UNIT - 2

Date: Page:

Chapter 1 → Regression

Regression Analysis
It is a mathematical measure of average relationship b/w two or more variables in terms of the original units of the data

Two types of variables :
i Dependant Variable — The variable whose value is influenced or is to be predicted
ii Independant Variable — The variable which influences the value or used for predition

* Dependant variable is also called regressed or ~~predictor~~ or explained variable
* Independant variable is also called regressor or preditor or explanatory variable.

Linear Regression
→ It has been a critical driving force behind many AI and data science applications
→ It is useful in business as it is simple, interpretable, efficient method to evaluate trends and make future estimates

Types
2 types of linear regression

i Simple Linear
→ It reveals correlation between dependant variable and an independant varible

→ It describes relationship strength between the given variables. Eg :- Relationship b/w pollution levels and rising temperature
→ The value of dependant variable is based on the value of independent variable.
Eg :- The value of pollution level at a Specific temperature.

ii Multiple Linear
→ It establishes the relationship b/w two or more independent variables and one dependant variable
→ Independant variable can be contiguous or categorical.
→ It helps foresee trends, determine future values, predict the impacts of changes.
→ Eg :- Calculating blood pressure (dependent) by considering height, weight, amount of exercise (independent).

Logistic Regression
→ Also known as logit model
→ Applicable where there is one dependent variable and more independent variable
→ The target value is discrete (binary or an ordinal value)
→ Eg :- Determining the likelihood of choosing an offer on your website (dependant) By considering sites customer came from, count of visits to the site, activity of the site (independent)

Types

2 types of Logistic

i Ordinal Regression

→ It involves one dependent opposite variable (yes, no / true, false) and one independent which can be ordinal or nominal

→ It interacts b/w dependent variable with multiple ordered levels with one or more independent variables.

→ Eg :- Survey where respondents must either agree or disagree.

ii Multinomial Logistic

→ It is performed when dependent variable is nominal with more than two levels.

→ It specifies relationship b/w one dependent variable and one or more continuous level independent variable.

→ Eg :- Used to model program choice made by school students based on vocational, sports, academic, extracurricular activity.

Lines of Regression

→ If variables in bivariate distribution are related, we find that points in scatter diagram will cluster round some curve called "curve of regression".

→ If curve is a straight line, then it is called the "line of regression" and there is said to be linear regression otherwise the regression is curvilinear regression

Page:

→ It gives the best estimate to the value of one variable for any specific value of the other variable.
→ Line of Regression is the line of "best fit".
→ It is based on principles of least square
→ Line of Regression are also called regression lines
→ These lines are used in regression analysis to provide insights into the nature and strength of associations b/w dependent and independent variables.

Types.
→ Regression Line are of two types
→ The two equations are not reversible or interchangeable because of assumptions for deriving these equations are different.

i Regression Line of Y on X (Y given X)

$$Y = a + bX.$$

→ This line predicts the value of dependent variable (Y) based on the values of independent variable (X)
→ a is y-intercept, predicted value of Y when X = 0
→ b is slope, the change in the mean of Y for a one unit change in X.

ii. Regression Line of X on Y (x given Y)

$$X = c + dY$$

→ This line predicts the value of independent variable (X) based on the values of dependent variable (Y)

→ c is x-intercept, predicted value of X when Y = 0

→ d is slope, the change in the value of X for a one unit change in Y.

Correlation Coefficient

→ Denoted by r.

→ It quantifies the strength and direction of a linear relationship b/w two variables.

→ It ranges from −1 to 1

$r = 1$ → Perfect positive linear relationship
$r = -1$ → Perfect negative linear relationship
$r = 0$ → No linear relationship

→ It helps to assess the degree to which changes in one variable correspond to changes in another, indicating linear relation

Properties.

1. Geometric Mean between regression coefficient
2. If one coefficient is greater than 1 then other must be less than 1
3. Arithmetic mean of regression coefficient is greater than correlation coefficient r
4. Regression coefficient are independent of the change of origin but not of scale.

Formulas used in this

1. Regression Coefficient of Y on X

$$Y = a + bX.$$

$$b_{yx} = \frac{n \sum XY - \sum X \sum Y}{n \sum X_2 - (\sum X)^2}$$

$$b_{yx} = \frac{Cov(X, Y)}{\sigma_x^2}$$

$$b_{yx} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2}$$

$$b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x} \Rightarrow \frac{\sum xy}{\sum x^2}$$

2. Regression Coefficient of X on Y

$$X = a + bY$$

$$b_{xy} = \frac{n \sum XY - \sum X \cdot \sum Y}{n \sum Y^2 - (\sum Y)^2}$$

$$b_{xy} = \frac{Cov(X, Y)}{\sigma_y^2}$$

$$b_{xy} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (Y - \bar{Y})^2}$$

$$b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y} \Rightarrow \frac{\sum xy}{\sum y^2}$$

3. Regression Eq. of Y on X

$$(Y-\bar{Y}) = b_{yx}(X-\bar{X})$$

$$(Y-\bar{Y}) = r \cdot \frac{\sigma_y}{\sigma_x}(X-\bar{X})$$

4. Regression Eq. of X on Y

$$(X-\bar{X}) = b_{xy}(Y-\bar{Y})$$

$$(X-\bar{X}) = r \cdot \frac{\sigma_x}{\sigma_y}(Y-\bar{Y})$$

5. Relation of 'r'

$$r^2 = b_{yx} \times b_{xy}$$

$$r = \sqrt{b_{yx} \cdot b_{xy}}$$

E1. In a partial destroyed lab record of an analysis correlation data, the following results are only legible: variance of $x = 9$, regression eqn :- $8x - 10y + 66 = 0$, $40x - 18y = 214$. What are :-.

$$(8x - 10y + 66 = 0) \times 5$$
$$40x - 18y = 214$$

$$\cancel{40x} - 50y = -330$$
$$\cancel{40x} - 18y = 214$$
$$- \quad + \quad -$$
$$-32y = -544$$

$\bar{y} = \frac{7544}{432}$

$= 17.$

$$\boxed{\bar{y} = 17}$$

$x = \frac{10y - 66}{8} \Rightarrow \frac{10 \times 17 - 66}{8}$

$$\boxed{\bar{x} = 13}$$

(a) Mean of x and y.

$\bar{x} = 13$, $\bar{y} = 17.$

(b) S.D of y. (c). coeff. of correl b/w x & y

(c) → Solⁿ

$8x - 10y = -66$

$66 + 8x = 10y$

$\frac{66}{10} + \frac{8}{10}x = y.$

$$\boxed{b_{yx} = \frac{8}{10}} \Rightarrow 0.8$$

$40x - 18y = 214$

$x = \frac{214}{40} + \frac{18y}{40}.$

$$\boxed{b_{xy} = \frac{18}{40}} \Rightarrow 0.45.$$

$r = \sqrt{b_{xy} \times b_{xy}}$
$= \sqrt{0.45 \times 0.8}$
$= \sqrt{0.36}$
$= \pm 0.6$.

$\boxed{r = 0.6}$

(b) → Sol^n

$\sigma_y = \dfrac{b_{yx} \cdot \sigma_x}{r}$

$= \dfrac{0.8 \times 3}{0.6}$ as $\boxed{\sigma_x^2 = 9}$

$= 4$.

$\boxed{\sigma_y = 4}$

Ex2. Find most likely price in Bombay corresponding to the price of Rs 70 at Calcutta.

	Calcutta {X}	Bombay {Y}
Av. Price	65	67
S.D.	2.5	3.5

Correlation coefficient b/w price of commodities in two cities is 0.8.

$\bar{X} = 65$ $\bar{Y} = 67$
$\sigma_x = 2.5$ $\sigma_y = 3.5$.
$r(X, Y) = 0.8$.

Problem :- Line of Regression of Y on X = 70.

$$y - \bar{y} = r \cdot \frac{\sigma y}{\sigma x} (x - \bar{x})$$

$$y - 67 = 0.8 \times \frac{3.5}{2.5} (x - 65)$$

$$y = 67 + 1.12 \cdot (70 - 65)$$

$$= 67 + 1.12 \times 5$$

$$= 67 + 5.6$$

$$= 72.6.$$

E3. Calculate regression coefficient and obtain line of regression for the following data.

