Index

3|1|25

Module 2:

Co-relation and Regression

Coefficient of corelation $= r$

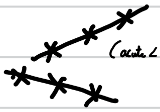
$-1 \le r \le 1$

(Not in exam): Scattered diagram

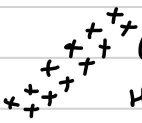
if all points lie on the line

① → bending towards right plane : perfect +ve correlation $r = +1$

② → bending towards left plane : perfect −ve correlation $r = -1$

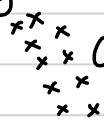


(acute ∠ w.r.t x-axis)

③  Close to each other but not on line towards right

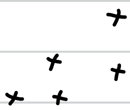
High degree +ve correlation

r tending to $+1$

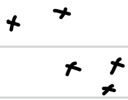
④  Close to each other but not on line towards left

High degree −ve correlation

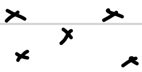
r tending to -1

⑤  Low degree +ve correlation

r tending to 0 (on the positive side)

⑥  Low degree −ve correlation

r tending to 0 (negative side)

⑦  No correlation

$r = 0$

r = Karl - Pearson's coefficient of correlation

$$= \boxed{\frac{\mathrm{cov}(x,y)}{\sigma_x \times \sigma_y}} = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{N \times \sigma_x \times \sigma_y}$$

N → No of observations

cov → covariance, measures the variation of two variables together from their origins

where $\sigma_x = \sqrt{\dfrac{\Sigma (x_i - \bar{x})^2}{N}}$ = standard deviation of x $\qquad \bar{x} = \dfrac{\Sigma x}{N}$

$\sigma_x^2 = \mathrm{var}(x)$

$\sigma_y = \sqrt{\dfrac{\Sigma (y_i - \bar{y})^2}{N}}$ = Standard deviation of y $\qquad \bar{y} = \dfrac{\Sigma y}{N}$

$\sigma_y^2 = \mathrm{var}(y)$

$\sigma_x = \sqrt{\dfrac{\Sigma x^2}{n} - \bar{x}^2}$

$\sigma_y = \sqrt{\dfrac{\Sigma y^2}{n} - \bar{y}^2}$

Alternative formula for r

$$\boxed{r = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma (x_i - \bar{x})^2}\sqrt{\Sigma (y_i - \bar{y})^2}}}$$

$$\boxed{r = \frac{\Sigma x_i y_i - N\bar{x}\bar{y}}{\sqrt{\Sigma x_i^2 - N(\bar{x})^2}\sqrt{\Sigma y_i^2 - N(\bar{y})^2}}}$$

Properties

① $-1 \le r \le 1$

② if x and y are independent then there is no correlation between x and y. $r = 0$

③ r is independent of change of scale and change of origin.

if $u = \dfrac{x - \bar{x}}{h}$, $v = \dfrac{y - \bar{y}}{k}$

then $r_{uv} = r_{xy}$

eg Calculate r for the following data   Using change of origin

x	y	$u_i = x_i - 30$	$v_i = y_i - 25$	$u_i \cdot v_i$	u_i^2	v_i^2
23	18	-7	-7	49	49	49
27	22	-3	-3	9	9	9
28	23	-2	-2	4	4	4
29	24	-1	-1	1	1	1
30	25	0	0	0	0	0
31	26	1	1	1	1	1
33	28	3	3	9	9	9
35	29	5	4	20	25	16
36	30	6	5	30	36	25
39	32	9	7	63	81	49
$N = 10$		$\Sigma u_i = 11$	$\Sigma v_i = 7$	$\Sigma u_i v_i = 186$	$\Sigma u_i^2 = 215$	$\Sigma v_i^2 = 163$

$\bar{u} = \dfrac{\Sigma u_i}{N} = 1.1$ $\bar{v} = \dfrac{\Sigma v_i}{N} = 0.7$

$N \bar{u} \cdot \bar{v} = 10 \times 1.1 \times 0.7 = 7.7$

$$r_{uv} = \frac{\Sigma u_i v_i - N \bar{u} \bar{v}}{\sqrt{\Sigma u_i^2 - N\bar{u}^2}\sqrt{\Sigma v_i^2 - N\bar{v}^2}} = \frac{186 - 7.7}{\sqrt{215 - 10 \times 1.21}\sqrt{163 - 10 \times 0.49}} = \frac{178.3}{\sqrt{202.9}\sqrt{158.1}}$$

$\underline{r_{uv} = 0.996}$ High degree +ve correlation

6|1|25

R = Spearman's Rank correlation coefficient

$$\boxed{R = 1 - \left[\frac{6 \Sigma d^2}{n^3 - n}\right]}$$

where d = difference between ranks
n = no of observations

$-1 \le R \le 1$

Q Calculate R

Std No	Ranks in Eng (R_1)	Ranks in Maths (R_2)	$d^2 = (R_1 - R_2)^2$ or $(R_2 - R_1)^2$
1	1	3	4
2	3	1	4
3	7	4	9
4	5	5	0
5	4	6	4
6	6	9	9
7	2	7	25
8	10	8	4
9	9	10	1
10	8	2	36

$n = 10$ $\qquad \Sigma d^2 = 96$

$$R = 1 - \left[\frac{6 \times 96}{1000 - 10}\right] = 1 - 0.5818 = 0.4182$$

Low degree +ve correlation

When Ranks are repeated then $R = 1 - \left[\dfrac{6\left[\Sigma d^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) - - -\right]}{n^3 - n}\right]$

m_i = no of times any entry is repeated

* Ranks remain same whether ascending or descending

Q Calculate R

X	R_1	Y	R_2	$d^2 = (R_1 - R_2)^2$
32	3	40	5	4
55	9	30	3.5	30.25
49	7.5	70	9	2.25
60	10	20	1	81
43	5.5	30	3.5	4
37	4	50	7	9
43	5.5	72	10	20.25
49	7.5	60	8	0.25
10	1	45	6	25
20	2	25	2	0

$n = 10$ $\qquad \Sigma d^2 = 176$

$m_1 = 2 \qquad m_1^3 - m_1 = 6 \qquad n^3 - n = 990$

$m_2 = 2 \qquad m_2^3 - m_2 = 6$

$m_3 = 2 \qquad m_3^3 - m_3 = 6$

$$R = 1 - \frac{6\left[\Sigma d^2 + \frac{3}{12}\left[m^3 - m\right]\right]}{n^3 - n} \quad (m_1 = m_2 = m_3) = 1 - \frac{6\left[176 + \frac{3}{12}(6)\right]}{990} = 1 - 1.0757$$

$= -0.075$

Low degree −ve correlation

$* \; \Sigma d = 0$

Q If $r_{xy} = 0.4$, $cov(x, y) = 1.6$ $\sigma_y^2 = 25$ find σ_x.

$$r_{xy} = \frac{cov(x, y)}{\sigma_x \cdot \sigma_y} \qquad \sigma_y = \sqrt{\sigma_y^2} = 5$$

$$0.4 = \frac{1.6}{\sigma_x \cdot 5} \Rightarrow \sigma_x = 0.8$$

Q if $R = 0.143$ $\Sigma d^2 = 48$ find n

$$R = 1 - \left[\frac{6 \, \Sigma d^2}{n^3 - n}\right]$$

$$0.143 = 1 - \left[\frac{6 \times 48}{n^3 - n}\right]$$

$$\frac{288}{n^3 - n} = 0.857 \Longrightarrow n^3 - n - 336.05 = 0$$

$x_1 = 7.00034$

$x_2 = -3.500171$

$x_3 = -3.500171$

$\therefore \; n = 7$ (Take positive value)

Q $N = 10$, $\Sigma x = 140$, $\Sigma y = 150$, $\Sigma (x-10)^2 = 180$, $\Sigma (y-15)^2 = 215$, $\Sigma (x-10)(y-15) = 60$ find r

Let $u = x - 10$

$v = y - 15$

$\Sigma u^2 = 180$ $\quad \Sigma v^2 = 215$ $\quad \Sigma u.v = 60$

$$\bar{u} = \frac{\Sigma u}{N} = \frac{\Sigma (x-10)}{N} = \frac{\Sigma x - \Sigma 10}{N} = \frac{140 - 100}{10} = 4$$

$$\bar{v} = \frac{\Sigma v}{N} = \frac{\Sigma (y-15)}{N} = \frac{\Sigma y - \Sigma 15}{N} = \frac{150 - 150}{10} = 0$$

$$r_{uv} = \frac{\Sigma u.v - N.\bar{u}.\bar{v}}{\sqrt{\Sigma u^2 - N(\bar{u})^2} \cdot \sqrt{\Sigma v^2 - N(\bar{v})^2}} = \frac{60 - 0}{\sqrt{20} \cdot \sqrt{215}} = \frac{60}{65.5743} = 0.9149$$

High degree +ve correlation.

9|1|25

Q A sample of 25 pairs of (x, y) gives the following result $\Sigma x = 127$ $\Sigma y = 100$ $\Sigma x^2 = 760$ $\Sigma y^2 = 449$ $\Sigma xy = 500$ later on it was found that 2 pairs of values were taken as $(8, 14)$ and $(8, 6)$ instead of $(8, 12)$ and $(6, 8)$ find correct correlation coefficient

Σx (correct) $= 127 - 8 - 8 + 8 + 6 = 125$

Σy (correct) $= 100 - 14 - 6 + 12 + 8 = 100$

Σx^2 (correct) $= 760 - 8^2 - 8^2 + 8^2 + 6^2 = 732$

Σy^2 (correct) $= 449 - 14^2 - 6^2 + 12^2 + 8^2 = 425$

Σxy (correct) $= 500 - (8\times14) - (8\times6) + (8\times12) + (6\times8) = 484$

correct $\bar{x} = \frac{\Sigma x}{25} = 5$ correct $\bar{y} = \frac{\Sigma y}{25} = 4$

$$r = \frac{\Sigma xy - N\bar{x}\bar{y}}{\sqrt{\Sigma x^2 - N\bar{x}^2}\sqrt{\Sigma y^2 - N\bar{y}^2}} = \frac{484 - 25\times5\times4}{\sqrt{732-625}\sqrt{425-400}}$$

$$= \frac{-16}{\sqrt{107}\times 5} = \frac{-16}{51.720} = -0.30935$$

Regression Lines

Regression line of y on x:

$$\boxed{\begin{array}{l} y - \bar{y} = b_{yx}(x - \bar{x}) \\ \text{where } b_{yx} \text{ is } r\frac{\sigma_y}{\sigma_x} \end{array}}$$

Regression line of x on y:

$$\boxed{\begin{array}{l} x - \bar{x} = b_{xy}(y - \bar{y}) \\ \text{where } b_{xy} \text{ is } r\frac{\sigma_x}{\sigma_y} \end{array}}$$

b_{xy} and b_{yx} are called regression coefficients and in fact they are the slopes of the respective lines

Properties

① If $r = 0$, this implies that $y = \bar{y}$ and $x = \bar{x}$.
The regression lines are perpendicular to each other.

② If $r = \pm 1$, the regression lines are parallel

③ $\boxed{r = \pm\sqrt{b_{xy} * b_{yx}}}$ (if both +ve then +, if both -ve then - sign)

④ If $b_{xy} > 1$ then $b_{yx} < 1$ (reciprocals)

⑤ $\frac{b_{xy} + b_{yx}}{2} \geq r$ (prove)

⑥ If θ is the angle between 2 regression lines then $\boxed{\tan\theta = \left(\frac{1-r^2}{r}\right)\cdot\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}}$

if $r = 0$ $\tan\theta = \infty$ $\theta = \pi/2$ perpendicular

if $r = \pm1$ $\tan\theta = 0$ $\theta = 0$ parallel

⑦ $\boxed{b_{xy} = \dfrac{\Sigma xy - N\bar{x}\bar{y}}{\Sigma y^2 - N\bar{y}^2}}$ $\boxed{b_{yx} = \dfrac{\Sigma xy - N\bar{x}\bar{y}}{\Sigma x^2 - N\bar{x}^2}}$

⑧ Regression coefficients are independent of change of origin but not change of scale.

if you know x and want to find y find y on x (b_{yx})
if you know y and want to find x find x on y (b_{xy})

eg A panel of 2 judges A and B graded the following performances judge A awarded 38 marks for 8th performance while judge B was absent, how many marks judge B would have awarded if he was present.

