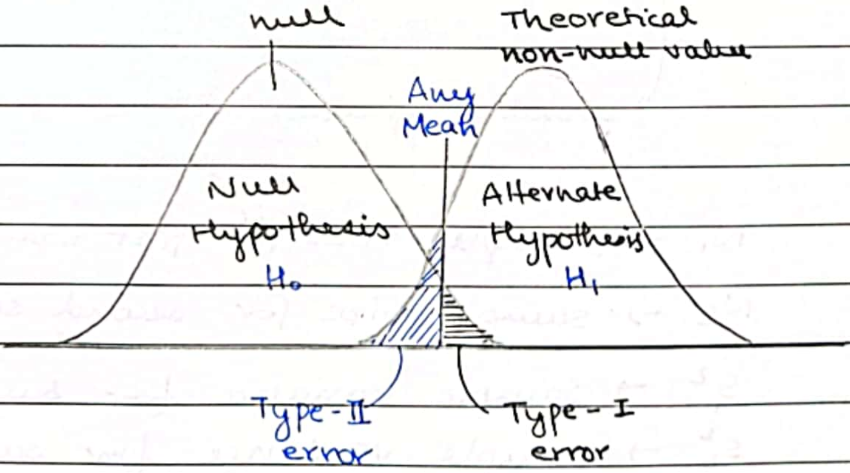
Trade-off : If we decrease Type-I error probability, then Type-II error probability will increase.

Errors in Hypothesis Testing →

Decision → Truth ↓	H_0 is not rejected	H_0 is rejected
H_0 is true	Correct decision $Pr = 1 - \alpha$	Type-I error $Pr = \alpha$
H_1 is true	Type-II error $Pr = \beta$	Correct Decision $Pr = 1 - \beta$

We can only control the Type-I error probability, not the Type-II error probability.

→ 'α' is also k/a level of significance of the test.



Test statistic (Z) → $\dfrac{\bar{x} - \mu}{SD / \sqrt{NOS}}$

SD = Standard Deviation

NOS = No. of samples taken from the total population

T - Test →

It is a type of inferential statistic used to determine if there is a significant difference b/w the mean of two groups, which may be related in certain features.

Test statistic (assuming unknown but equal variances for two groups) –

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{S_p^2}{N_1} + \frac{S_p^2}{N_2}}} \sim t_{N_1+N_2-2}$$

where, $$S_p = \frac{(N_1-1)\,S_1^2 + (N_2-1)\,S_2^2}{N_1+N_2-2}$$

→ pooled variance

N_1 → sample size for first sample

N_2 → sample size for second sample

S_1^2 → sample variance for sample 1.

S_2^2 → sample variance for sample 2.

Summary of Useful Hypothesis Tests →

Type of test	Characteristic	Example	Application
z-test	Sum of independent normal variables	Test for a mean or comparison between two group means (variance known)	Test coefficients of a regression model
t-test	Ratio of a standard normal variable and chi-square variables with p degrees of freedom	Test for a mean or comparison between two group means (variance unknown)	Test coefficients of a regression model
chi-square test (p degrees of freedom)	Sum of p independent standard normal variables	Test for variance	Test quality of regression model
F-test (p_1 and p_2 degrees of freedom)	Ratio of two chi-square variables	Test for comparing variances of two groups	Choose between regression models having different number of parameters