X	1	2	3	4	5	6	7
Y	9	8	10	12	11	13	14.

	X	Y	X^2	Y^2	XY
	1	9	1	81	9
	2	8	4	64	16
	3	10	9	100	30
	4	12	16	144	48
	5	11	25	122	55
	6	13	36	169	78
	7	14	49	196	98
Total	28	77	140	875	334

$$\bar{X} = \frac{28}{7} \Rightarrow 4 \qquad \bar{Y} = \frac{77}{7} \Rightarrow 11$$

(i) Regression Coefficient of X on Y

$$b_{xy} = \frac{n \sum XY - \sum X \sum Y}{n \sum Y^2 - (\sum Y)^2}$$

$$= \frac{7(334) - (28)(77)}{7(875) - (77)^2}$$

$$= \frac{2338 - 2156}{6125 - 5929}$$

$$= \frac{182}{196}$$

$$= 0.929$$

(ii) Regression Eq^n of X on Y

$$X - \bar{X} = b_{xy} (Y - \bar{Y})$$

$$X - 4 = 0.929 (Y - 11)$$

$$X - 4 = 0.929Y - 10.219.$$

(iii) Regression Coefficient of Y on X

$$b_{yx} = \frac{n \sum YX - \sum X \sum Y}{n \sum X^2 - (\sum X)^2}$$

$$= \frac{7(334) - (28)(77)}{7(140) - (28)^2}$$

$$= \frac{2338 - 2156}{980 - 784}$$

$$= \frac{182}{196}$$

$$= 0.929$$

(iv) Regression Eq^n of Y on X

$$Y - \bar{Y} = b_{yx} (X - \bar{X})$$
$$Y - 11 = 0.929 (X - 4)$$
$$Y - 11 = 0.929 X - 3.716$$

E4. Find linear regression equation for the following two sets of data.

	X	Y	X^2	Y^2	XY
	2	3	4	9	6
	4	7	16	49	28
	6	5	36	25	30
	8	10	64	100	80
Total	20	25	120	183	144

Linear Regression Eq^n $\boxed{Y = a + bX.}$

$$b = \frac{n \Sigma XY - \Sigma X \Sigma Y}{n \Sigma X^2 - (\Sigma X)^2}$$
$$= \frac{4 \times 144 - (20)(25)}{4 \times 120 - (20)^2}$$
$$= \frac{576 - 500}{480 - 400}$$
$$= \frac{76}{80}$$
$$= 0.95.$$

$$a = \frac{\Sigma Y \Sigma X^2 - \Sigma X \Sigma XY}{n \Sigma X^2 - (\Sigma X)^2}$$

$$a = \frac{25 \times 120 - 20 \times 144}{4 \times 120 - 400}$$

$$= \frac{3000 - 2880}{480 - 400}$$

$$= \frac{12\cancel{0}}{8\cancel{0}}$$

$$= 1.5$$

$$y = a + bX$$

$$\boxed{y = 1.5 + 0.95x}$$

E5. Calculate two regression equations of X on Y and Y on X from given data, taking deviation from actual means of X and Y. Also estimate the likely demand when price is Rs 20.

Price	Demanded Amount	x	x^2	y	y^2	xy
10	40	-3	9	-1	1	3
12	38	-1	1	-3	9	3
13	43	0	0	2	4	0
12	45	-1	1	4	16	-4
16	37	3	9	-4	16	-12
15	43	2	4	2	4	4
78	246.	0	24	0	50	-6

Mean of $x = \frac{78}{6} \Rightarrow 13$.

Mean of $y = \frac{246}{6} = 41$

(ii) Regression eqn of X on Y

$$b_{xy} = \frac{\Sigma xy}{\Sigma y^2} \Rightarrow \frac{-6}{50} \Rightarrow -0.12$$

$$X - \bar{X} = b_{xy} (Y - \bar{Y})$$
$$X - 13 = -0.12 (Y - 41)$$
$$X - 13 = -0.12Y + 4.92$$
$$X = -0.12Y + 17.92$$

(iii) Regression eqn of Y on X

$$b_{yx} = \frac{\Sigma xy}{\Sigma x^2} \Rightarrow \frac{-6}{24} \Rightarrow -0.25$$

$$Y - \bar{Y} = b_{yx} (X - \bar{X})$$
$$Y - 41 = -0.25 (X - 13)$$
$$Y - 41 = -0.25X + 3.25$$
$$Y = -0.25X + 44.25$$

(iv) When price is Rs 20 ie $X = 20$.

$$Y = -0.25 \times 20 + 44.25$$
$$= -5 + 44.25$$
$$= 39.25$$

E6. Find mean of X and Y and coefficient of correlation b/w them from regression eqns.

$$2Y - X - 50 = 0$$
$$3Y - 2X - 10 = 0$$

$$\left\{\begin{array}{l}(2y - x - 50 = 0) \times 2 \\ 3y - 2x - 10 = 0\end{array}\right.$$

$$4y - \cancel{2x} = 100$$
$$\underline{\overset{-}{3}y \overset{+}{-} \cancel{2x} = \overset{-}{10}}$$
$$y = 90$$

$$2 \times 90 - x = 50$$
$$-x = 50 - 180$$
$$x = 130$$

$\boxed{\bar{x} = 130}$ $\boxed{\bar{y} = 90}$

Eqn of y on x

$$2y - x - 50 = 0$$
$$2y = x + 50$$
$$y = \frac{x}{2} + 25$$

$$b_{yx} = \frac{1}{2} \Rightarrow 0.5$$

Eqn of x on y

$$3y - 2x - 10 = 0$$
$$-2x = -3y + 10$$
$$2x = 3y - 10$$
$$x = \frac{3}{2}y - 5$$

$$b_{xy} = \frac{3}{2} \Rightarrow 1.5$$

$$r = \sqrt{1.5 \times 0.5} \Rightarrow \sqrt{0.75}$$
$$\Rightarrow 0.866$$

Chapter 2 → Discrete Distributions

Probability Distribution

→ They are mathematical functions that describe the likelihood of different outcomes in a random experiment.

→ They are used for understanding, modeling uncertainty in various real-world scenes

Probability Distribution

- **Discrete**
  - → Binomial
  - → Bernoulli
  - → Poisson
- **Continuous**
  - → Uniform
  - → Exponential
  - → Normal

Discrete Distribution

→ The random variable can only take on distinct, separate values.

→ These value often represent outcomes of events, and probability of each outcome is assigned to specific value.

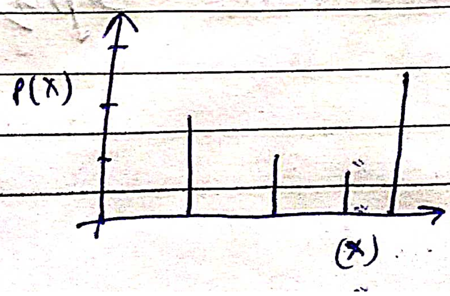
→ Used when dealing with countable outcomes or events with clear boundary

→ It has three basic types

(a) Bernoulli

(b) Binomial

(c) Poisson



Bernoulli Distribution

→ It models a single binary experiment with two possible outcome either success / failure

→ It serves as foundation for more complex distribution

→ Named after Jacob Bernoulli who introduced it in 18th century.

→ Event X → success $= 1$

$P(X=1) = p$

→ failure $= 0$

$P(X=0) = q \Rightarrow 1-p$.

→ $p + q = 1$. "always"

Mean and Variance and PMF

- Probability mass function $= \begin{cases} q = 1-p & \text{if } x = 0 \\ p & \text{if } x = 1 \end{cases}$

$$PMF = p^x (1-p)^{1-x}$$

- Mean $(\mu) = p$
  = probability of success

- Variance $(\sigma^2) = pq$
  $= P(\text{success}) * P(\text{failure})$

- SD $= \sqrt{\sigma^2}$
  $= \sqrt{pq}$

E1. A basketball player can shoot a ball into the basket with probability 0.6. What is the probability that he misses the shots?