Perf. No.	X(A)	Y(B)	$u = x - \bar{x}$	$v = y - \bar{y}$	uv	u^2
1	36	35	3	2	6	9
2	32	33	-1	0	0	1
3	34	31	1	-2	-2	1
4	31	30	-2	-3	6	4
5	32	34	-1	1	-1	1
6	32	32	-1	-1	1	1
7	34	36	1	3	3	1
8	X (38)	Y	–	–		
	$\Sigma x = 231$	$\Sigma y = 231$	$\Sigma u = 0$	$\Sigma v = 0$	$\Sigma uv = 13$	$\Sigma u^2 = 18$

$y \to x$ b_{yx} $v \to u$ b_{vu}

$\bar{x} = \dfrac{\Sigma x}{7} = 33$ $\bar{y} = \dfrac{\Sigma y}{7} = 33$

$b_{vu} = \dfrac{\Sigma uv - N\bar{u}\bar{v}}{\Sigma u^2 - N\bar{u}^2} = \dfrac{13 - 0}{18 - 0}$

$\bar{u} = \dfrac{\Sigma u}{7} = 0$ $\bar{v} = \dfrac{\Sigma v}{7} = 0$

$= 0.72$

∴ Regression line of v on u is $v - \bar{v} = b_{vu}(u - \bar{u})$

$\Rightarrow v = 0.72u$

$y - \bar{y} = 0.72(x - \bar{x})$

$x = 38$

$y - 33 = 0.72(38 - 33)$

$y - 33 = 0.72(5)$

$y = 33 + 3.6$

$y = 36.6 \simeq \underline{\underline{37}}$

10|1|25

Q Following is the information about marks of 60 students, estimate marks of students in maths if they scored 60 in english and marks of a student in english who scored 70 marks in maths

	Maths	English
Mean	($\bar{x}$) 80	($\bar{y}$) 50
SD	15 (σ_x)	10 (σ_y)

$r = 0.4$

① $x = ?$ $y = 60$ ② $x = 70$ $y = ?$

x on y b_{xy} y on x b_{yx}

$b_{yx} = r\frac{\sigma_y}{\sigma_x} = 0.4 \times \frac{10}{15} = 0.266$

$b_{xy} = r\frac{\sigma_x}{\sigma_y} = 0.4 \times \frac{15}{10} = 0.6$

Regression line of x on y [y=60]

$x - \bar{x} = b_{xy}(y - \bar{y})$

$x - 80 = 0.6(60 - 50)$

$\underline{x = 86}$

Regression line of y on x [x=70]

$y - \bar{y} = b_{yx}(x - \bar{x})$

$y - 50 = 0.266(70 - 80)$

$y = 0.27(-10) + 50$

$\underline{y = 47.3}$

Q Given that $6y = 5x + 90$ $15x = 8y + 130$ $\sigma_x^2 = 16$ find ① $\bar{x}$ & $\bar{y}$ ② r ③ σ_y^2

$6y - 5x - 90 = 0$

$15x - 8y - 130 = 0$

} simultaneously solve these on the calculator

① $\underline{\bar{x} = 30 \quad \bar{y} = 40}$ as they will be the points where these lines intersect

$y = \frac{5x}{6} + \frac{90}{6}$ ↳ b_{yx}

$x = \frac{8y}{15} + \frac{130}{15}$ ↳ b_{xy}

As coefficient of x, y [$y - \bar{y} = b_{yx}(x - \bar{x})$]

$r = +\sqrt{b_{yx} \cdot b_{xy}} = \sqrt{\frac{5}{6} \times \frac{8}{15}}$

$= \sqrt{\frac{1}{3} \times \frac{4}{3}} = 0.667$

other way round would've been $r = 1.5$ so not possible

② $\underline{r = 0.66}$

$b_{yx} = r\frac{\sigma_y}{\sigma_x}$

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

$\sigma_y = 5.05$

③ $\underline{\sigma_y = 5}$

Q if $\sigma_x = \sigma_y = \sigma$ $\theta = \tan^{-1}(3)$ find r.

$$\tan\theta = \frac{(1-r^2)}{r} \cdot \frac{\sigma_x \sigma_y}{\sigma_x^2+\sigma_y^2} \qquad \tan\theta = 3$$

$$3 = \frac{1-r^2}{r} \cdot \frac{\sigma^2}{2\sigma^2} \implies 6r = 1 - r^2$$

$$\implies r^2 + 6r - 1 = 0$$

$$r = 0.16$$

Q $\bar{x} = 5$ $\bar{y} = 10$ and line of regression of y on $x \Rightarrow (y-\bar{y}) = b_{yx}(x-\bar{x}) \parallel 20y = 4x + 40$ find y for $x = 30$

parallel lines have equal slopes

From the given data, both lines have identical slopes.

$$\therefore b_{yx} = \frac{4}{20}$$

$$y - \bar{y} = \frac{4}{20}(x - \bar{x})$$

$$y - 10 = \frac{4}{20}(30 - 5)$$

$$y = 21.25$$

Q if tangent of angle made by the line of regression of y on x is 0.6 and $\sigma_y = 2\sigma_x$ find r

tan θ where θ is angle between y on x and x-axis which makes it the slope aka b_{yx}

tan θ formula is used for when the angle is between two regression lines.

$$b_{yx} = 0.6$$

$$b_{yx} = r\frac{\sigma_y}{\sigma_x}$$

$$0.6 = r\,\frac{2\sigma_x}{\sigma_x}$$

$$\underline{r = 0.3}$$

10|1|25

Module 1:

Probability Distribution

Random variable: Is a function which assigns a real number to an outcome of an experiment.

eg Three coins are thrown, if we assign a random variable x be the no of tails

Let x = no of tails

$S = \{(HHH), (HHT), (HTH), (THH), (HTT), (THT), (TTH), (TTT)\}$

$X: \quad 0, \quad 1, \quad 2, \quad 3$

$P(X=x) = 1/8, \quad 3/8, \quad 3/8, \quad 1/8$

Sum of probability is always 1.

$x = 0, \quad A = \{(HHH)\} \quad n(A) = 1$

Distribution of x is called probability distribution

13|1|25

1.1 Conditional Probability

If S is a sample space and A, B be an event of sample s and let B be an event which has already occurred then probability of an event A when B has already occurred is called conditional probability it is given as

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

multiplication theorem: A & B (2 events)

$$\boxed{\begin{aligned} P(A \cap B) &= P(A)\,P(B|A) \\ &= P(B)\,P(A|B) \end{aligned}}$$

Note: ① $P(A|B) \leq P(A)$

② An event A is said to be independent if $P(A|B) = P(A)$

B if $P(B|A) = P(B)$

③ If A and B are independent events then
$P(A \cap B) = P(A)\,P(B)$

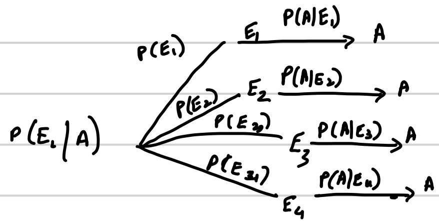
No problems on conditional probability

Baye's Theorem

Statement: If $E_1, E_2, E_3 \cdots E_n$ are mutually disjoint events i.e. $E_i \cap E_j = \phi, \forall i \neq j$, $P(E_i) \neq 0, \forall i$ then for every arbitrary event A, which is a subset of union of E_i ($A \subseteq \bigcup_{i=1}^{n} E_i$) such that $P(A) > 0$

we have $\boxed{P(E_i/A) = \dfrac{P(E_i)\, P(A|E_i)}{\sum_{i=1}^{\infty} P(E_i)\, P(A|E_i)}}$ $= \dfrac{P(E_i)\, P(A|E_i)}{P(A)}$

$\sum_{i=1} P(E_i)\, P(A/E_i) = P(A)$



Q Why $P(A) = \sum_{i=1}^{n} P(E_i)\, P(A/E_i)$

$\because A \subseteq \bigcup_{i=1}^{n} E_i$

we have $P[A \cap (\cup E_i)]$

$= P(A)$

$P(A) = P[A \cap (E_1 \cup E_2 \cup E_3 \cup \cdots \cup E_n)]$

$= P[(A \cap E_1) \cup (A \cap E_2) \cup \cdots \cup (A \cap E_n)]$

$= P(A \cap E_1) + P(A \cap E_2) - + P(A \cap E_n)$

$= \Sigma P(E_i \cap A)$

$= \sum_{i=1}^{n} P(E_i)\, P(A|E_i)$

Q Suppose that a product is produced in 3 factories x y z it is known that factory x produces 3 times as many as factory y and factories y and z produce same no of products assume that 3% of products produced by x and z each are defective while 5% produce of y is defective. All the items produced in 3 factories are stocked stock and an item is selected at random.

① Probability item is defective

② if selected at random and defective, probability that it is produced by x, y, z respectively

Let the no of items produced by y and z be n

then no of items produced by x be 3n

Let E_1 E_2 E_3 denote events

$E_1 \to x$, $E_2 \to y$, $E_3 \to z$

A = defective

Then $P(E_1) = \dfrac{3n}{5n} = 0.6$ $\qquad P(E_2) = \dfrac{n}{5n} = 0.2 = P(E_3)$

$P(A|E_1) = 0.03 = P(A/E_3)$
$P(A/E_2) = 0.05$

$P(A) = \Sigma P(E_i)\ P(A/E_i)$
$= 3/5 \times 0.03 + 1/5 \times 0.05 + 1/5 + 0.03$
$= 0.0034$

$$P(E_1|A) = \frac{P(E_1)\ P(A/E_1)}{\Sigma P(E_i)\ P(A/E_i)} = \frac{3/5 \times 0.03}{0.0034} = 0.529$$

$$P(E_2/A) = \frac{P(E_2)\ P(A/E_2)}{\Sigma P(E_i)\ P(A/E_i)} = \frac{1/5 \times 0.05}{0.0034} = 0.294$$

$$P(E_3/A) = \frac{P(E_3)\ P(A/E_3)}{\Sigma P(E_i)\ P(A/E_i)} = \frac{1/5 \times 0.03}{0.0034} = 0.176$$

Three candidates for principal x y z whose chances of getting appointment R are 4:2:3 probability that Mr. X is selected who would introduce co-education is 0.3 Probability of x and z doing the same are 0.5 and 0.8 resp.
1. What is the probability that co-education takes place
2. Already exists then Mr. Z is principal

Coeducation in college = A

E_1 — x, E_2 — y, E_3 — z selected for principal

$P(E_1) = 4/9$ $\quad P(A/E_1) = 0.3$
$P(E_2) = 2/9$ $\quad P(A/E_2) = 0.5$
$P(E_3) = 3/9$ $\quad P(A/E_3) = 0.8$

To find $P(E_1|A)$
$P(A) = \Sigma P(E_i)\ P(A/E_i)$
$= 4/9 \times 0.3 + 2/9 \times 0.5 + 3/9 \times 0.8$
$= 0.13 + 0.11 + 0.27$
$P(A) = 0.51$

$$P(E_3|A) = \frac{P(E_3)P(A/E_3)}{P(A)} = \frac{0.27}{0.51} = 0.529$$

$P(E_2|A) = 0.217$

$P(E_1|A) = 0.522$

14|1|25

The probability of X Y Z becoming managers are 4/9, 2/3, 1/3 (3/9) probability that a bonus scheme will be introduced if X Y Z become managers are 3/10, 1/2, 4/5
① What is the probability that the bonus scheme will be introduced. ② If already introduced then what is the probability of X is the manager.

Let E = Bonus scheme introduced

$E_1 = X$ is manager $\quad E_2 = Y$ is manager $\quad E_3 = Z$ is manager

Event E_i	Prior Prob $P(E_i)$	Cond Prob $P(E/E_i)$	Joint Prob $P(E \cap E_i)$	Posterior Prob $P(E_i/E)$
E_1	$4/9$	$3/10$	$12/90$	$6/23$
E_2	$2/9$	$1/2$	$10/90$	$5/23$
E_3	$1/3$	$4/5$	$24/90$	$12/23$

$P(A) = \Sigma P(E_i) P(E|E_i)$

$\Sigma P(E \cap E_i) = P(A)$

$= 46/90$

Final Ans: $\frac{6}{23}$

A factory produces certain types of outputs by machines m_1 m_2 m_3, daily productions by the 3 machines are 3000, 2500, 4500 units respectively. It shows that 1% of the output produced by m 1 is defective while 1.2% and 2% are defective (by m 2 and m 3) an item is drawn at random and found to be defective. ① probability of it being produced m_1 m_2 m_3.

A = defective units

$P(E_1) = 3/10 \qquad P(A|E_1) = 1/100 \qquad P(A) = \frac{3}{1000} + \frac{3}{1000} + \frac{9}{1000} = \frac{15}{1000}$

$P(E_2) = 25/100 \qquad P(A|E_2) = 12/1000$

$P(E_3) = 45/100 \qquad P(A|E_3) = 2/100$

$P(E_1|A) = 3/15 = 0.2 \qquad [(P(E_1) \cdot P(A|E_1)) / P(A)]$

$P(E_2|A) = 3/15 = 0.2 \qquad [(P(E_2) \cdot P(A|E_2)) / P(A)]$

$P(E_3|A) = 3/5 = 0.6 \qquad [(P(E_3) \cdot P(A|E_3)) / P(A)]$

There are 2 bags A and B, A contains n white and 2 black balls and B contains 2 white and n black balls, 2 balls are removed from one of the bags without replacement. if both the balls drawn are white and the probability that bag A was chosen to draw the balls is 6/7. find the value of n.

Let E = both balls drawn are white

E_1 = Bag A is selected

E_2 = Bag B is selected

$P(E_1|E) = 6/7$

$P(E_1) = P(E_2) = 1/2$

$$P(E|E_1) = \frac{{}^nC_2}{{}^{(n+2)}C_2} = \frac{\frac{n(n-1)}{2}}{\frac{(n+2)(n+1)}{2}} = \frac{n(n-1)}{(n+2)(n+1)}$$

$$P(E|E_2) = \frac{{}^2C_2}{{}^{(n+2)}C_2} = \frac{1}{\frac{(n+2)(n+1)}{2}} = \frac{2}{(n+2)(n+1)}$$

$$P(A) = \sum_{i=1}^{2} P(E_i)\,P(E/E_i)$$

$$= \frac{n(n-1)}{2(n+2)(n+1)} + \frac{2}{2(n+2)(n+1)} = \frac{2+n(n-1)}{(n+2)(n+1)}$$

$$P(E_1|E) = \frac{P(E_1)\,P(E|E_1)}{P(A)} = \frac{1/2 * \frac{n(n-1)}{(n+2)(n+1)}}{\frac{2+n(n-1)}{2(n+2)(n+1)}} = \frac{n(n-1)}{2+n(n-1)} = \frac{6}{7}$$

$$7n^2 - 7n = 12 + 6n^2 - 6n$$

$$n^2 - n - 12 = 0$$

$$n = 4$$

16|1|25

The contents of box 1, 2 and 3 are as follows. A box is chosen at random and 2 balls are drawn, they happen to be white and red. What is the probability that they came from box 1, 2, 3

I	II	III
1 White	2 White	4 White
2 Black	1 Black	5 Black
3 red	1 Red	3 Red

Let E be the event that 2 balls drawn are white and red

E_1 = Box 1 selected E_2 = Box 2 selected E_3 = Box 3 selected

$$P(E_1) = P(E_2) = P(E_3) = 1/3$$

$$P(E/E_1) = \frac{{}^1C_1 \times {}^3C_1}{{}^6C_2} = \frac{3}{15} = \frac{1}{5}$$

$$P(E/E_2) = \frac{{}^2C_1 \times {}^1C_1}{{}^4C_2} = \frac{2 \times 1}{6} = \frac{1}{3}$$

$$P(E/E_3) = \frac{{}^4C_1 \times {}^3C_1}{{}^{12}C_2} = \frac{4 \times 3}{66} = \frac{12}{66} = \frac{2}{11}$$

$$P(A) = \frac{1}{3}\cdot\frac{1}{5} + \frac{1}{3}\cdot\frac{1}{3} + \frac{1}{3}\cdot\frac{2}{11}$$

$$= 1/15 + 1/9 + 2/33$$

$$= \frac{118}{445}$$

$$P(E_1/E) = \frac{P(E_1)P(E/E_1)}{\Sigma P(E_i)P(E/E_i)} = \frac{1/3 \cdot 1/5}{118/495} = \frac{33}{118}$$

$$P(E_2/E) = \frac{P(E_2)P(E/E_2)}{\Sigma P(E_i)P(E/E_i)} = \frac{1/3 \cdot 1/3}{118/495} = \frac{55}{118}$$

$$P(E_3/E) = \frac{P(E_3)P(E/E_3)}{\Sigma P(E_i)P(E/E_i)} = \frac{1/3 \cdot 2/11}{118/495} = \frac{30}{118}$$

From a vessel containing 3 white 5 black balls, 4 balls are transferred into an empty vessel, and from this vessel a ball is drawn and found to be white, what is the probability that 3 are white and 1 is black.

A = white ball is drawn

E_1 = 3 white 1 black

E_2 = 2 white 2 black

E_3 = 1 white 3 black

$${}^nC_r = \frac{n!}{(n-r)!\,r!}$$

$P(A|E_1) = ?$

$$P(E_1) = \frac{{}^3C_3 \times {}^5C_1}{{}^8C_4} = \frac{1 \times 5}{70} = \frac{5}{70} \qquad P(E_1/A) = \frac{3}{4}$$

$$P(E_2) = \frac{{}^3C_2 \times {}^5C_2}{{}^8C_4} = \frac{3 \times 10}{70} = \frac{3}{7} \qquad P(E_2/A) = \frac{2}{4}$$

$$P(E_3) = \frac{{}^3C_1 \times {}^5C_3}{{}^8C_4} = \frac{3 \times 10}{70} = \frac{3}{7} \qquad P(E_3/A) = \frac{1}{4}$$

$$P(A) = \frac{15}{280} + \frac{3}{14} + \frac{3}{28} = \frac{105}{280} = \frac{15}{40} = \frac{3}{8}$$

$$P(A|E_1) = \frac{P(E_1) \cdot P(E_1/A)}{P(A)} = \frac{5}{70} \cdot \frac{3}{4} \cdot \frac{8}{3} = \frac{5}{35} = \frac{1}{7}$$

1.2 Probability Distribution

Random Variable: is a function which assigns a real number to an outcome of an experiment

eg: 3 coins are tossed.

$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$

$n(S) = 2^3 = 8$

Let x = no of tails

$x: 0 \ 1 \ 2 \ 3$

$P(x=0) = 1/8$ $\quad P(x=2) = 3/8$ $\quad \Sigma P_i = 1$

$P(x=1) = 3/8$ $\quad P(x=3) = 1/8$

$P(x = x_i) \rightarrow$ Pmf $\rightarrow$ probability mass function (of a discrete random variable x)

& $(X, P(x=x_i)) \rightarrow$ Probability distribution of a random variable x

eg 2 dice are thrown, write down the probability distribution of sum of the numbers appearing on the top of the dice.

$S = \{11 \ 12 \ 13 \ 14 \ 15 \ 16$
$21 \ 22 \ 23 \ 24 \ 25 \ 26$
$31 \ 32 \ 33 \ 34 \ 35 \ 36$
$41 \ 42 \ 43 \ 44 \ 45 \ 46$
$51 \ 52 \ 53 \ 54 \ 55 \ 56$
$61 \ 62 \ 63 \ 64 \ 65 \ 66\}$

X = Sum of numbers appearing on top.

$X: 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9 \ 10 \ 11 \ 12$

$P(X=x_i) = \frac{1}{36} \ \frac{2}{36} \ \frac{3}{36} \ \frac{4}{36} \ \frac{5}{36} \ \frac{6}{36} \ \frac{5}{36} \ \frac{4}{36} \ \frac{3}{36} \ \frac{2}{36} \ \frac{1}{36}$

A random variable x has following probability distribution

x :	0	1	2	3	4	5	6	7
$P(x=x_i)$	0	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2+k$

find ① k ② $P\left(\frac{1.5 < x < 4.5}{x > 2}\right)$ ③ Smallest λ, such that $P(x \le \lambda) > 1/2$

Since $\Sigma P_i = 1 \Rightarrow$

$0 + k + 2k + 2k + 3k + k^2 + 2k^2 + 7k^2 + k = 1$

$10k^2 + 9k - 1 = 0$

$10k^2 + 10k - k - 1 = 0 \Rightarrow 10k(k+1) - 1(k+1) = 0 \Rightarrow k = -1, \ 1/10 \quad \therefore k = 1/10$ ①

② $P\left(\frac{1.5 < x < 4.5}{x > 2}\right) = \frac{P[(1.5 < x < 4.5) \cap (x > 2)]}{P(x > 2)}$

$\frac{P(3,4)}{P(3,4,5,6,7)} = \frac{P(x=3) + P(x=4)}{P(x=3) \cdots + + \cdots P(x=7)}$

$= \frac{5k}{10k^2 + 6k} = \frac{5}{10k+6} = \frac{5}{7}$

③ if $\lambda = 1 : P(x \le 1) = k = 1/10 \ne 1/2$

if $\lambda = 2 : P(x \le 2) = 3k = 3/10 \ne 1/2$

if $\lambda = 3 : P(x \le 3) = 5k = 1/2 \ne 1/2$

if $\lambda = 4 : P(x \le 4) = 8k = 8/10 > 1/2$

$\therefore$ Smallest value $= 4$

A random variable takes values $-2, -1, 0, 1, 2$ such that $P(x>0) = P(x=0) = P(x<0)$ $P(x=-2) = P(x=-1)$ $P(x=2) = P(x=1)$ find ① $P\left(\frac{-1 \le x \le 1}{-2 \le x \le 0}\right)$ ② $P\left(\frac{x=1}{0 \le x \le 2}\right)$

Let $P(x=0) = k$

We have $\underbrace{P(x=-2) + P(x=-1)}_{P(x<0)} + P(x=0) + \underbrace{P(x=1) + P(x=2)}_{P(x>0)} = 1$

$P(x<0) = P(x=0) = P(x>0) = k$

$3k = 1 \implies k = 1/3$

$k = P(x<0)$

$= P(x=-2) + P(x=-1)$

$1/3 = 2y \implies y = 1/6$

$k = P(x>0)$

$= P(x=2) + P(x=1)$

$\implies z = 1/6$

x :	-2	-1	0	1	2
$P(x)$:	$1/6$	$1/6$	$1/3$	$1/6$	$1/6$

① $P\left(\frac{-1 \le x \le 1}{-2 \le x \le 0}\right) = \frac{P[(-1 \le x \le 1) \cap (-2 \le x \le 0)]}{P[-2 \le x \le 0]} = \frac{P(-1 \le x \le 0)}{P(-2 \le x \le 0)} = \frac{1/2}{2/3} = \frac{1}{2} \cdot \frac{3}{2}$

$= \frac{3}{4}$

② $P\left(\frac{x=1}{0\le x\le 2}\right) = \frac{P(x=1)\cap(0\le x\le 2)}{P(0\le x\le 2)} = \frac{1/6}{2/3} = \frac{1}{6}\cdot\frac{3}{2} = \frac{1}{4}$

17/1/25

$X: \; x_1 \quad x_2 \quad --- \quad x_n$

$P(X=x_i): P_1 \quad P_2 \quad -- \quad P_n$

$P(X=x_i) \rightarrow$ Pmf

$(X, P(X_i)) \rightarrow$ Probability distribution

Cumulative Distribution Function

Let X be a discrete random variable, consider $[F(x) \;\; CDF]$

$F(x_i) = P(x \le x_i)$

$F(x_1) = P(x \le x_1)$

$F(x_2) = P(x \le x_1) + P(x \le x_2)$

$F(x_3) = P(x \le x_1) + P(x \le x_2) + P(x \le x_3)$

$\boxed{\therefore F(x_n) = P(x \le x_3) \; --- \; P(x \le x_n)}$

'F' is called cumulative distribution function at a random discrete variable 'x'.
Cumulative distribution function is cdf.

eg if $x: 1 \; 2 \; 3 \; 4$ such that $2P(x=1) = 3P(x=2)$
$= P(x=3) = 5P(x=4)$ find pmf and cdf.
$= k$

x :	1	2	3	4
$P(x=x_i)$:	$k/2$	$k/3$	k	$k/5$
$P(x)$ =	$15/61$	$10/61$	$30/61$	$6/61$

$\frac{k}{2} + \frac{k}{3} + k + \frac{k}{5} = 1$

$15k + 10k + 30k + 6k = 30$

$k = 30/61$

$F(x) = \frac{15}{61} \quad \frac{25}{61} \quad \frac{55}{61} \quad \frac{61}{61} = 1$

keep adding consecutively.

Probability Density Function [f(x) PDF]

Consider a continuous function $y = f(x)$ $\because$ $a \le x \le b$ satisfying the conditions:

① f is integrable

② $f(x) \geq 0$ ③ $\boxed{\int_a^b f(x)\,dx = 1}$

④ $\boxed{\int_a^b f(x)\,dx = P(\alpha \le x \le \beta)\ \&\ (a < \alpha < \beta < b)}$

Then f is called probability density function (pdf) of a continuous random variable x.

eg if $f(x) = kx^2$, $0 \le x \le 2$ is a pdf.
find ① k ② $P(0.2 \le x \le 0.5)$ ③ $P[(x \geq 3/4)/(x \geq 1/2)]$

Since f(x) is a pdf

$\therefore \int_0^2 f(x)\,dx = 1$

$= \int_0^2 kx^2\,dx = 1 \Rightarrow \left(\frac{kx^3}{3}\right)_0^2 = 1$

① $\frac{8k}{3} - \frac{0}{3} = 1 \Rightarrow k = 3/8$

② $P(0.2 \le x \le 0.5)$

$= \int_{0.2}^{0.5} \frac{3}{8} x^2\,dx \Rightarrow \frac{3}{8}\left[\frac{x^3}{3}\right]_{0.2}^{0.5} \Rightarrow \frac{3}{8}\left[\frac{0.125}{3} - \frac{0.008}{3}\right]$

$\Rightarrow \frac{0.117}{8} = \underline{0.0146}$

③ $P[(x \geq 3/4)/(x \geq 1/2)]$

$P[(x \geq 3/4)/(x \geq 1/2)] = \frac{P(x \geq 3/4) \cap P(x \geq 1/2)}{P(x \geq 1/2)}$

$= \frac{P(x \geq 3/4)}{P(x \geq 1/2)} = \frac{\int_{3/4}^2 f(x)\,dx}{\int_{1/2}^2 f(x)\,dx}$

$= \left[\frac{3}{8} \cdot \frac{x^3}{3}\right]_{3/4}^2 \Big/ \left[\frac{3}{8} \cdot \frac{x^3}{3}\right]_{1/2}^2$

$= \frac{3}{8}\left(\frac{7.57}{3}\right) \Big/ \left[\frac{3}{8} \cdot \frac{7.87}{3}\right]$

$= \frac{7.57}{8} \cdot \frac{8}{7.87} = 0.96$

Cumulative Distribution Function for continous random variable 'x'.

Let X be a continuous random variable with pdf $f(x)$, then $\boxed{F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\,dx}$

this 'F' is called cumulative distribution function for the continous random variable.

Properties

1. $0 \le F(x) \le 1$
2. $F'(x) = f(x)$
3. $\boxed{P(a \le x \le b) = F(b) - F(a)}$

Q CDF of a continous random variable x is given by

$$F(x) = \begin{cases} 0 & x < 0 \\ x^2 & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

find (i) pdf (ii) $P(1/2 \le x \le 4/5)$

$$\Rightarrow f(x) = F'(x)$$

$$= \begin{cases} 0 & x < 0 \\ 2x & 0 \le x \le 1 \\ 0 & x > 1 \end{cases}$$

(ii) $P\left(1/2 \le x \le 4/5\right) = F(4/5) - F(1/2)$

$$= (4/5)^2 - (1/2)^2$$

$$= \frac{16}{25} - \frac{1}{4}$$

$$= \frac{64-25}{100} = \frac{39}{100}$$

find cdf for the following pdf

$$f(x) = \begin{cases} \frac{1}{2} x^2 \cdot e^{-x}, & 0 < x < \infty \\ 0, & \text{otherwise} \end{cases}$$

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

$$= \int_0^x \frac{1}{2} x^2 e^{-x}\,dx = \frac{1}{2}\left[x^2(-e^{-x}) - 2x(+e^{-x}) + 2(-e^{-x})\right]_0^x$$

$$= \frac{1}{2}\left[x^2e^{-x} + 2xe^{-x} + 2e^{-x}\right]$$

$$= \frac{x^2e^{-x}}{2} + xe^{-x} + e^{-x}$$

20|1|25

Expectation and Variance

Let x be a discrete random variable then expectation of x $\boxed{E(x) = \sum_i x_i p_i = \text{Mean}(x)}$

and variance of x $\boxed{\text{var}(x) = E(x^2) - [E(x)]^2 = \sum_i x_i^2 p_i - [E(x)]^2}$

Note: For continuous $\Sigma \to \int$

If X is continuous random variable then $\boxed{E(x) = \int x f(x) dx}$

and $\boxed{\text{var}(x) = E(x^2) - [E(x)]^2 = \int_{-\infty}^{\infty} x^2 f(x) dx - [E(x)]^2}$

eg
X : x_1 x_2 x_3
P : p_1 p_2 p_3
f : f_1 f_2 f_3

$$\text{Mean} = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{\Sigma x_i p_i}{\underbrace{\Sigma p_i}_{=1}}$$

eg A fair coin is tossed till head appears, what is the expectation of the no of tosses required.