$P(\text{success}) = p$
$= 0.6$
$P(\text{failure}) = 1 - p$
$= 1 - 0.6$
$= 0.4$

E2. If Bernoulli distribution has a parameter 0.45 then find its mean.

$P(\text{success}) = 0.45$
$\text{Mean} = p$
$= 0.45$

E3. If a Bernoulli distribution has a parameter 0.72 then find its variance

$P(\text{success}) = 0.72$
$P(\text{failure}) = 1 - 0.72$
$= 0.28$

$\text{Variance} = pq$
$= 0.72 \times 0.28$
$= 0.2016$

E4. Approx. 1 in 200 American adults are lawyers. One American adult is randomly selected. What is the distribution of the no. of lawyers.

$P(x=1) = \frac{1}{200}$ = success ie lawyer

$P(x=0) = \frac{199}{200}$ = failure ie not lawyer

Binominal Distribution

→ It gives only two possible results in an experiment either success or failure.
→ Two parameters :- n and p are used.
→ 'n' → no of times experiment execute
→ 'p' → probability of any one outcome
→ All the trials are independent
→ Number of trials 'n' is finite

Probability Mass Function

$$P(x) = {}^{n}C_{x}\, p^{x}\, q^{n-x}$$

x :- experiment to be found out.
n :- no. of times

E1 Find probability of head occuring 2 times while tossing a coin thrice.

$$P(x=2) = {}^{3}C_{2}\left(\frac{1}{2}\right)^{2}\left(\frac{1}{2}\right)^{3-2}$$

$$= 3 \times \frac{1}{4} \times \frac{1}{2}$$

$$= \frac{3}{8}.$$

E2. Show that $P(X)$ is PMF.
To prove this, PMF must have two condition satisfying — (a) $P(X) \geq 0$
(b) $\sum_{x=0}^{n} P(X) = 1.$

$$\sum_{X=0}^{n} P(X) = \sum_{X=0}^{n} {}^{n}C_{x} \cdot p^{x} \cdot q^{n-x}$$

$$= q^{n} + {}^{n}C_{1} p^{1} q^{n-1} + \dots p^{n}$$

by binominal expansion

$$= (p+q)^{n}$$

$$= 1^{n}$$

$$= 1.$$

Hence; $P(X)$ is PMF.

Mean and Variance

$$P(X) = {}^{n}C_{x} \, p^{x} \, q^{n-x}$$

- Mean $(\mu) = E(x) = \sum_{X=0}^{n} X P(X)$

$$\boxed{\mu = np.}$$

- Variance $(\sigma^2) = E(x^2) - (E(x))^2$

$$\boxed{\sigma^2 = npq}$$

- $S.D = \sqrt{\sigma^2}$

$$\boxed{S.D = \sqrt{npq}}$$

E1. If a coin is tossed 5 times, find probability

(a) exactly 2 heads.

$n = 5$, $p = \frac{1}{2}$, $q = \frac{1}{2}$, $x = 2$

$$P(x=2) = \cancel{50} \times {}^5C_2 \cdot \left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right)^{5-2}$$

$$= \frac{5!}{2!\,3!} \cdot \frac{1}{4} \cdot \frac{1}{8}$$

$$= \frac{5 \times \cancel{4} \times \cancel{3} \times \cancel{2}}{\cancel{2} \times \cancel{3} \times \cancel{2}} \cdot \frac{1}{4} \cdot \frac{1}{8_4}$$

$$= \frac{5}{16}.$$

(b) Atleast 4 heads.

$n = 5, \quad p = \frac{1}{2}, \quad q = \frac{1}{2}, \quad x = 4, 5.$

$$P(x=4) = {}^5C_4 \times \left(\frac{1}{2}\right)^4 \times \left(\frac{1}{2}\right)^{5-4}$$

$$= \frac{5!}{4!\,1!} \times \frac{1}{16} \times \frac{1}{2}$$

$$= \frac{5 \times \cancel{4 \times 3 \times 2}}{\cancel{4 \times 3 \times 2}} \times \frac{1}{16} \times \frac{1}{2}$$

$$= \frac{5}{32}$$

$$P(x=5) = {}^5C_5 \left(\frac{1}{2}\right)^5 \cdot \left(\frac{1}{2}\right)^{5-5}$$

$$= 1 \times \frac{1}{32}$$

$$= \frac{1}{32}.$$

$$P(\text{atleast 4 head}) = \frac{5}{32} + \frac{1}{32}$$

$$= \frac{6}{32} \Rightarrow \frac{3}{16}.$$

Page:

(c) Getting atmost 2 heads.

$n = 5, \; p = \frac{1}{2}, \; q = \frac{1}{2}, \; x = 0, 1, 2$

$P(x=0) = {}^5C_0 \cdot \left(\frac{1}{2}\right)^0 \cdot \left(\frac{1}{2}\right)^{5-0}$

$= \frac{1}{32}.$

$P(x=1) = {}^5C_1 \cdot \left(\frac{1}{2}\right)^1 \cdot \left(\frac{1}{2}\right)^{5-1}$

$= 5 \times \frac{1}{2} \times \frac{1}{16}$

$= \frac{5}{32}$

$P(x=2) = {}^5C_2 \cdot \left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right)^{5-2}$

$= \frac{5 \times \cancel{4} \times \cancel{3} \times \cancel{2}}{\cancel{3} \times \cancel{2} \times \cancel{2}} \times \frac{1}{4} \times \frac{1}{\cancel{8}_4}$

$= \frac{5}{16}.$

$P(\text{getting atmost 2 head}) = \frac{1}{32} + \frac{5}{32} + \frac{5}{16}$

$= \frac{16}{32}$

$= \frac{1}{2}.$

Ex2. A fair coin is tossed 10 times, what are the probability of getting.

(a) exactly 6 heads.

$n = 10, \; p = \frac{1}{2}, \; q = \frac{1}{2}, \; x = 6.$

Page:

$$P(x=6) = {}^{10}C_6 \cdot \left(\frac{1}{2}\right)^6 \cdot \left(\frac{1}{2}\right)^{10-6}$$

$$= \frac{\cancel{10}^{5} \times \cancel{9}^{3} \times \cancel{8} \times 7}{\cancel{4} \times \cancel{3} \times \cancel{2}} \times \frac{1}{\cancel{64}} \times \frac{1}{\cancel{16}\,8}$$

$$= \frac{105}{512}$$

(b) Atleast six heads.

$n=10$, $p=\frac{1}{2}$, $q=\frac{1}{2}$, $x=6,7,8,9,10$.

$P(x=6) = {}^{10}C_6 \left(\frac{1}{2}\right)^{10}$ $P(x=9) = {}^{10}C_9 \left(\frac{1}{2}\right)^{10}$

$P(x=7) = {}^{10}C_7 \left(\frac{1}{2}\right)^{10}$ $P(x=10) = {}^{10}C_{10} \left(\frac{1}{2}\right)^{10}$.

$P(x=8) = {}^{10}C_8 \left(\frac{1}{2}\right)^{10}$

$$P(\text{atleast six heads}) = \left(\frac{1}{2}\right)^{10} \left({}^{10}C_6 + {}^{10}C_7 + {}^{10}C_8 + {}^{10}C_9 + {}^{10}C_{10}\right)$$

$$= \left(\frac{1}{2}\right)^{10} \times (210 + 120 + 45 + 10 + 1)$$

$$= \left(\frac{1}{2}\right)^{10} \times 386$$

$$= \frac{193}{512}.$$

E3 60% of people who purchase sports cars are men. Find probability exactly 7 are men if 10 sports car owners are selected randomly

$n=10$, $x=7$, $p=\frac{6}{10}$ $q=\frac{4}{10}$.

$$P(x=7) = {}^{10}C_7 \times \left(\frac{6}{10}\right)^7 \times \left(\frac{4}{10}\right)^{10-7}$$

$$= \frac{10 \times \cancel{9}^3 \times \cancel{8}^4}{\cancel{3} \times \cancel{2}} \times \left(\frac{6}{10}\right)^7 \times \left(\frac{4}{10}\right)^3$$

$$= 120 \times 0.0279936 \times 0.064.$$

$$= 0.215$$

E4 The probability that a person can achieve a target is 3/4. The count of tries is 5. What is probability that he will attain the target atleast thrice.?