	H	TH	TTH	
x :	1	2	3	.....
$P(X=x)$:	$1/2$	$1/2^2$	$1/2^3$	

x : No of tosses required

$$E(x) = \sum_i x_i p_i$$

$$S = 1\times\frac{1}{2} + 2\times\frac{1}{2^2} + 3\times\frac{1}{2^3} + \text{-----}$$

$$\frac{S}{2} = \frac{1}{2^2} + \frac{2}{2^3} + \frac{3}{2^4} \text{-----}$$

$$S - \frac{S}{2} = \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} \text{-----}$$

..... GP sum $= a/1-r$

$$\frac{S}{2} = \frac{1/2}{1-1/2} = 1$$

$S = 2$ ....converging

$\therefore$ $E(x) = 2$

Q find the E(n) ① of the sum ② of the product of number appearing on throw of n dice

1 2 - - - - n

X_i : no appearing on the i^{th} die

:	1	2	3	4	5	6
$P(x=x_i)$	1/6	1/6	1/6	1/6	1/6	1/6

$$E(x_i) = \sum_{i=1}^{6} x_i p_i$$

$$= \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} = \frac{1}{6}[1+2+3+4+5+6]$$

$$= \frac{1}{6}\left[\frac{6 \times 7}{2}\right] = \frac{7}{2}$$

$\frac{n(n+1)}{2}$ = sum of n natural numbers

$$\therefore E\left(\sum_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} E(x_i) = \sum_{i=1}^{n} 7/2 = \frac{7n}{2}$$

$$E\left(\sum_{i=1}^{n} x_i\right) = \frac{7n}{2}$$

$$E\left(\prod_{i=1}^{n} x_i\right) = \prod_{i=1}^{n} E(x_i) = \left(\frac{7}{2}\right)^n$$

$$E\left(\prod_{i=1}^{n} x_i\right) = \left(\frac{7}{2}\right)^n$$

Q find the expectation of number of failures preceding the first success in an infinite series of independent trials with probabilities P and Q of success and failure respectively

X = no of failures preceding first success

x :	0	1	2	3	- - - - -
$P(x=x)$:	p	qp	q^2p	q^3p	- - - - - - .

$(1-x)^{-1} = 1 + x + x^2 - - - -$

$(1-x)^{-2} = 1 + 2x + 3x^2 - - -$

$$E(x) = \sum x_i p_i$$

$$= qp + 2q^2p + 3q^3p + - - - - - -$$

$$p(1-q)^{-2} = \frac{qp}{(1-q)^2} = \frac{q}{p}$$

Q X is a continous random variable with pdf $f(x) = \begin{cases} kx^2(1-x^3) & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$

find ① k ② $P(0 \leq x \leq 1/2)$ ③ mean and var

① f is pdf

$$\int_0^1 f(x)\, dx = 1 \Rightarrow k \int_0^1 kx^2(1-x^3)\, dx = 1 \Rightarrow k\left[\frac{x^3}{3} - \frac{x^6}{6}\right]_0^1 = 1$$

$$k\left(\frac{1}{3}-\frac{a}{3}-\frac{1}{6}+\frac{a}{6}\right)=1$$

$$k=6$$

② $P(0 \le x \le 1/2) = \int_0^{1/2} f(x)\,dx$

$$\int_0^{1/2} 6x^2(1-x^3)\,dx = 6\left[\frac{x^3}{3}-\frac{x^6}{6}\right]_0^{1/2} \Rightarrow \frac{6}{6}\left[2x^3-x^6\right]_0^{1/2} = \left[\frac{1}{4}-\frac{1}{64}\right] = \frac{15}{64}$$

$$\frac{15}{64}$$

③ $E(x) = \int_0^1 x f(x)\,dx = ?$

$E(x^2) = \int_0^1 x^2 f(x)\,dx = ?$

$var(x) = E(x^2) - [E(x)]^2$

$$E(x) = \int_0^1 x \cdot 6x^2(1-x^3)\,dx = \int_0^1 6x^3 - 6x^6\,dx = 6\left[\frac{x^4}{4}-\frac{x^7}{7}\right]_0^1 = 6\left[\frac{1}{4}-\frac{1}{7}\right] = \frac{6\times 3}{28} = \frac{9}{14}$$

$$E(x^2) = \int_0^1 x^2 f(x)\,dx = \int_0^1 6x^4 - 6x^7\,dx = 6\left[\frac{x^5}{5}-\frac{x^8}{8}\right]_0^1 = 6\left[\frac{8-5}{40}\right] = \frac{18}{40} = \frac{9}{20}$$

$$var(x) = \frac{9}{20}-\frac{81}{196} = \frac{9}{4}\left(\frac{1}{5}-\frac{9}{49}\right) = \frac{9}{4}\left(\frac{49-45}{245}\right) = \frac{9}{245} = 0.0367$$

$E(x) = \frac{9}{14}$ $\quad var(x) = 0.037$ $\quad P(0 \le x \le 1/2) = 15/64$ $\quad k = 6$

23/1/25

Joint Probability Distribution

Let X and Y be 2 random variables, defined on the same sample space S. Then the function (x,y) that assigns a point in $\mathbb{R}^2 = \mathbb{R}\times\mathbb{R}$ (Cartesian product) is called 2-dimensional random variable

$\mathbb{R}$ of random discrete variable (pmf, etc)

Joint Probability Mass function

If (x,y) two dimensional discrete random variables with probabilities P_{ij} satisfying the following conditions:

① $P_{ij} \ge 0, \forall\, i,j$

② $\boxed{\sum_{j=1}^{\infty}\sum_{i=1}^{\infty} P(x_i, x_j) = 1}$

Then this P is called joint Pmf of x,y and probability distribution of X, Y is given by:

X/Y	y_1	y_2	$\cdots$	y_n	Total
x_1	P_{11}	P_{12}	$\cdots$	P_{1n}	$P_i = \sum_{j=1}^{\infty} P_{ij}$
x_2	P_{21}	P_{22}	$\cdots$	P_{2n}	P_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
x_m	P_{m1}	P_{m2}	$\cdots$	P_{mn}	P_n
Total	P_1'	P_2'		P_m'	1

Marginal Probability Distribution of x is given by:

$X : x_1 \; x_2 \; \cdots\cdots \; x_m$

$P(X=x) : P_1 \; P_2 \; \cdots\cdots \; P_m$ y varies

$Y : y_1 \; y_2 \; \cdots \; y_n$

$P(Y=y) = P_1' \; P_2' \; \quad P_n'$ x varies

Conditional Probability

$$\boxed{P(X=x_i/Y=y_j) = \frac{P(X=x_i, Y=y_j)}{P(Y=y_j)}}$$

The joint pmf of X, Y is given by $P(X=x, Y=y) = \dfrac{x+3y}{24}$, $x, y = 1, 2$
find jpd of X, Y and marginal pd of X, Y

X/Y	1	2	Total
1	$4/24$	$5/24$	$9/24$
2	$7/24$	$8/24$	$15/24$
Total	$11/24$	$13/24$	1

j.p.d

$X : \; 1, \; 2$

$P(X=x) : \frac{11}{24} \; \frac{13}{24}$

$Y : \; 1, \; 2$

$P(Y=y) : 9/24 \; 15/24$

JPD of X_1 and X_2 is given by $P(X_1 = x_1, X_2 = x_2) = \frac{1}{27}(x_1 + 2x_2)$ where $x_1, x_2 = 0, 1, 2$ where mp of x_1 and x_2 in terms of x_1 and x_2.

Marginal prob of X_1

$$P(X_1 = x_1) = \sum_{x_2=0}^{2} P(X_1 = x_1, X_2 = x_2)$$

$$= \frac{1}{27}\sum_{x_2=0}^{2}(x_1 + 2x_2) = \frac{1}{27}(x_1 + x_1 + 2 + x_1 + 4) = \frac{3x_1 + 6}{27} = \frac{x_1 + 2}{9}$$

$$P(X_2 = x_2) = \sum_{x_1=0}^{2} P(X_1 = x_1, X_2 = x_2)$$

$$= \frac{1}{27} \sum_{x_1=0}^{2} (x_1 + 2x_2) = \frac{1}{27}(0 + 2x_2 + 1 + 2x_2 + 2 + 2x_2) = \frac{6x_2+3}{27} = \frac{2x_2+1}{9}$$

X_1 :	0	1	2
$P(X_1=x_1)$:	$2/9$	$3/9$	$4/9$

Y_1 :	0	1	2
$P(Y_1=y_1)$:	$1/9$	$3/9$	$5/9$

Three balls are drawn at random without replacement from a box containing 2 white, 4 black and 3 red. X denotes white balls drawn and Y is red balls drawn. ① find jpd of X & Y ② $P(X \le 1)$ ③ $P(X \le 1, Y \le 2)$ ④ $P(Y \le 2 \text{ given } X \le 1)$ ⑤ $P(X+Y=2)$ ⑥ X & Y independent.

Y \ X	0	1	2	Total
0	$1/21$	$1/7$	$1/21$	$5/21$
1	$3/14$	$2/7$	$1/28$	$15/28$
2	$1/7$	$1/7$	0	$3/14$
3	$1/84$	0	0	$1/84$
Total	$35/84$	$42/84$	$7/84$	1

$$P(X=0, Y=0) = \frac{{}^4C_3}{{}^9C_3} = \frac{4}{84} = \frac{1}{21}$$

$$P(X=0, Y=1) = \frac{{}^3C_1 \times {}^4C_2}{{}^9C_3} = \frac{18}{84} = \frac{3}{14}$$

$$P(x=1, y=0) = \frac{{}^2C_1 \times {}^4C_2}{{}^9C_3} = \frac{2\times 6}{84} = \frac{1}{7}$$

$$P(x=2, y=0) = \frac{{}^2C_2 \times {}^4C_1}{{}^9C_3} = \frac{1\times 4}{84} = \frac{1}{21}$$

$$P(x=1, y=1) = \frac{{}^2C_1 \times {}^3C_1 \times {}^4C_1}{{}^9C_3} = \frac{2\times3\times4}{84} = \frac{2}{7}$$

$$P(x=2, y=1) = \frac{{}^2C_2 \times {}^5C_1}{{}^9C_3} = \frac{1\times 3}{84} = \frac{1}{28}$$

$$P(x=0, y=2) = \frac{{}^4C_2 \times {}^3C_2}{{}^9C_3} = \frac{6\times 3}{84} = \frac{3}{14}$$

② $P(x \le 1) = P(x=0) + P(x=1)$

$$= \frac{35}{84} + \frac{42}{84} = \frac{77}{84}$$

③ $P(x \le 1, y \le 2) = \sum_{j=0}^{2} P(x=0, y=j) + \sum_{j=0}^{2} P(x=1, y=j)$

$$= \left(\frac{1}{21} + \frac{3}{14} + \frac{1}{7}\right) + \left(\frac{1}{7} + \frac{2}{7} + \frac{1}{14}\right)$$

$$= \frac{19}{21}$$

④ $P(y \le 2 / x \le 1) = \dfrac{P(x \le 1, y \le 2)}{P(x \le 1)}$

$$= \frac{19/21}{77/84} = \frac{76}{77}$$

⑤ $P(x+y\le 2) = \sum_{j=0}^{2} P(x=0, y=j) + \sum_{j=0}^{1} P(x=1, y=j) + P(x=2, y=0)$

$= \left(\frac{1}{21} + \frac{3}{14} + \frac{1}{7}\right) + \left(\frac{1}{7} + \frac{2}{7}\right) + \frac{1}{21}$

$= \frac{37}{42}$

⑥ If x, y are independent then $P(x,y) = P(x)\,P(y)$

$P(0,1) = 3/4$

$P(0) \times P(1) = \frac{35}{84} \times \frac{15}{28}$

$\because P(0,1) \ne P(0)P(1) \Rightarrow x, y$ are not independent

24/1/25

2-D continous probability distribution

Let (x,y) be 2-dimensional continous random variable

Let $f_{xy}(x,y)$ be a function satisfying the following condition

1. $f_{xy}(x,y) \ge 0$

2. $\boxed{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{xy}(x,y)\,dx\,dy = 1}$

3. $\boxed{\int_{a}^{b}\int_{c}^{d} f_{xy}(x,y)\,dx\,dy = P(a\le x\le b,\ c\le y\le d)}$

The $f_{xy}(x,y)$ is called 2-dimensiond pdf.

Marginal pdf of x is given by:

$$\boxed{f_x(x) = \int_{-\infty}^{\infty} f_{xy}(x,y)\,dy,\quad f_x(x)\ge 0 \quad \text{and} \quad \int_{-\infty}^{\infty} f_x(x)\,dx = 1}$$

Marginal pdf of y is given by :

$$\boxed{f_y(y) = \int_{-\infty}^{\infty} f_{xy}(x,y)\,dx,\quad f_y(y)\ge 0 \quad \text{and} \quad \int_{-\infty}^{\infty} f_y(y)\,dy = 1}$$

2-D pdf is given by: $f_{xy}(x,y) = \begin{cases} 15e^{-3x-5y}, & x>0,\ y>0 \\ 0, & \text{otherwise} \end{cases}$

find ① $P(1<x<2,\ 0.2<y<0.3)$ ② $P(x<2,\ y>0.2)$ ③ Marginal pdf of x and y.