$n = 5$, $x = 3, 4, 5$, $p = \frac{3}{4}$, $q = \frac{1}{4}$

$$P(x=3) = {}^5C_3 \cdot \left(\frac{3}{4}\right)^3 \cdot \left(\frac{1}{4}\right)^{5-3}$$

$$= \frac{5 \times \cancel{4}}{2} \times \frac{3 \times 3 \times 3}{\cancel{4} \times 4 \times 4} \times \frac{1}{16}.$$

$$= \frac{135}{512}$$

$$P(x=4) = {}^5C_4 \cdot \left(\frac{3}{4}\right)^4 \cdot \left(\frac{1}{4}\right)^{5-4}$$

$$= 5 \times 0.308025 \times 0.25$$

$$= 0.385$$

$$= \frac{405}{1024}$$

$$P(x=5) = {}^5C_5 \left(\frac{3}{4}\right)^5 \cdot \left(\frac{1}{4}\right)^{5-5}$$

$$= \frac{243}{1024}$$

P(target atleast three) $= \frac{135}{512} + \frac{405}{1024} + \frac{243}{1024}$

$= \frac{459}{512}$.

E5. A and B play a game in which their chance of winning are in the ratio 3:2. Find A's chance of winning atleast three games out of five games played.

$n = 5$, $x = 3, 4, 5$, $p = \frac{3}{5}$, $q = \frac{2}{5}$.

$P(x=3) = {}^5C_3 \cdot \left(\frac{3}{5}\right)^3 \cdot \left(\frac{2}{5}\right)^{5-3}$

$= \frac{\cancel{5} \times \cancel{4}^{2}}{\cancel{2}} \cdot \frac{27}{125} \cdot \frac{4}{\cancel{25}\,5}$

$= \frac{216}{625}$.

$P(x=4) = {}^5C_4 \cdot \left(\frac{3}{5}\right)^4 \cdot \left(\frac{2}{5}\right)^1$

$= \cancel{5} \times \frac{81}{625} \times \frac{2}{\cancel{5}}$

$= \frac{810}{625}$. $\frac{162}{625}$

$P(x=5) = {}^5C_5 \cdot \left(\frac{3}{5}\right)^5 \cdot \left(\frac{2}{5}\right)^0$

$= \frac{243}{3125}$.

P(winning atleast 3 games) $= \frac{216}{625} + \frac{162}{625} + \frac{243}{3125}$

$= 0.68$

Poisson Distribution

→ It is a limiting case of Binominal distribution
→ Discovered by Simeon Denis Poisson in 1837
→ The events are independent
→ It is limited when the no. of trials n is indefinitely large.
→ In this, the trials taken place are extra large and the success rate is quite low.

Probability Mass Function

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad x = 0, 1, 2, 3 \ldots\ldots, \infty$$

E1. Prove that poisson distribution is a limiting case of binominal distribution under the following conditions → (i) $n \to \infty$
(ii) $p \to 0$
(iii) $np = \lambda$ (finite)

$$\lim_{n\to\infty} p(x)$$

$$\lim_{n\to\infty} {}^nC_x \, p^x \, q^{n-x}$$

$$\lim_{n\to\infty} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$\lim_{n\to\infty} \frac{n(n-1)\ldots\ldots(n-(x-1))\,(n-x)!}{x!\,(n-x)!} p^x (1-p)^{n-x}$$

Now $\boxed{p = \frac{\lambda}{n}}$

$$\lim_{n\to\infty} \frac{n(n-1)\cdots\cdots(n-(x-1))\cancel{(n-x)!}}{x!\ \cancel{(n-x)!}} \left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$\lim_{n\to\infty} \frac{\cancel{n^x}\left(1-\frac{1}{n}\right)\cdots\cdots 1-\left(\frac{x-1}{n}\right)}{x!} \left(\frac{\lambda^x}{\cancel{n^x}}\right)\left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$\lim_{n\to\infty} \frac{1}{x!}\lambda^x\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-x}$$

$$\frac{\lambda^x}{x!}\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n \underbrace{\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-x}}_{\text{Value is 1.}}$$

$$\frac{\lambda^x}{x!}\cdot e^{-\lambda}$$

Mean and Variance

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad x = 0, 1, 2, \ldots \infty$$

- Mean $= E(x) = \lambda$
- Variance $= V(x) = E(x) = \lambda$.
- SD $= \sqrt{\lambda}$

→ Mean of poisson distribution is equal to variance.

Date:
Page:

E1 Given that 2% of the fuses manufactured by a firm are defective. Find probability that a box containing 200 fuses has.

(a) Atleast 1 defective fuses.

$n = 200$, $p = 0.02$, $\lambda = np \Rightarrow 4$.

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

$$P(x \geq 1) = 1 - P(x < 1)$$
$$= 1 - P(0)$$
$$= 1 - \frac{4^0 e^{-4}}{0!}$$
$$= 1 - e^4.$$

(b) 3 or more defective fuses.

$$P(x \geq 3) = 1 - P(x < 3)$$
$$= 1 - [P(0) + P(1) + P(2)]$$
$$= 1 - \left[\frac{4^0 e^{-4}}{0!} + \frac{4^1 e^{-4}}{1!} + \frac{4^2 + e^{-4}}{2!}\right]$$
$$= 1 - e^{-4}(1 + 4 + 8)$$
$$= 1 - e^4 \times 13.$$
$$= 1 - 13e^{-4}$$

(c) No defective fuses.

$$P(x = 0) = \frac{4^0 e^{-4}}{0!}$$
$$= e^{-4}.$$

E2. A random variable X has a poison distribution with parameter λ such that $P(x=1) = 0.2$ P(x=2). Find $P(x=0)$

$$P(x=1) = 0.2\, P(x=2)$$

$$\frac{\lambda^1 \cancel{e^{-\lambda}}}{1!} = 0.2 \times \frac{\lambda^2 \cancel{e^{-\lambda}}}{\cancel{2!}\ 10}$$

$$\cancel{\lambda} = \frac{\lambda^2}{10}$$

$$\lambda = 10.$$

$$\boxed{e = 2.718}$$

$$P(x=0) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$= \frac{10^0 e^{-10}}{0!}$$

$$= e^{-10}$$

$$= 0.0000454$$

E3 Telephone calls arrive at an exchange according to poisson process at a rate $\lambda = 2$/min. Calculate the probability that exactly two calls will be recieved during each of the first 5 minutes of the hour.

$$P(x=2) = \frac{2^2 e^{-2}}{2!}$$

x = no. of calls in 1 minute

$$= 2e^{-2}$$

$$P(y=5) = 2^5 e^{-10}$$

$$= 0.00145$$

y = no. of minutes

$n = 5 \quad p = 2e^{-2}$

Binomial distribution

Date:
Page:

E4. In a cafe, the customer arrives at a mean rate of 2 per min. Find probability of arrival of 5 customers in 1 minute using poisson.

$\lambda = 2$. $x = 5$.

$$P(x=5) = \frac{2^5 e^{-2}}{5!}$$

$$= \frac{32\, e^{-2}}{5!}$$

$$= 0.036$$

P (arrival of 5 customer in 1 min) = 3.6%.

E5. Find the mass probability of function at $x = 6$; if mean is 3.4.

$\lambda = 3.4$, $x = 6$.

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

$$= \frac{(3.4)^6 e^{-3.4}}{6!}$$

$$= 0.072$$

$$= 7.2\%$$

E6. If 3% of electronic units manufactured by company are defective. Find probability that in a sample of 200 units, less than 2 bulbs are defective.

$p = 0.03$ $n = 200$ $\lambda = 0.03 \times 200 \Rightarrow 6$

$x = 0, 1$

$$P(x<2) = P(x=0) + P(x=1)$$
$$= \left(\frac{6^0 \cdot e^{-6}}{0!}\right) + \left(\frac{6^1 \cdot e^{-6}}{1!}\right)$$
$$= e^{-6} + 6e^{-6}$$
$$= 7e^{-6}$$
$$= 0.01727$$

E7. In a book of 520 pages, 390 typo graphical errors occur. Assuming Poisson law for the no. of error per page, find probability that a random sample of 5 pages will contain no error.

Errors per pages $\lambda = \frac{390}{520} \Rightarrow 0.75$.

$$P(x \text{ page error}) = \frac{\lambda^x e^{-\lambda}}{x!}$$
$$= \frac{(0.75)^x \, e^{-0.75}}{x!}$$

For ramplom sample of 5 pages ; no error is

$$[P(X=0)]^5 = \left(e^{-0.75}\right)^5$$
$$= e^{-3.75}.$$

E8. In certain factory turning out blades there is a small chance of 0.002. For any blade to be defective. The blades are supplied in packet of 10. using poisson distribution. find the approximatily no. of packets containing

(a) no defective blades.

$n = 10$, $p = 0.002$, $N = 10000$

$\lambda = np$

$= 10 \times 0.002 = 0.02$.

$$P(0) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$= \frac{(0.02)^0 \, e^{-0.02}}{0!}$$

$$= 0.9802.$$

No defective blades number $= 10000 \times 0.9802$

$= 9802$

(b) One defective blades.

$$P(1) = \frac{(0.02)^1 \, e^{-0.02}}{1!}$$

$$= 0.02 \times 0.9802$$

$$= 0.0196$$

$\Rightarrow$ One defective blades number $= 10000 \times 0.0196$

$= 196.$

Chapter 3 → Continuous Distribution

Continuous distribution

→ They are fundamental concept in probability theory.

→ It deals with random variable taking on values within a continuous range.

→ They encompass an uncountably infinite no. of possible values.

→ It is solved using integral "$\int$."

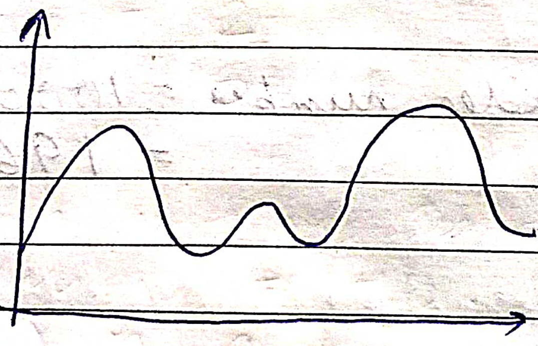
Points to remember.

1. $f(x)$ → probability density function.
   $F(x)$ → cummulative density function.

$$\frac{dF(x)}{dx} \Rightarrow f(x)$$

'$\int$'

2. Graph is always continuous and there is no discrete value.



Ex :- $1 < x < 5$.

* 3. PDF in the interval $[a, b]$ is

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

4. CDF for [a, b] interval is given by
$$F(x) = \frac{x-a}{b-a} \quad \text{for } a \le x \le b$$

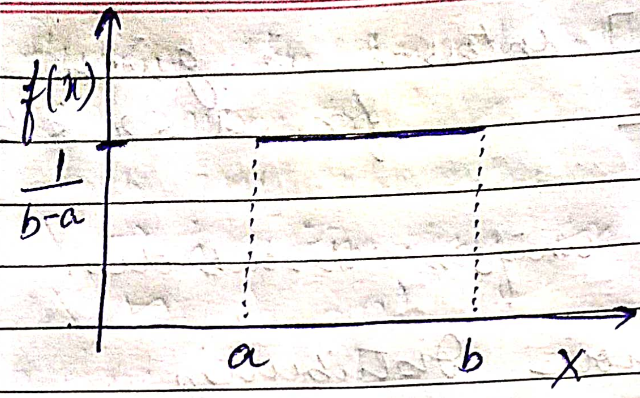
* :- for special case i.e uniform distribution

Types of Continuous Distribution
It has three major types.
1. Uniform Distribution.
2. Exponential Distribution
3. Normal distribution.

Uniform Distribution
→ It is related to events which are equally likely to occur.
→ Defined by two parameters x and y or a and b
→ Where a is minimum value and b is the maximum value.
→ PDF must be constant over entire range
→ PDF over [a, b] is
$$f(x) = \frac{1}{b-a}$$
→ CDF over [a, b] is
$$F(x) = \frac{x-a}{b-a}$$
→ Mean $\mu = \frac{a+b}{2}$

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

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



E1. Consider continuous uniform distribution over the interval [2, 7]. The PDF for this distribution?

$$f(x) = \frac{1}{7-2}$$

$$= \frac{1}{5} \quad \text{for } 2 \le x \le 7.$$

It means that any value within the interval [2, 7] has equal probability of 1/5

E2. Consider above E1 and find probability that X is less then or equal to 4.

$$F(4) = \frac{4-2}{7-2}$$

$$= \frac{2}{5}.$$

E3. If X is uniformly distributed with mean = 1 and variance = 4/3. find $P(X<0)$.

Let $X \sim [a, b]$ then

$$p(x) = \frac{1}{b-a} \quad \text{for } a \le x \le b.$$

$$\text{Mean} = \frac{b+a}{2} = 1$$

$b + a = 2$ — (1)

Variance $= \frac{(b-a)^2}{12} = \frac{4}{3}$

$= (b-a)^2 = 16$

$b - a = \pm 4$ — (2)

Combine (1) & (2)

$$\begin{aligned} b + a &= 2 \\ b - a &= 4 \\ \hline 2b &= 6 \end{aligned}$$

$\boxed{b = 3}$

$3 + a = 2$

$\boxed{a = -1}$

$p(x) = \frac{1}{3-(-1)}$

$= \frac{1}{4}$ for $-1 < x < 3$.

$P(X<0) = \int_{-1}^{0} p(x)\,dx$

$= \int_{-1}^{0} \frac{1}{4} dx$

$= \frac{1}{4} [x]_{-1}^{0}$

$= \frac{1}{4} (0 - (-1))$

$= \frac{1}{4}$.

Eg. Subway trains on a certain line run every half hour between mid night and six in the morning. What is probability that a man entering the station at a random time during this period will have to wait atleast 20 minutes?

$X \rightarrow$ Waiting time for next train

$X \sim (0, 30)$

$$f(x) = \frac{1}{b-a}$$

$$= \frac{1}{30-0} = \frac{1}{30}. \quad \text{for } 0 < x < 30$$

P(Waiting atleast 20 minutes)

$$F(x) = \int_{20}^{30} f(x)\,dx$$

$$P(X \geq 20) = \int_{20}^{30} \frac{1}{30}\,dx$$

$$= \frac{1}{30}\,[x]_{20}^{30}$$

$$= \frac{1}{30}(30-20)$$

$$= \frac{1}{3\cancel{0}} \times 1\cancel{0}$$

$$= \frac{1}{3}.$$

Exponential Distribution

→ The probability distribution of time between events in the poisson point process is called exponential distribution.

→ It is special case of gamma distribution

→ It is a process in which events happens continuously and independently at a constant average rate.

→ PDF is given by

$$f_X(x \mid \lambda) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

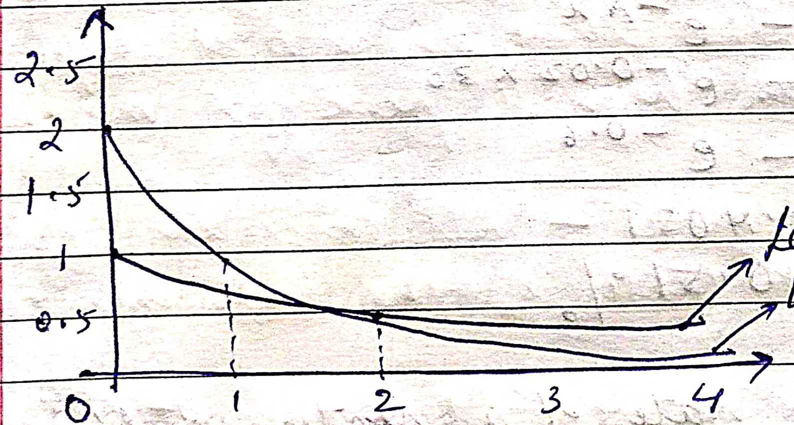
λ is distribution rate.