① $P(1<x<2,\ 0.2<y<0.3)$

$= \int_{0.2}^{0.3}\int_{1}^{2} 15e^{-3x-5y}\,dx\,dy = 15\int_{0.2}^{0.3} e^{-5y}\left(\frac{-e^{-3x}}{3}\right)_1^2 dy$

$$= -5\int_{0.2}^{0.3} e^{-5y}(e^{-6}-e^{-3})\,dy = (e^{(-6)}\cdot e^{(-3)})(e^{-1.5})$$

(2) $P(x<2,\ y>0.2) = \int_0^2\int_{0.2}^{\infty} 15e^{-3x-5y}\,dx\,dy$

$$= 15\int_0^2 e^{-3x}\int_{0.2}^{\infty} e^{-5y}\,dx\,dy$$

$$= -3\int_0^2 e^{-3x}\left[e^{-5(\infty)+0.2(5)}\right]dx$$

$$= -3\int_0^2 e^{-3x}[e^{-\infty}\cdot e]\,dx$$

$$= -3e\int_0^2 e^{-3x}\,dx$$

27/1/25

Given $f_{x,y}(x,y) = \begin{cases} c\,x(x-y), & 0<x<2,\ -x<y,\ -x<y<x \\ 0, & \text{otherwise} \end{cases}$

(i) evaluate c (ii) find $f_x(x)$ (iii) $f_{y/x}(y/x)$ (iv) $f_y(y)$

$\Rightarrow$ (1) $\because \iint f_{xy}(x,y)\,dx\,dy = 1$

$\Rightarrow \int_0^2\int_{y=-x}^{x} c\,x(x-y)\,dx\,dy = 1 \Rightarrow \int_0^2\int_{y=-x}^{x} Cx^2 - Cxy\,dx\,dy = 1 \Rightarrow \int_0^2\left[Cx^3 - \frac{Cx^3}{2} + Cx^3 + \frac{Cx^3}{2}\right]dx = 1$

$\Rightarrow \int_0^2 2Cx^3\,dx = 1 \Rightarrow c = 1/8$

(2) $f_x(x)$ = marginal pdf of x

$$= \int_{y=-x}^{x} f_{xy}(x,y)\,dy \qquad = \int_{y=-x}^{x} 1/8 \cdot x(x-y)\,dy$$

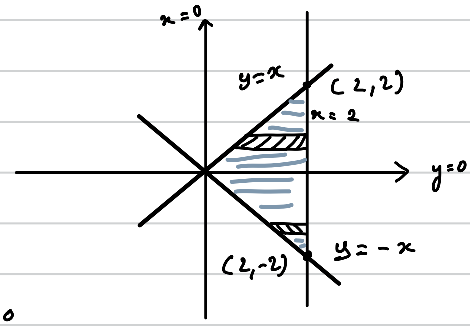
$$= \frac{x^3}{4}, \qquad 0<x<2$$

(3) $f_{y/x}(y/x)$

$$\frac{f_{xy}(x,y)}{f_x(x)} = \frac{1/8\ x(x-y)}{x^3/4}$$

$$= \frac{x-y}{2x^2}$$

④ $f_y(y)$ = marginal pdf of y



$$= \int_{x=x_1}^{x_2} f_{xy}(x,y)\,dx$$

$$= \begin{cases} \displaystyle\int_{x=y}^{2} \frac{1}{8}\, x(x-y)\,dx \\ \displaystyle\int_{x=-y}^{2} \frac{1}{8}\, x(x-y)\,dx \end{cases} = \begin{cases} \frac{1}{3} - \frac{y}{4} + \frac{5}{48}y^3, & 0 \le y \le 2 \\ \frac{1}{3} - \frac{y}{4} + \frac{1}{48}y^3, & -2 \le y \le 0 \end{cases}$$

$$f_y(y) = \begin{cases} \frac{1}{3} - \frac{y}{4} + \frac{5}{48}y^3, & 0 \le y \le 2 \\ \frac{1}{3} - \frac{y}{4} + \frac{1}{48}y^3, & -2 \le y \le 0 \end{cases}$$

Given $f_{xy}(x,y) = xy^2 + \frac{x^2}{8}$ $\quad 0 \le x \le 2 \quad 0 \le y \le 1$

Compute ① $P(x>1)$ ② $P(y<1/2)$ ③ $P(x>1,\, y<1/2)$ ④ $P(x>1 / y<1/2)$ ⑤ $P(x<y)$ ⑥ $P(x+y\le 1)$

① $P(x>1) = \int_{y=0}^{1}\int_{x=1}^{2} f_{xy}(x,y)\,dx\,dy$

$$= \int_{y=0}^{1}\int_{x=1}^{2} xy^2 + \frac{x^2}{8}\,dx\,dy \Rightarrow \int_{y=0}^{1}\left[\frac{x^2y^2}{2} + \frac{x^3}{24}\right]_1^2 dy = \int_0^1\left(\frac{4y^2}{2} + \frac{8}{24} - \frac{y^2}{2} - \frac{1}{24}\right)dy$$

$$= \int_0^1 \frac{36y^2+7}{24}\,dy = \frac{1}{24}\left[\frac{36y^3}{3} + 7y\right]_0^1 = \frac{1}{24}\left[12+7\right] = \frac{19}{24}$$

② $P(y<1/2) = \int_{y=0}^{1/2}\int_{x=0}^{2} xy^2 + \frac{x^2}{8}\,dx\,dy$

$$\int_0^{1/2}\left[\frac{x^2y^2}{2} + \frac{x^3}{24}\right]_0^2 dy = \int_0^{1/2} \frac{4y^2}{2} + \frac{8}{24}\,dy = \int_0^{1/2} 2y^2 + \frac{1}{3}\,dy = \left[\frac{2y^3+y}{3}\right]_0^{1/2}$$

$$= \frac{1/4 + 1/2}{3} = \frac{3/4}{3} = \frac{1}{4}$$

③ $P(x>1,\, y<1/2)$ $\qquad = 5/24$

$$= \int_{y=0}^{1/2}\int_{x=1}^{2} xy^2 + \frac{x^2}{8}\,dx\,dy = \int_{y=0}^{1/2}\left[\frac{x^2y^2}{2} + \frac{x^3}{24}\right]_1^2 dy = \int_0^{1/2}\left[\frac{4y^2}{2} + \frac{8}{24} - \frac{y^2}{2} - \frac{1}{24}\right]dy$$

$$= \left[\frac{2y^3}{3} + \frac{y}{3} - \frac{y^3}{6} - \frac{y}{24}\right]_0^{1/2} = \left[\frac{1/4}{3} + \frac{1/2}{3} - \frac{1/8}{6} - \frac{1/2}{24}\right] = \left[\frac{3/4}{3} - \frac{1}{24}\right] = \frac{6-1}{24}$$

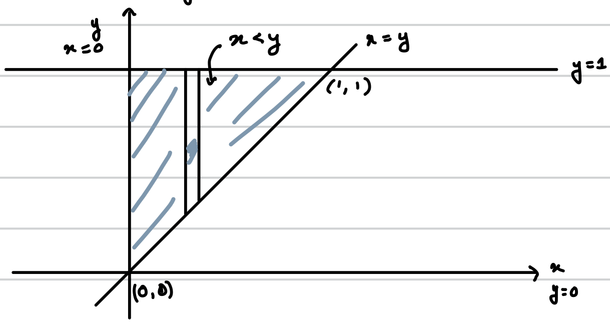
$$= \frac{5}{24}$$

④ $P(x>1 / y<1/2)$

$$= \frac{P(x>1,\ y<1/2)}{P(y<1/2)} = \frac{5/24}{1/4}$$

$$= \frac{5}{6}$$

⑤ $P(x<y) = \int_{x=0}^{1}\int_{y=x}^{1} xy^2 + \frac{x^2}{8}\, dy\, dx$



$$= \int_0^1 \left[\frac{xy^3}{3} + \frac{x^2y}{8}\right]_x^1 dx$$

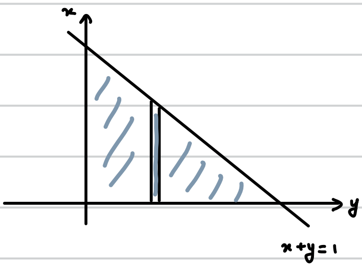
$$= \int_0^1 \frac{x}{3} + \frac{x^2}{8} - \frac{x^4}{3} - \frac{x^3}{8}\, dx$$

$$= \int_0^1 \left[\frac{1}{3}\left[\frac{x^2}{2} - \frac{x^5}{5}\right] + \frac{1}{8}\left[\frac{x^3}{3} - \frac{x^4}{4}\right]\right]$$

$$= \left[\frac{x^2}{6} - \frac{x^5}{15} + \frac{x^3}{24} - \frac{x^4}{32}\right]_0^1 = \left[\frac{1}{6} - \frac{1}{15} + \frac{1}{24} - \frac{1}{32}\right]$$

$$= \frac{5-2}{30} + \frac{(4-3)}{96} = \frac{3}{30} + \frac{1}{96} = \frac{1}{10} + \frac{1}{96} = \frac{106}{960} = \frac{53}{480}$$

⑥ $P(x+y \le 1) = \int_{x=0}^{1}\int_{y=0}^{1-x} xy^2 + \frac{x^2}{8}\, dy\, dx$



$$= \int_{x=0}^{1} \left[\frac{xy^3}{3} + \frac{x^2y}{8}\right]_0^{1-x} dx$$

$$= \int_{x=0}^{1} \frac{x(1-x)^3}{3} + \frac{x^2(1-x)}{8}\, dx$$

$$= \int_{x=0}^{1} \frac{x - x^4 - 3x^2 + 3x^3}{3} + \frac{x^2 - x^3}{8}\, dx$$

$$= \left[\frac{x^2}{6} - \frac{x^5}{15} - \frac{x^3}{3} + \frac{x^4}{4} + \frac{x^3}{24} - \frac{x^4}{32}\right]_0^1$$

$$= \frac{13}{480}$$

1.4 Binomial Distribution

(discrete)

Let X be a discrete random variable then x follows a binomial distribution with parameter n and p.
(n → no of trials, p → no of successes)

if pmf of x is given by $\boxed{P(X=x) = {}^nC_x\, p^x q^{n-x}}$ where $\boxed{q = 1-p}$

it is called $\boxed{X \sim B(n, p)}$ (∼ follows)

$$S.d. = \sqrt{Var(x)} = \sqrt{npq}$$

Mean and variance of Binomial Distribution

$\boxed{E(x) = \Sigma x_i p_i = np}$ (Mean)

$\boxed{Var(x) = E(x^2) - [E(x)]^2 = npq}$

Properties

① If $x_1 \sim B(n_1, p_1)$, $x_2 \sim B(n_2, p_2)$ then $x_1 + x_2 \sim B(n_1+n_2, p)$ where $p = p_1 = p_2$

eg if x is binomially distributed where $E(x) = 2$, $Var(x) = 4/3$ find probability distribution of x and P(at least 1 success)

Solution Let X = no of successes

Given $E(x) = 2 = np$
$Var(x) = \frac{4}{3} = npq$
$\Big\} \rightarrow q = \frac{2}{3}$, $p = 1-q = \frac{1}{3}$, $n = 6$

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

$$= {}^6C_x \left(\tfrac{1}{3}\right)^x \left(\tfrac{2}{3}\right)^{6-x}$$

X:	0	1	2	3	4	5	6
P(X=x):	$\frac{64}{729}$	$\frac{192}{729}$	$\frac{240}{729}$	$\frac{160}{729}$	$\frac{60}{729}$	$\frac{12}{729}$	$\frac{1}{729}$

---- put x = 0 to 6

$$P(\text{at least 1 success}) = 1 - P(\text{no success})$$
$$= 1 - 64/729$$
$$= 665/729$$

Q Probability of a man aged 60 living upto 70 is 0.65. What is the probability that out of 10 such men, 7 live upto 70.

Solution: $n = 10$ X = no of men who will live upto 70. $p = 0.65$

$P(X \geq 7)$

$$P(X=x) = {}^nC_x\, p^x q^{n-x}$$
$$= {}^{10}C_x (0.65)^x (0.35)^{10-x}$$

$$P(X \geq 7) = P(x=7) + P(x=8) + P(x=9) + P(x=10)$$
$$= 0.51381$$

$$= 0.25221 + 0.17565 + 0.07249 + 0.01346$$

Q Ratio of probability of 3 successes in 5 independent trials to the probability of 2 successes is $1/4$. Find the probability of 4 successes in 6 independent trials.

Solution $n = 5$

$$\frac{P(X=3)}{P(X=2)} = \frac{1}{4} = \frac{{}^5C_3\, p^3 q^2}{{}^5C_2\, p^2 q^3} = \frac{1}{4}$$

$4p = q$

$4p = 1-p \Rightarrow p = 1/5,\ q = 4/5$

$\therefore P(X=4) = {}^nC_x\, p^x q^{n-x}$

$= {}^6C_4 (1/5)^4 (4/5)^2$

$= 48/3125$

$= 0.01536$

Q Communication system has 4 components, each of which functions with probability p, the total system will work efficiently if atleast half of its components are functioning. For what value of p a 5 component system will work more efficiently than a 3 component system.

Solution : Let X = no of components

A 5-component system will work more efficiently than a 3-component system if
P(5-component system working efficiently) > P(3-component system working efficiently)

i.e. $P(X=3,4,5) > P(X=2,3)$

i.e. $\sum_{x=3}^{5} {}^5C_x\, p^x q^{5-x} > \sum_{x=2}^{3} {}^3C_x\, p^x q^{3-x}$

i.e. ${}^5C_3 p^3 q^2 + {}^5C_4 p^4 q + {}^5C_5 p^5 > {}^3C_2 p^2 q + {}^3C_3 p^3$

i.e. $10p^3(1-p)^2 + 5p^4(1-p) + p^5 - 3p^2(1-p) - p^3 > 0$

i.e. $10p^3 - 20p^4 + 10p^5 + 5p^4 - 5p^5 + p^5 - 3p^2 + 3p^3 - p^3 > 0$

i.e. $12p^3 - 15p^4 + 6p^5 - 3p^2 > 0$

i.e. $3p^2(4p - 5p^2 + 2p^3 - 1) > 0$

i.e. $3p^2(p - 1/2)(p-1)^2 > 0$

i.e. $p - 1/2 > 0$

i.e. if $\underline{p > 1/2}$

> if sum of all coefficients = 0, root = 1
>
> if sum of even = odd coefficients, root = -1

Q Let X, Y be 2 independent binomial variants with parameters $n_1 = 6, p_1 = 1/2, n_2 = 4, p_2 = 1/2$ find probability $(X+Y=3)$.

Solution $p = p_1 = p_2$

$X \sim B(n_1, p_1) = X \sim B(6, 1/2)$

$Y \sim B(n_2, p_2) = Y \sim B(4, 1/2)$

$Z = X + Y$

$Z \sim B(n_1 + n_2, p) \Rightarrow Z \sim B(10, 1/2)$

$\therefore P(z=3) = {}^{10}C_3 (1/2)^3 (1/2)^7$

$= \frac{15}{128} = 0.11718$

31/1/25

Q. 7 dice are thrown 729 times. How many times do you expect at least 4 dice to show 3 or 5.

$n = 7 \qquad X =$ no of dice showing 3 or 5

We want to find $P(X \geq 4)$

$p =$ probability of success $= 1/6 + 1/6 = 2/6 = 1/3$

$q = 1 - p = 1 - 1/3 = 2/3$

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

$= {}^{7}C_x (1/3)^x (2/3)^{7-x}$

$P(X \geq 4) = P(X=4) + P(X=5) + P(X=6) + P(X=7)$

$= {}^{7}C_4 (1/3)^4 (2/3)^3 + {}^{7}C_5 (1/3)^5 (2/3)^2 + {}^{7}C_6 (1/3)^6 (2/3)^1 + {}^{7}C_7 (1/3)^7 (2/3)^0$

$= \frac{379}{3^7}$

$\therefore$ Expected no of times at least 4 dices to show 3 or 5.

$= NP(X \geq 4)$

$= 729 \times \frac{379}{3^7}$

$\simeq 126$

Q. 3 fair coins are tossed 3000 times. Find frequency distribution of no of heads and no of tails also find mean and standard deviation of the distribution

$n = 3, \; N = 3000, \; X =$ no of heads

$p = q = 1/2$

X :	0	1	2	3
$P(X=x)$:	$1/8$	$3/8$	$3/8$	$1/8$
NP :	$3000/8$	$9000/8$	$9000/8$	$3000/8$
	$= 375$	$= 1125$	$= 1125$	$= 375$

mean $= E(x) = np = 3/2 = 1.5$

$s.d. = \sqrt{\frac{Var}{npq}} = \sqrt{3 \times 1/2 \times 1/2} = \frac{\sqrt{3}}{2} = 0.866$

Fitting of Binomial Distribution

Q 7 coins are tossed and no of heads obtained is noted. The experiment is repeated 128 times. Fit a binomial distribution if ① coins are unbiased ② nature of the coin is not known

X:	0	1	2	3	4	5	6	7	Total
Freq:	7	6	19	35	30	23	7	1	$128 = 2^7$

① $n = 7 \quad p = q = 1/2$

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

$= {}^7C_x (1/2)^x (1/2)^{7-x}$

X:	0	1	2	3	4	5	6	7
$P(X=x)$:	$1/2^7$	$7/2^7$	$21/2^7$	$35/2^7$	$35/2^7$	$21/2^7$	$7/2^7$	$1/2^7$
NP:	1	7	21	35	35	21	7	1

② $\bar{x} = \dfrac{\sum x_i f_i}{\sum f_i} = \dfrac{433}{128}$

$= 3.38 = np$

$3.38 = 7p$

$p = 0.48$

$q = 1 - p$

$= 1 - 0.48$

$= 0.52$

$P(X = x) = {}^7C_x (0.48)^x (0.52)^{7-x}$

X =	0	1	2	3	4	5	6	7
$P(X=x)$:	0.02	0.066	0.183	0.243	0.261	0.144	0.04	0.00588
	1.305	8.5	23.54	36.22	33.43	18.52	5.61	0.75
	1	9	24	36	33	19	6	1

3|2|25

Poissons Distribution (discrete)

A discrete random variable X, is said to follow poissons distribution with parameter 'm' if $P(X=x) = \frac{e^{-m} m^x}{x!}$, $x = 0, 1, 2 \cdots$

where $m = E(x) = \text{mean}(x) = \text{Var}(x) = np$ = finite

$$\therefore \boxed{\text{P.M.F of PD} = \frac{e^{-m} m^x}{x!}}$$

$n \to$ large $\to$ tends to ∞

$p \to$ small $\to$ tends to 0

if $x_1, x_2 \cdots x_n$ are poisson variants with parameters $m_1, m_2 \cdots m_n$. then $y = x_1 + x_2 + \cdots x_n$ is also a poisson variant with parameter $M = m_1 + m_2 + \cdots m_n$, but if $y = a_1 x_1 + a_2 x_2 + \cdots a_n x_n$ such that $(a_1, a_2 \cdots a_n \neq 1)$ then it is not a poissons variant.

Q If mean of the poissons distribution is 4. find $P(m - 2\sigma < x < m + 2\sigma)$

mean = $m = \text{Var}(x) = 4$

$\sigma = 2$

$\therefore P(m - 2\sigma < x < m + 2\sigma)$

$P(4 - 4 < x < 4 + 4) = P(0 < x < 8)$

$= P(x = 1, 2, \cdots 7)$

$= P(x=1) + P(x=2) \cdots + P(x=7)$

$= e^{-4}\left[\frac{4^1}{1!} + \frac{4^2}{2!} + \frac{4^3}{3!} + \frac{4^4}{4!} + \frac{4^5}{5!} + \frac{4^6}{6!} + \frac{4^7}{7!}\right]$

$= 0.93.$

Q A car hire firm has 2 cars. Then no of demands for a car on each day follows a poissons distribution with mean 1.5, Calculate the proportion (probability) that ① Neither car is used. ② Some demand is refused.

Let X = no of demands

$m = \text{mean}(x) = \text{Var}(x) = 1.5$

$\frac{e^{-1.5} \cdot (1.5)^x}{x!}$

1) $P(X=0)$ = Neither car is used

$\frac{e^{-1.5} \cdot (1.5)^0}{0!} = e^{-1.5}$

$= 0.22$

2) $P(X>2)$ = Some demand is refused.

$= 1 - P(X \leq 2) = 1 - [P(x=0) + P(x=1) + P(x=2)]$

$= 1 - \left[0.22 + e^{-1.5}\left(\frac{1.5}{1} + \frac{(1.5)^2}{2}\right)\right]$

$= 1 - [0.22 + 0.58] = 1 - 0.805$

$= 0.194$

Q If X_1, X_2, X_3 are poisson variants with parameters $m_1 = 1$, $m_2 = 2$, $m_3 = 3$ then find
① $P[(x_1 + x_2 + x_3) \geq 3]$ ② $P[(x_1 = 1)/(x_1 + x_2 + x_3 = 3)]$

We have $X_1 + X_2 + X_3 = x$ as a poisson variant where $m = m_1 + m_2 + m_3 = 1 + 2 + 3 = 6$

$$\therefore P(x = x) = \frac{e^{-m} m^x}{x!} = \frac{e^{-6} 6^x}{x!}$$

① $P[(x_1 + x_2 + x_3) \geq 3] = P[x \geq 3]$

$= 1 - [P(x < 3)] = 1 - [P(x = 0, 1, 2)]$

$= 1 - \left[e^{-6}\left(\frac{6^0}{0!} + \frac{6^1}{1!} + \frac{6^2}{2!}\right)\right] = 1 - \left[e^{-6}(25)\right]$

$= 1 - 0.0619$

$= 0.938$

$= 0.94$

② $P[(x_1 = 1)/(x_1 + x_2 + x_3 = 3)] = \dfrac{P[(x_1 = 1) \cap (x_1 + x_2 + x_3 = 3)]}{P[x_1 + x_2 + x_3 = 3]} = \dfrac{P[(x_1 = 1) \cap (x_2 + x_3 = 2)]}{P[x_1 + x_2 + x_3 = 3]}$

$$= \frac{P(x_1 = 1) \times P(x_2 + x_3 = 2)}{P[x_1 + x_2 + x_3 = 3]} = \frac{\frac{e^{-1} 1}{1!} \times \frac{e^{-5} 5}{2!}}{\frac{e^{-6} 6^3}{3!}}$$

$$= \frac{e^{-1}}{1} \cdot \frac{25 e^{-5}}{2} \cdot \frac{6}{e^{-6} \cdot 216} = \frac{75 e^{-6}}{216 e^{-6}} = \frac{75}{216} = 0.34722$$

Q Insurance company found that 0.01% of the population meet with an accident every year. If 1000 policy holders are selected randomly what is the probability that no more than 2 clients meet with an accident next year.

X = no of clients

$n = 1000$

$p = 0.01\% = 0.01/100$

$m = np = 0.1$

$$P(X = x) = \frac{e^{-0.1} (0.1)^x}{x!}$$

we want $P(x \leq 2) = P(x = 0) + P(x = 1) + P(x = 2)$

$$= e^{-0.1}\left[\frac{(0.1)^0}{0!} + \frac{(0.1)^1}{1!} + \frac{(0.1)^2}{2!}\right] = e^{-0.1}\left[1 + 0.1 + 0.005\right]$$

$= 0.999$

4|2|25

Q Using poissons distribution find the approximate value of $300C_2(0.02)^2(0.98)^{298} + 300C_3(0.02)^3(0.98)^{297}$

We observe that these are binomial proprties →

$n = 300 \quad p = 0.02 \quad q = 0.98$

$m = np = 300 \times 0.07 = 6$

$\therefore P(x=x) = \dfrac{e^{-m} m^x}{x!} = \dfrac{e^{-6} 6^x}{x!}$

$\therefore$ given term $= \dfrac{e^{-6}6^2}{2!} + \dfrac{e^{-6}6^3}{3!}$

$= 0.044617 + 0.089235$

$= 0.13385$

Q In a certain factory producing blades, there's 1/500 chance of a blade being defective. Blades are supplied in a pack of 10. Find approximate no of packets containing ① no defective blade ② 1 defective blade ③ 2 defective blades in a 10,000 packet consignment.

Let X = no of defective blades

$p = 1/500$

$n = 10 \quad N = 10000$

$m = np = 0.02$

$P(X=x) = \dfrac{e^{-m} m^x}{x!} = \dfrac{e^{-0.02} \cdot 0.02^x}{x!}$

$P(X=0) = \dfrac{e^{-0.02} \cdot 0.02^0}{0!} = 0.98$

①

Expected Number of packets with 0 defective blades $= NP(X=0)$

$= 10000 \times 0.98$

$= 9800$

② $= \dfrac{e^{-0.02} \cdot 0.02^1}{1!} = 0.0196$

$NP(X=1) = 10000 \times 0.0196$

$= 196$

③ $= \dfrac{e^{-0.02} \cdot 0.02^2}{2!} = 0.000196$

$NP(X=2) = 10000 \times 0.000196$

$= 1.96$

$\simeq 2$

Q Fit a poissons distribution to the following data.

No of deaths:	0	1	2	3	4	Total
Frequency :	123	59	14	3	1	200

$$m = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{100}{200} = 0.5 \qquad N = 200$$

$$\therefore \; P(X=x) = \frac{e^{-m} \cdot m^x}{x!} = \frac{e^{-0.5} \cdot 0.5^x}{x!}$$

x :	0	1	2	3	4
$P(X=x)$:	0.6065	0.3032	0.07581	0.0126	0.001579
NP :	121.3	60.64	15.162	2.52	0.3158
	121	61	15	2	0

Uniform Distribution (discrete)

Discrete variable X is said to follow a uniform distribution if $\boxed{P(X=x_i) = \frac{1}{n}, \; \forall x_i = 1, 2, \text{---} n}$

$$\boxed{E(x) = \Sigma x_i P_i}$$

$$= \frac{1}{n}[1+2+3\text{----}+n] = \frac{1}{n} \cdot \frac{n \cdot (n+1)}{2}$$

$$\boxed{E(x) = \frac{n+1}{2}}$$

$$\boxed{Var(x) = E(x^2) - (E(x))^2}$$

$$\boxed{Var(x) = \frac{n^2-1}{12}}$$

Uniform Distribution (continous)

A continous random variable X is said to follow uniform distribution if its' PDF is given by $\boxed{f(x) = \begin{cases} k & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}}$

Now
$$\boxed{\begin{aligned} &\int_a^b f(x)\,dx = 1 \\ &\Longrightarrow k(b-a) = 1 \\ &\therefore k = 1/b-a \end{aligned}}$$

$$\boxed{E(x) = \int_a^b x f(x)\,dx} = \frac{1}{b-a}\int_a^b x\,dx = \boxed{\frac{b^2-a^2}{2(b-a)}}$$

$$\boxed{E(x) = \frac{b+a}{2}}$$

$$\boxed{Var(x) = E(x^2) - [E(x)]^2 = \frac{(b-a)^2}{12}}$$

4/2/25

Q Roll a 6 faced dice (fair) X denotes the number on top.
find 1) P(even number) 2) $P(X<3)$ 3) $E(x)$ and $Var(x)$

1) $X \sim U(1,6)$
$P(X=x) = 1/6$
$P(X = \text{even number}) = P(X=2) + P(X=4) + P(X=6)$
$= 1/6 + 1/6 + 1/6$
$= 3/6$
$P(X = \text{even number}) = 1/2$

2) $P(X<3) = P(X=1) + P(X=2)$
$= 1/6 + 1/6$
$= 2/6$
$P(X<3) = 1/3$

3) $E(x) = \frac{n+1}{2} = \frac{7}{2}$

$Var(x) = \frac{n^2-1}{2} = \frac{35}{2}$

Q A telephone number is selected at random X denotes the last digit of the selected no. find probability that the last digit is: find ① $X=6$ ② $X<3$ ③ $X \geq 8$
$X \sim U(0,9)$
$P(X=x) = 1/10$
① $P(X=6) = 1/10$
② $P(X<3) = P(X=0) + P(X=1) + P(X=2)$
$= 1/10 + 1/10 + 1/10$
$P(X<3) = 3/10$

③ $P(X \geq 8) = P(X=8) + P(X=9)$
$= 1/10 + 1/10$
$P(X \geq 8) = 1/5$

Q A man with n keys and wants to open the door with trying the keys independently and 1-by-1 find mean and variance of n of trials that are required to open the door if unsuccessful keys are kept aside.
$E(x) = \frac{n+1}{2}$ $\quad Var(x) = \frac{n^2-1}{2}$

For continuous uniform distribution (CUD),

For any sub-interval of $[c, d] \subset [a, b]$

$$P(c \leq x \leq d) = \int_c^d f(x)\,dx$$

$$\boxed{P(c \leq x \leq d) = \frac{1}{b-a}(d-c)}$$

Q Suppose there are 30 participants in a quiz, question is given to all 30 and time alloted is 25 secs find the probability of participants responding in 6 secs.

$$X - U[\underset{a}{0}, \underset{b}{25}]$$

$$P[X = x] = \frac{1}{25-0} = \frac{1}{25}$$

Interval of successful event is $[0, 6]$

$$\therefore P(x < 6) = \frac{1}{25}(6-0)$$

$$= \frac{6}{25}$$

$\therefore$ Expected no of participants $= \frac{6}{25} \times 30$

$$= \underline{\underline{7}}$$

Q A random number N is taken from 690 to 850 in uniform distribution. find the probability that $N > 790$.