→ Mean $\mu = \dfrac{1}{\lambda}$

→ Variance $\sigma^2 = \dfrac{1}{\lambda^2}$

→ S.D $= \sqrt{\dfrac{1}{\lambda^2}}$

→ Graph is denoted as



→ CDF is defined by

$$F(x) = 1 - e^{-\lambda x} \qquad x \geq 0$$

Date:

Page:

E1. Suppose we have a light bulb that has a constant failure rate and time until failure follows an exponential distribution with a [illegible] of $\lambda = 0.02$. What is probability that the light bulb will last more than 50 hours without failing?

$$P(X>50) = \int_{50}^{\infty} \lambda e^{-\lambda x} dx$$

$$= \int_{50}^{\infty} 0.02 e^{-0.02 x} dx$$

$$= 0.02 \int_{50}^{\infty} e^{-0.02x} dx$$

$$= 0.1353$$

$$= 13.53\%$$

E2. Continuing to E1, find probability that the light bulb will last less than or equal to 30 hours without failing?

$$F(x) = 1 - e^{-\lambda x}$$

$$F(30) = 1 - e^{-0.02 \times 30}$$

$$= 1 - e^{-0.6}$$

$$= 0.4051$$

$$= 40.51\%$$

E3. If the failure rate of a system follows an exponential distribution with $\lambda = 0.1$, then the mean time to failure and the variance is?

$\lambda = 0.1$

$$\text{Mean} = \frac{1}{\lambda} = \frac{1}{0.1} \Rightarrow 10.$$

$$\text{Variance} = \frac{1}{\lambda^2} = \frac{1}{(0.1)^2} \Rightarrow 100.$$

Eg. Assume that, you usually get 2 phone calls per hour. Calculate probability that a phone call will come within the next hour.

2 call per hour = 1 call per half-hour

$$\lambda = \frac{1}{2} = 0.5.$$

$$p(x) = \int_{x=0}^{1} 0.5\, e^{-0.5x} \qquad 0 \leq x \leq 1$$

$$= 0.5 \int_{x=0}^{1} e^{-0.5x}$$

$$= 0.3934$$

Applications

1. helps to find distance b/w mutations on a DNA strand
2. Calculating time until ~~rada~~ radioactive particle decays
3. helps to computate the monthly and annual highest values of regular rainfall.

⟹ The key property is memoryless as the past has no impact on its future behaviour and each instant is like the starting of the new random period.

Normal Distribution

→ Also called Gaussian distribution
→ It is widely used in statistics.
→ It is characteristic by its bell shaped cu
→ It is applicable to a diverse range of phenomena in physics, finance.
→ PDF is given by

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

x :- variable
μ :- mean
σ :- standard deviation.

→ Empirical Rule is of 68 − 95 − 99.7 ru

Properties.

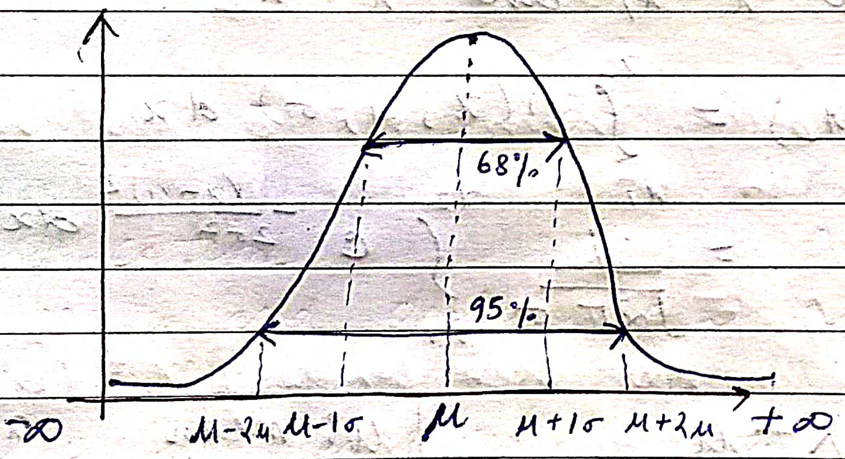
1. Symmetry − The normal curve is symmetric i.e through central peak; both sides look similar and thus probability is also same
2. Bell Shaped − The curve has a bell shaped appearance, with highest peak at mean.
3. Mean, Median, Mode − All three are equal and located at the center of the distribution
4. Area under the curve − It represents the probability of a random variable falling within a certain range.

5. Standard Deviation – The spread / dispersion of the curve is determined by S.D. The larger the SD, the wider than curve

6. Asymptotic to X-axis – Curve extend infinitely in both direction but never touches the x-axis

7. Empirical Rule – 68 – 95 – 99.7. It means that 68% of data falls within one SD, 95% of data falls within two SD and 99.7% falls within three SD.

8. Z-Scores – They are used to standardize values on the normal curve. It represents the no. of SD a data point is from mean.



Q1. Calculate PDF of normal distribution using data $x = 3$, $\mu = 4$ and $\sigma = 2$.

$$f(3,4,2) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(3-2)^2}{2\times2^2}}$$

$$= \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{8}} \Rightarrow 1.106.$$

Page:

E2. If the value of random variable is 2, mean is 5 and SD is 4. Calculate PDF.

$$f(2,2,4) = \frac{1}{4\sqrt{2\pi}} e^{-\frac{(2-2)^2}{2\times4^2}}$$

$$= \frac{1}{4\sqrt{2\pi}} e^{0}$$

$$= 0.0997$$

Areas under the normal curve

→ Curve of any continuous probability distribution or density function is constructed so that area under the curve bounded by two ordinates $x = x_1$ and $x = x_2$ equals the probability that random variable X assumes a value b/w $x = x_1$ and $x = x_2$.

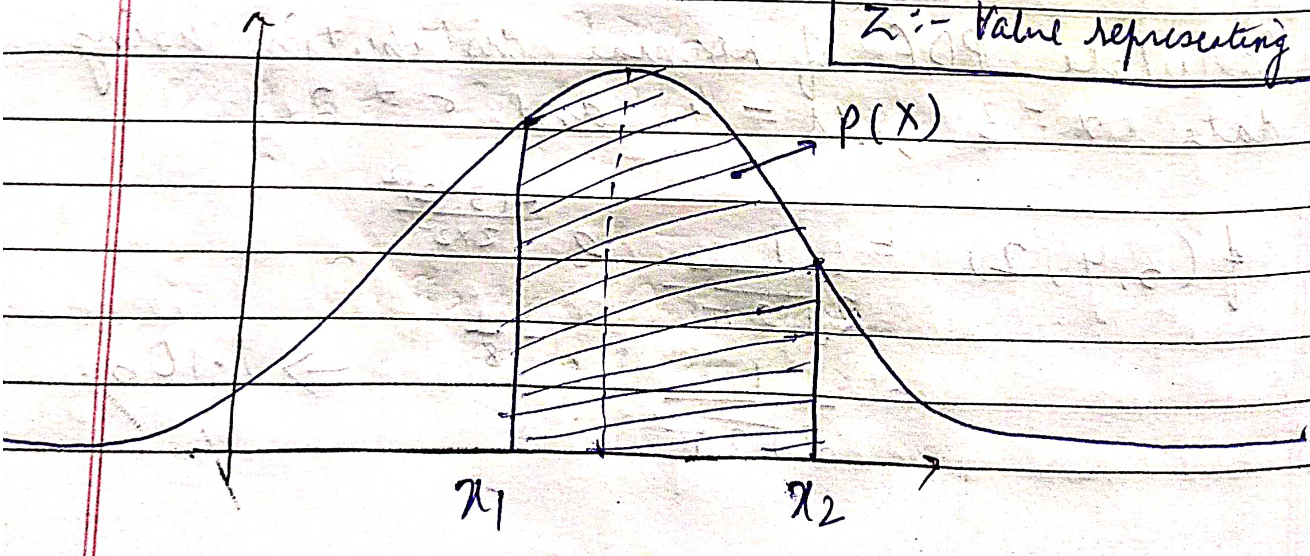
$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} n(x, \mu, \sigma)\, dx$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{x_1}^{x_2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