Interval of distribution $= [690, 850]$

Interval of successful event $= [790, 850]$

$$P(\underset{c}{790} < x < \underset{d}{850}) = \frac{850-790}{850-690} = \frac{1}{160}[850-790] = \frac{60}{160} = \frac{3}{8}$$

$$P(N > 750) = \underline{\underline{0.375}}$$

Q A flight is about to land and expected time to land is 30 min. Find the probability of the landing happening between 25 and 30 minutes.

Given interval $= [0, 30]$

$$P(X = x) = \frac{1}{30}$$

$$P(\text{successful event}) = \frac{5}{30} = \frac{1}{6}$$

Q If 'X' is uniformly distributed in -2 to 2. find ① $P(X<1)$ ② $P(|X-1| \geq 1/2)$

$\because x<1 \quad \therefore x \in [-2,1] \subset [-2,2]$

① $P(X<1) = \frac{1}{2-(-2)} [1-(-2)]$

$= \frac{3}{4}$

② if $|x-1| \geq 1/2$

$= \pm (x-1) \geq 1/2$

$x-1 \geq 1/2$ and $1-x \leq 1/2$

$x \geq 3/2$ and $x \leq 1/2$

$x \in [3/2, 2]$ and $x \in [-2, 1/2]$

$P[|x-1| \geq 1/2] = P[3/2 < x < 2] + [-2 \leq x \leq 1/2]$

$= \frac{3}{4}$

Exponential Distribution (continuous)

A continuous random variable X is said to follow exponential distribution with a parameter λ. If $pdf = f(x) = \begin{cases} \lambda . e^{-\lambda x} & x \geq 0 \\ 0 & o/w \end{cases}$

Note: $\int_0^\infty f(x)\,dx = 1$

$E(x) = 1/\lambda$

$var(x) = 1/\lambda^2$

Note: $P(X>k)$

$= \int_k^\infty \lambda e^{-\lambda x} dx = e^{-\lambda k}$

Q The mileage which car owners get with a certain type of tire is a random variable having an exponential distribution with mean 40,000 km. Find the probability ① P(one these tires will last at least 20,000)

② P(at most 30,000)

X = mileage

mean = 40,000

$\frac{1}{\lambda} = 40,000$

$f(x) = \frac{1}{40000} . e^{\frac{1}{40000}x}$

① $P(X \geq 70,000) = \int_{20,000}^{\infty} \frac{1}{40,000} . e^{-\frac{1}{40000}x} dx$

$= e^{-\frac{1}{40000} \times 20000}$

$= e^{-1/2}$

② $P(X \leq 30,000) = \int_0^{30,000} \frac{1}{40,000} e^{-\frac{1}{40000} \times x} dx \quad = 1 - e^{-0.75}$

6/2/25

If the number of kilometers a car can run before its battery wears out is exponentially distributed with an average value of 10,000 km. if the owner decides to take a 5000 km trip. What is the probability that he will be able to complete the trip without replacing the battery.

X = no of km the car can run

$avg = \frac{1}{\lambda} = 10{,}000$, $f(x) = \lambda e^{-\lambda x}$

$$P(x > 5{,}000) = \int_{5000}^{\infty} \frac{1}{10000} e^{-\frac{1}{10000}x} dx$$

$$= e^{-\lambda 5000}$$

$$= e^{-\frac{1}{10000} \times 50000}$$

$$= e^{-1/2}$$

Q if X is exponentially distributed prove that probability that X exceeds its expected value less than 0.75 .

$$P(x > 1/\lambda) < 0.5$$

$$\int_{1/\lambda}^{\infty} \lambda e^{-\lambda x} dx = e^{-\lambda \cdot \frac{1}{\lambda}}$$

$$\Longrightarrow e^{-1}$$

$$\frac{1}{e} < 0.5$$

Q The amount of times that a watch will run without reset is a random variable having an exponential distribution with mean 120 days. Find the probability that ① watch has to be reset in less than 24 days ② will not have to reset in 180 days.

X = amount of time

① $P(X < 24)$

$$= 1 - P(x \geq 24)$$

$$= 1 - \int_{24}^{\infty} \lambda e^{-\lambda x} dx$$

$$= 1 - e^{-1/120 \times 24}$$

$$= 0.181$$

② $P(x \geq 180)$

$$= \int_{180}^{\infty} f(x) dx$$

$$= e^{-1/120 \times 180}$$

$$= 0.2231$$

Memory less property

X is exponentially distributed, then $P(x > s+t \,/\, x > s) = P(x>t)$ ↗ given that

Q The length of the shower during monsoon has an exponential distribution with parameter 2. Find the ① probability that the shower will last more than 3 minutes ② if it has already lasted for 2 minutes what is the probability that it will last for atleast 1 more minute.

① $P(X>3) = \int_3^{\infty} 2e^{-2x}\,dx$

$= e^{-2\times 3}$

$= e^{-6}$

② $P(X > 2+1 / x > 2)$

$= P(X>1) = e^{-2\times 1}$

$= e^{-2}$

Normal Distribution (continous)

A continous random variable X is said to follow a normal distribution with parameter 'm' (mean) and 'σ' (standard deviation). if its pdf is given by

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}, \quad -\infty < x < \infty}$$

Property: If $x_1, x_2, x_3 --- x_n$ are normal variants with parameters $m_1, m_2, m_3 -- m_n$ and s.d. $\sigma_1, \sigma_2, \sigma_3 -- \sigma_n$. Then,

$y = a_1x_1 + a_2x_2 + a_3x_3 --- a_nx_n$ is also a normal variant with parameter $m = a_1m_1 + a_2m_2 + \ldots a_nm_n$ and s.d. $\sigma^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 -- a_n^2\sigma_n^2$

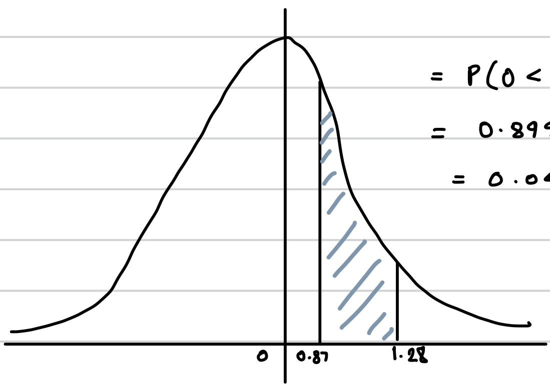
If $X \sim N(m, \sigma)$, then $z = \frac{X-m}{\sigma}$ is called standard normal variant (SNV) and $z \sim N(0,1)$

i.e. with mean = 0 and s.d. = 1.

Bell-shaped curves that will be symmetric about m (0 in standardisation)

Q Find probability of $(0.87 < z < 1.28)$

z so mean = 0.



$= P(0 < z < 1.28) - P(0 < z < 0.87)$

$= 0.8997 - 0.8078$

$= 0.0419$

10/2/25

Q Find the value of c such that

① $P(-c < z < c) = 0.95$

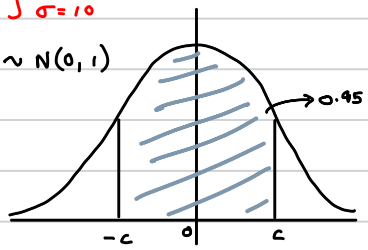
② $P(|z| < c) = 0.01$

③ $P(x > c) = 0.02$ } $m = 120$

④ $P(x < c) = 0.05$ } $\sigma = 10$

(i) $z = \dfrac{x - m}{\sigma} \sim N(0,1)$

$[x \sim N(m, \sigma)]$



we have,

$P(0 < z < c) = \dfrac{0.95}{2} = 0.475$

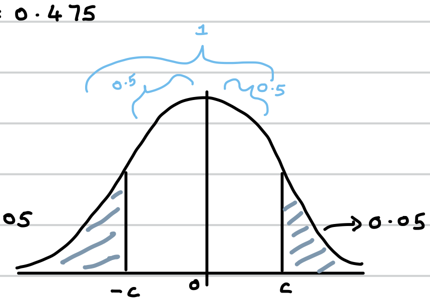
$\therefore c = 1.9 + 0.06 = 1.96$

So that $P(0 < z < c) = 0.475$

② $P(|z| > c) = 0.01$

$P(z > c \;\&\; z < -c)$

$P(z > c) = \dfrac{0.01}{2} = 0.005$



$P(0 < z < c) = 0.5 - 0.005$

$= 0.495$

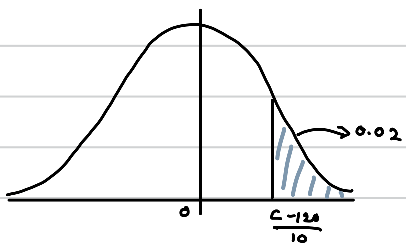
$\therefore c = 2.58$

So that $P(0 < z < c) = 0.495$

3 $P(x > c) = 0.02$

$m = 120 \quad \sigma = 10$

$x \to z \qquad z = \dfrac{x - m}{\sigma} = \dfrac{x - 120}{10}$



$P\left(\dfrac{x-120}{10} > \dfrac{c-120}{10}\right) = P\left(z > \dfrac{c-120}{10}\right) = 0.02$

$\therefore P\left(0 < z < \dfrac{c-120}{10}\right) = 0.5 - 0.02 = 0.48$

$\therefore \dfrac{c-120}{10} = 2.06$

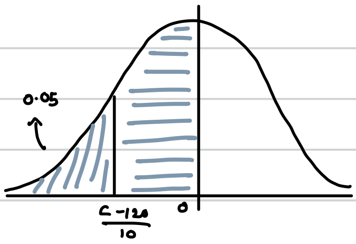
$c = 2.06$

$\therefore c = 10 \times 2.06 + 120 = 140.6$

④ $P(x < c) = 0.05$

$m = 120 \qquad \sigma = 10$

$P\left(z < \frac{c-120}{10}\right) = P\left(0 < z < \frac{c-120}{10}\right)$



$= 0.5 - 0.05$

$= 0.45$

$\therefore \frac{c-120}{10} = -1.65$ ↳ because on the other end

$c = 103.5$

17/2/25 ISE portion: Module 1 & 2 5 out of 8 q's

Q Monthly salary X is normally distributed between with mean = 3000 Rs and σ = 250 Rs. What should the minimum salary of the worker be so that he belongs to top 5% workers.

SNV $\Rightarrow z = \frac{X-m}{\sigma}$

Let X_1 be minimum salary of top 5% workers

$\therefore$ The monthly salary of the worker should be greater than X_1 so that the worker belongs to the top 5% workers.

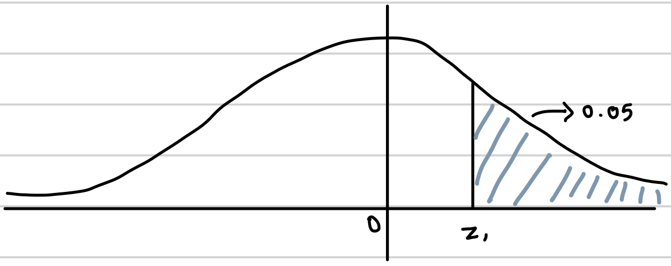
We want to find z_1 such that probability of $z > z_1$ = 5% = 0.05

↳ $\frac{x_1 - m}{\sigma}$

$\Rightarrow P(0 < z < z_1) = 0.5 - P(z > z_1)$

$= 0.5 - 0.05$

$P(0 < z < z_1) = 0.45$



$\therefore z_1 = 1.65$

i.e. $\frac{X_1 - 3000}{250} = 1.65$

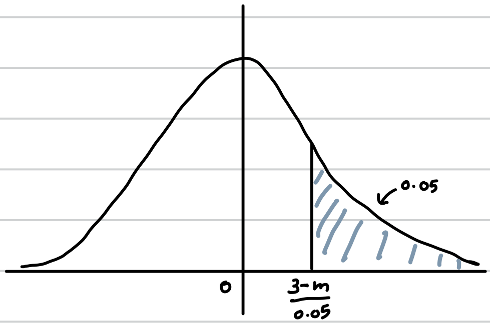
$X_1 = \underline{\underline{3412}}$

Q The diameter of the tops of the cans are normally distributed with s.d. 0.05 cm at what mean should the diameter be set so that not more than 5% of the tops can be produced which will have diameter exceeding 3 cm.

Let X be diameter of the top

$P(X > 3) = 0.05$ (5%)

$P\left(\frac{x-m}{\sigma} > \frac{3-m}{\sigma}\right) = 0.05$



$P\left(\frac{x-m}{0.05} > \frac{3-m}{0.05}\right) = 0.05 \implies P\left(z > \frac{3-m}{0.05}\right) = 0.05$

$\therefore P\left(0 < z < \frac{3-m}{0.05}\right) = 0.5 - 0.05$
$= 0.45$

$\Rightarrow \frac{3-m}{0.05} = 1.65$

$m = 2.9175$

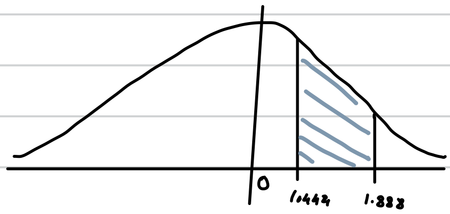
9 If X_1 and X_2 are two normal variants with mean 30 and 25 and variance 16 and 12. $y = 3x_1 - 2x_2$. Find probability of y between 60 and 80.

Let m be mean of y and σ be s.d of y.

$\therefore m = 3 \times 30 + (-2)(25) = 40$

$\sigma^2 = 9 \times 16 + 4 \times 12 = 192$

$P(60 \leq y \leq 80) = P\left(\frac{60-m}{\sigma^2} \leq y \leq \frac{80-m}{\sigma^2}\right)$



$P\left(\frac{60-40}{13.85} \leq y \leq \frac{80-40}{13.85}\right) = P(1.444 < z < 2.888)$

$= P(0 < z < 2.888) - P(0 < z < 1.444)$

$= 0.4980 - 0.4251 = 0.0729$

The probability of y between 60 and 80 is $\underline{0.0729}$.