→ $Z = \frac{X - \mu}{\sigma}$

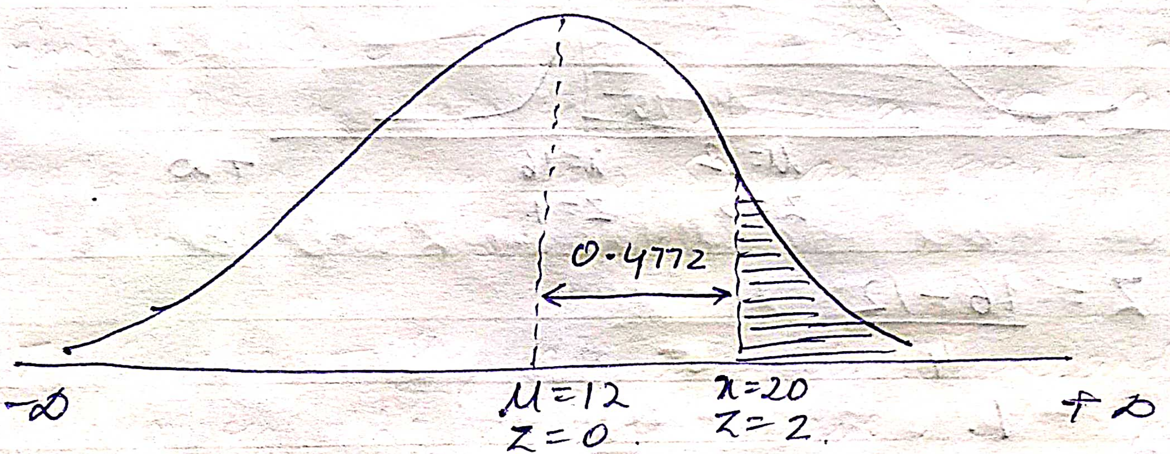
μ :- mean
σ :- SD
Z :- Value representing area

x :- Req. value



→ Total area must be 1.

E1. Calulate probability when x is greater than equal to 20 when mean = 12, SD = 4. Given that $Z(0 \text{ to } 1) = 0.31413$ and $Z(0 \text{ to } 2) = 0.4772$

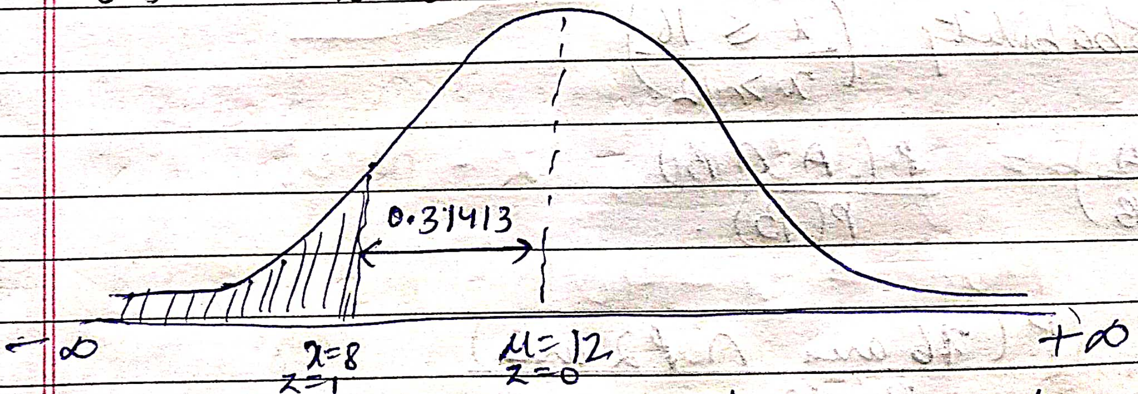


$$Z = \frac{20-12}{4} \Rightarrow \frac{8}{4}$$
$$= 2.$$

From $Z = 0$ to $Z = +\infty$ value is 0.5

$\therefore$ $0.5 - 0.4772 \Rightarrow 0.228$.

E2. Calulate probability when x is less than equal to 8 with other values in E1.

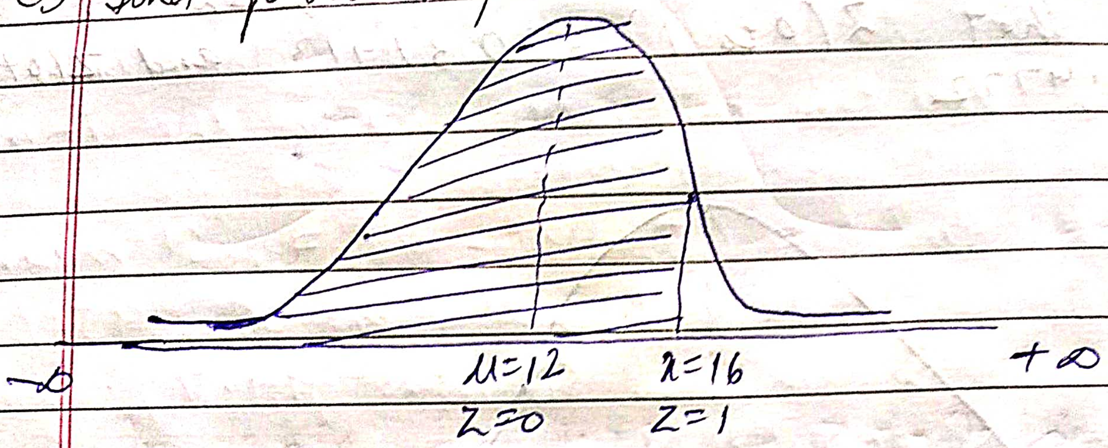


$$Z = \frac{8-12}{4} \Rightarrow \frac{-4}{4} \Rightarrow -1.$$

-ve represent left side

Probability $= 0.5 - 0.3413$
$= 0.1587$

E3 Find probability when $x \le 16$ as same in E1



$$Z = \frac{16-12}{4}$$
$$= 1.$$

Probability $= 0.5 + 0.3413$
$= 0.8413$

E4. Find probability when $x = 16$ rest values are same as in E1.

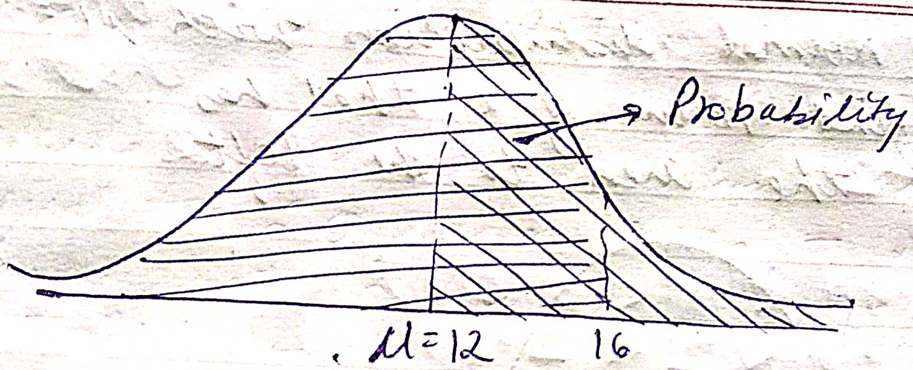
Probability $= 0$.
As, there is no area under $\boxed{x = 16}$

E5 Probability $\left(\frac{x \le 16}{x \ge 12}\right)$

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

$$\therefore \frac{P(\cancel{16}\text{ area} \cap P2\text{ area})}{P(12\text{ area})}$$

Date: Page:

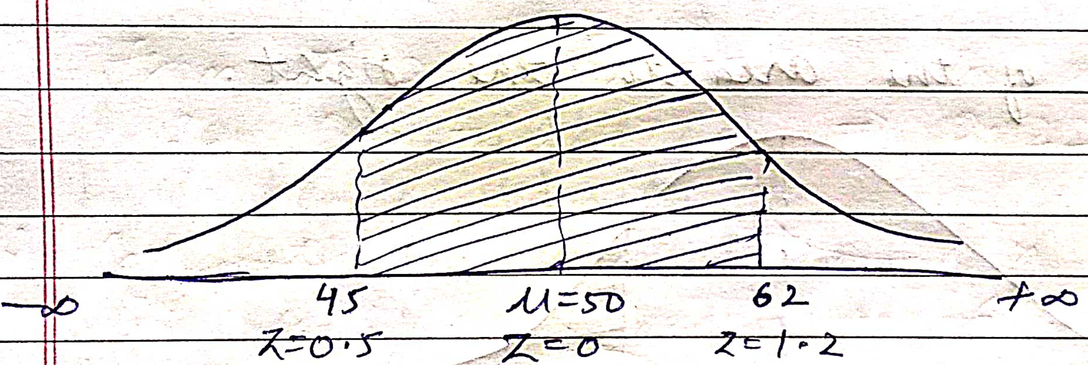


$$Z = \frac{16 - 12}{4} = 1$$

$$\text{Probability} = \frac{0.3413}{0.5}$$

$$= 0.6826.$$

E6 Given a random variable X having a normal distribution with $\mu = 50$ and $\sigma = 10$. find probability that X value is b/w 45 & 62



$$Z = \frac{45 - 50}{10} \Rightarrow -0.5$$

$$Z = \frac{62 - 50}{10} \Rightarrow 1.2$$

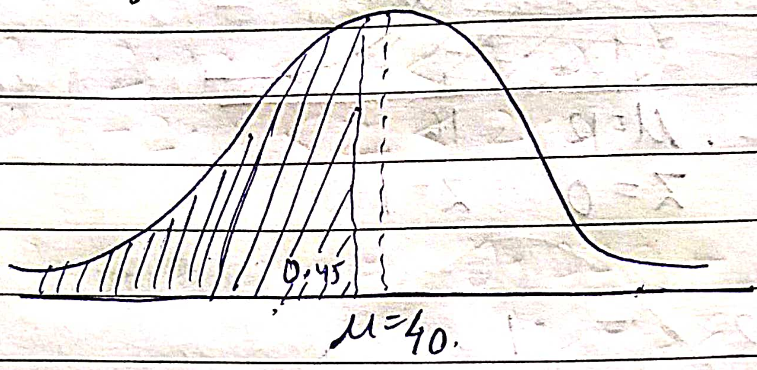
$$\text{Probability} = P(-0.5 < Z < 1.2)$$

$$= 0.8849 - 0.3085$$

$$= 0.5764$$

E7 Given a normal distribution with $\mu = 40$ and $\sigma = 6$, find value of x that has

(a) 45% of the area to the left.



In table for 0.45, $z = -0.13$.

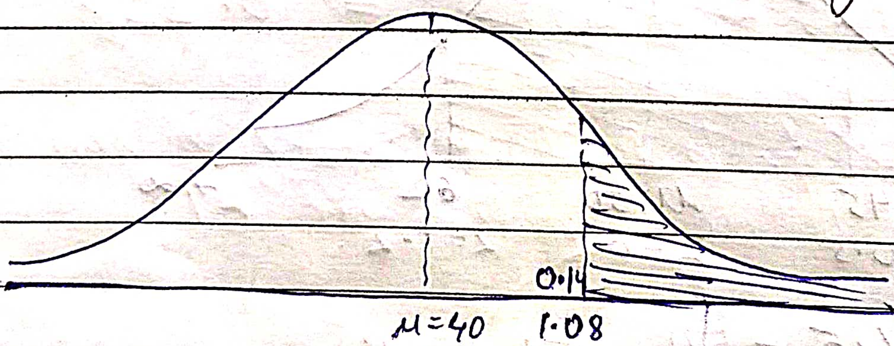
$\therefore z = \frac{x - \mu}{\sigma}$

$-0.13 = \frac{x - 40}{6}$

$-0.78 + 40 = x$

$x = 39.22$.

(b) 14% of the area to the right.



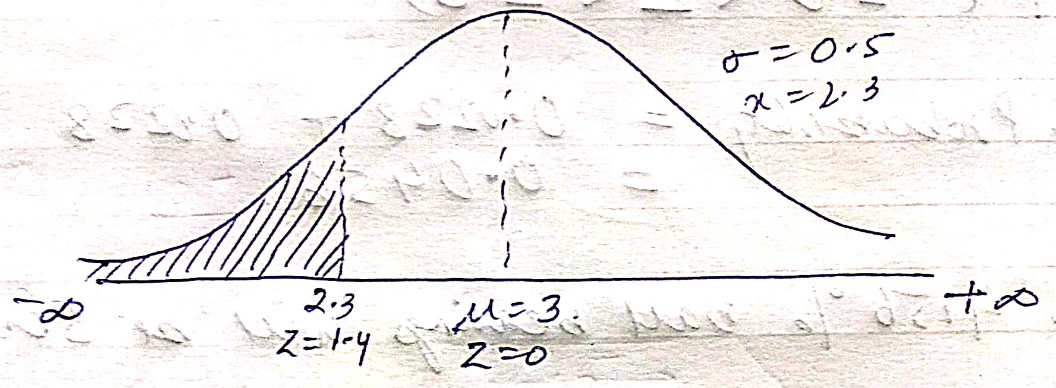
In table for 0.14; $z = 1.08$.

$1.08 = \frac{x - 40}{6}$

$x = 40 + 6.48$

$= 46.48$

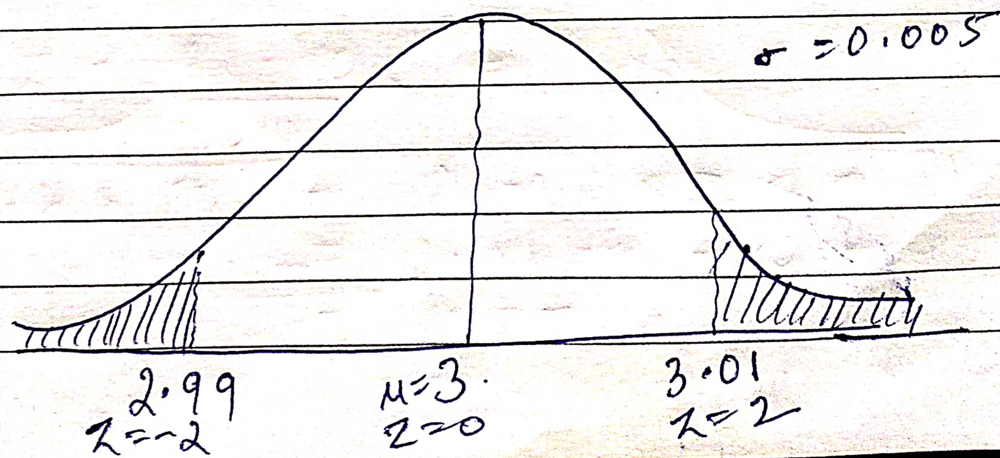
E8. A certain type of storage battery last on average 3 years with SD of 0.5 year. Assuming that battery life is normally distributed, find the probability that battery will last less than 2.3 years.



$$Z = \frac{2.3 - 3}{0.5} \Rightarrow \frac{-0.7}{0.5} \Rightarrow 1.4.$$

Probability = Z value of 1.4 $(Z < -1.4)$
= 0.0808

E9. In an industrial process, the buyer sets ball bearing diameter to be 3.0 ± 0.01 cm. The implication is that no part falling outside these will be accepted. Mean = 3.0 and SD = 0.005. Find how many ball bearings will be scrapped?



$$Z_1 = \frac{2.99 - 3.0}{0.005} = -2 \qquad Z_2 = \frac{3.01 - 3.0}{0.005} = 2$$

$$P(2.99 < X < 3.01)$$
$$P(-2.0 < Z < 2.0)$$

$\therefore$ Probability $= 0.0228 + 0.0228$
$= 0.0456$.

$\therefore$ 4.56 % ball bearings will be Scrapped.