$m = a_1 m_1 + a_2 m_2$

$\sigma^2 = a_1^2 \sigma_1 + a_2^2 \sigma_2$

20/2/25 9 Random variable x and Y are normally distributed with mean and standard deviation as (52, 3) and (50, 2). Find the probability that randomly chosen pair X and Y differ by 1.7 or more.

$(\underset{m_1}{52}, \underset{\sigma_1}{3})$ $(\underset{m_2}{50}, \underset{\sigma_2}{2})$

$P(|x-y| \geq 1.7)$

Let $X - Y = u$

$m =$ mean of $u = 2$ $[X - Y$ so $52 - 50]$

$\sigma^2 =$ var of $u = 9 + 4$

$\therefore \sigma = \sqrt{13}$

$P(|U| \geq 1.7) = 1 - P(|u| < 1.7)$

$= 1 - P(-1.7 < u < 1.7)$

$= 1 - P\left(\frac{-1.7-2}{\sqrt{13}} < z < \frac{1.7-2}{\sqrt{13}}\right)$

$|a| < \delta$

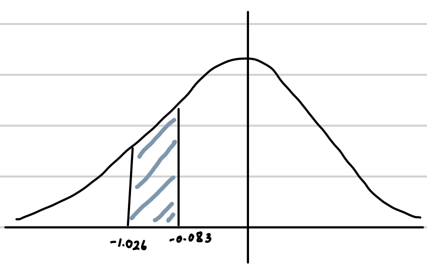
$-\delta < |a| < \delta$

$= 1 - P(-1.026 < z < -0.083)$

$= 1 - [P(0 < z < 1.026) - P(0 < z < 0.083)]$

$= 1 - [0.3461 - 0.0319] = 1 - 0.3142$

$= \underline{0.6858}$



Q if X and Y are normal variants with $m_1 = 8$ $m_2 = 12$ $\sigma_1 = 2$ and $\sigma_2 = 4\sqrt{3}$ find the value of α such that $P[(2X - Y) \le 2\alpha] = P[(X + 2Y) \ge 3\alpha]$

Let $2X - Y = U$

$X + 2Y = V$

$m_u = 16 - 12 = 4$ $\qquad m_v = 8 + 24 = 32$

$\sigma_u^2 = 16 + 48 = 64$ $\qquad \sigma_v^2 = 4 + 192 = 196$

$\sigma_u = 8$ $\qquad \sigma_v = 14$

Given $P(U \le 2\alpha) = P(V \ge 3\alpha)$

$P\left(z \le \frac{2\alpha - 4}{2}\right) = P\left(z \ge \frac{3\alpha - 32}{14}\right)$

$\frac{2\alpha - 4}{8} = \frac{3\alpha - 32}{14}$

$\alpha = 6$

Q Income group of 10,000 people is normally distributed with mean 520 and s.d. 60 find
1. No of people with income 400 and 550.
2. Lowest income of richest 500

1. $P(400 < x < 550)$

$P\left(\frac{400 - 520}{60} < z < \frac{550 - 520}{60}\right)$

$= P(-2 < z < 0.5) = P(0 < z < 0.5) + P(0 < z < 2)$

$= 0.1915 + 0.4772$

$= 0.6687$

No of people $= 10000 \times 0.6682$

$= 6687$

2. Let X_2 = Lowest income of richest 500

$P(X \ge X_2) = \frac{500}{10000} = 0.05$

$P\left(z \ge \frac{X_2 - 520}{60}\right) = 0.05$

$P\left(0 < z < \frac{X_2 - 520}{60}\right) = 0.5 - 0.05$

$= 0.45$

$\frac{X_2 - 510}{60} = 1.65$

$X_2 = 619$

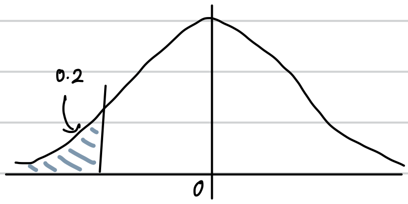
In a competitive exam, top 15% of students will get grade A while bottom 20% will be declared failed. Grades are normally distributed with mean 75 and sd = 10. Determine the lowest percentage of marks to recieve grade A and to pass resp.

$P(X \geq x_1) = 15\% = 0.15$

$P\left(z \geq \frac{x_1 - 75}{10}\right) = 0.5 - 0.15 = 0.35$

$\therefore \frac{x_1 - 75}{10} = 1.04 \implies x_1 = 85\%$

Let x_2 be lowest % marks that passes



$P(X < x_2) = 20\% = 0.2$

$P\left(z < \frac{x_2 - 75}{10}\right) = 0.2$

$P\left(0 < z < \frac{x_2 - 75}{10}\right) = 0.5 - 0.2 = 0.3$

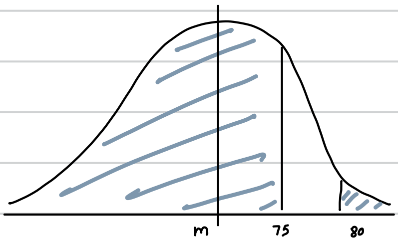
$\frac{x_2 - 75}{10} = -0.85$

$x_2 - 75 = -8.5$

$x_2 = -8.5 + 75$

$x_2 = 66.6$

21|2|25 Q Find the mean and standard deviation of the scores where 58% of the candidates obtained marks below 75. 4% got above 80 and rest between 75 and 80.



$P\left(0 < z < \frac{75 - m}{\sigma}\right) = 0.08$

$\frac{75 - m}{\sigma} = 0.2$

$m + 0.2\sigma = 75$ (1)

$P(m < x < 75) = 58 - 50 = 8\% = 0.08$

$P(x > 80) = 4\% = 0.04$

$P\left(z > \frac{80-m}{\sigma}\right) = 0.04$

$\therefore P\left(0 < z < \frac{80-m}{\sigma}\right) = 0.5 - 0.04$

$= 0.46$

$\therefore \frac{80-m}{\sigma} = 1.76$

$\therefore m + 1.76\sigma = 80$ ②

from ① and ②

$m = 74.35 \quad \sigma = 3.225$

Q Probability that an electronic component will fail in less than 1200 hours of use is 0.25.

$p = 0.25 \quad q = 0.75$

Using normal approximation to the binomial distribution find probability that amongst 200 such components fewer than 45 will fail in less than 1200 hours of continous use.

Let X = no of components that fail < 1200 hrs

Add 0.5 if it is lesser than when using normal approximation to binomial distribution

$p = 0.25 \quad q = 0.75 \quad n = 200$

$m = np = 50 \quad \sigma^2 = npq = 37.5$

$\sigma = 6.123$

Required probability $= P(x < 45)$

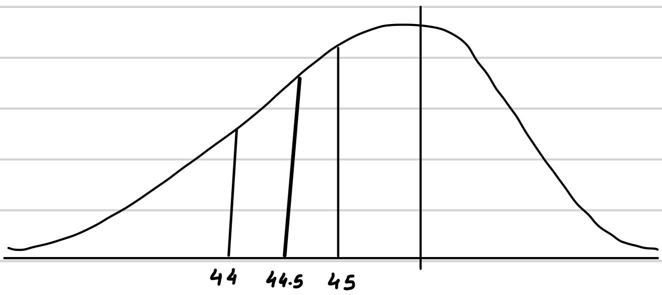
$= P(x \leq 44.5)$

$= P\left(z \leq \frac{44.5-50}{6.123}\right) = P(z \leq -0.898)$

$= 0.5 - P(0 < z < 0.898)$

$= 0.5 - 0.3159$

$= 0.1841$



27/2/25 1.3 not in ISE

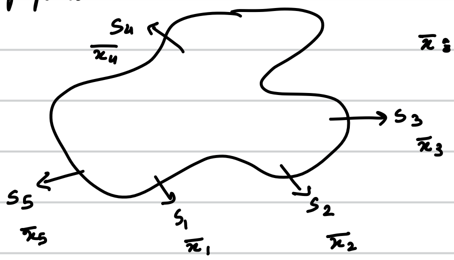
10/3/25

Module 3:

Sampling

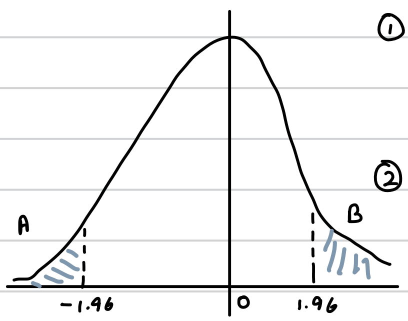
Large sampling: $(n \geq 30)$: if $\bar{x}$ is a mean of each sample of size n drawn from the population whose mean is μ and s.d. σ then $\bar{x}$ is normally distributed with mean μ and s.d. $\sigma/\sqrt{n}$. $\bar{x} \sim N(\mu, \sigma/\sqrt{n})$

population:



$\bar{x} : \bar{x}_1, \bar{x}_2, \text{---} \bar{x}_n$

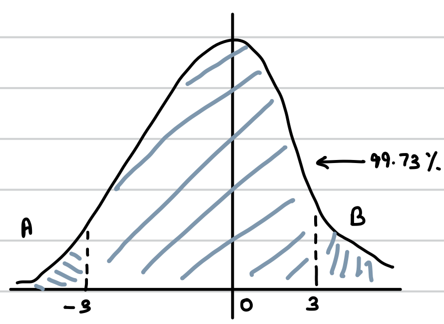
Therefore the SNV is $\boxed{z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{\bar{x} - \mu(\sqrt{n})}{\sigma}}$



① $P(z > 1.96, z < -1.96) = 0.05$
$= P(|z| > 1.96)$

② Similarly $P(|z| > 2.58) = 0.01$

③ $P(|z| > 3) = 0.0027$



The levels marked by the probabilities 0.05 and 0.01 which decide the significance of an event are called level of significance (LOS). Normally expressed in terms of percentage (5% and 1% LOS).

Confidence limits: The limits within which we expect z to lie with specified probabilities are called confidence limits. Thus for $P(|z| > 1.96) = 0.05$, ± 1.96 are confidence limits. This means we are confident that (95%) in 95 out of 100 cases of the sample mean $\bar{x}$ lies between ± 1.96.

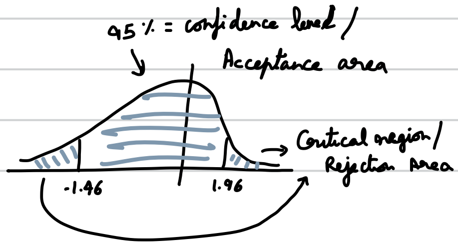
Critical region: The regions beyond the confidence limits/levels are called critical regions.

11/3/25

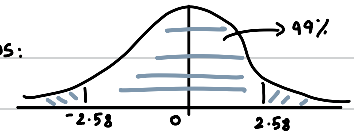
$\bar{x} \sim N(\mu, \sigma/\sqrt{n})$

then $z = \dfrac{\bar{x} - \mu}{(\sigma/\sqrt{n})}$

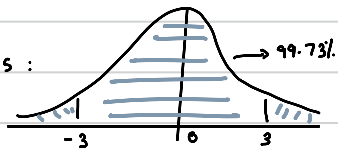
① at 5% LOS:



② at 1% LOS:



③ at 0.27% LOS :



Interval estimation : We know that at 5% LOS

$$|z| < 1.96$$

$$\left| \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right| < 1.96$$

$$|x - a| < \delta$$

$$x \in (a-\delta, a+\delta)$$

i.e. $|\bar{x} - \mu| < 1.96 \cdot \sigma/\sqrt{n}$

$|\mu - \bar{x}| < 1.96 \cdot \sigma/\sqrt{n}$

$$\therefore \mu \in (\bar{x} - 1.96 * \sigma/\sqrt{n}, \bar{x} + 1.96 * \sigma/\sqrt{n})$$

or

$$\mu \in (\bar{x} - 1.96 * s/\sqrt{n}, \bar{x} + 1.96 * s/\sqrt{n})$$

Example : Measurements of weights of random sampling of 200 ball bearings showed a mean of 0.824 N and s.d. of 0.042 N. find 95% confidence limits for the mean weight of all bearings.

$n = 200$

$\bar{x} = 0.824$

$s = 0.042$ N

at 5% LOS

$$\mu \in (\bar{x} - 1.96 * s/\sqrt{n}, \bar{x} + 1.96 * s/\sqrt{n})$$

$$\mu = (0.824 - 1.96 * 0.042/\sqrt{200}, 0.824 + 1.96 * 0.042/\sqrt{200})$$

$$\mu \in (0.8235, 0.8352)$$

Q A random sample of size 65 in the process of estimating mean annual income of 950 families, mean = 4730 s.d. = 765. find 95% confidence interval for the population mean.

$n = 65$

$N = 950$

$\bar{x} = 4730$

$s = 765$

$$\frac{n}{N} = 0.068 \,(> 0.05)$$

then $$\hat{\sigma}_{\bar{x}} = \sqrt{\frac{N-n}{N-1}} \times \frac{s}{}$$

$$\mu \in \left(\bar{x} - 1.96 * \sqrt{\frac{N-n}{N-1}} * \frac{s}{\sqrt{n}} \,,\; \bar{x} + 1.96 * \sqrt{\frac{N-n}{N-1}} * \frac{s}{\sqrt{n}}\right)$$

$$\mu \in (4628 \,,\, 4831)$$

Testing of Hypothesis

Q. A random sample of 50 items gives a mean of 6.2 and s.d. of $\sqrt{10.24}$. Can the sample (it) be regarded as drawn from from normal population with mean 5.4 at 5% LOS

$n = 50$

$\bar{x} = 6.2$

$s = \sqrt{10.24}$

$\mu = 5.4$

Let null hypothesis $H_0 : \mu = 5.4$

alternate hypothesis $H_a : \mu \neq 5.4$

(Two tailed test if equated)

$$Z_{cal} = \left| \frac{\bar{x} - \mu}{s/\sqrt{n}} \right|$$ always in mod if two tailed

$$= 1.767$$

at 5% LOS, $Z_\alpha = 1.96$

Since $Z_{cal} < Z_\alpha$, it lies in the acceptance area.

Therefore the null hypothesis is accepted.

$\mu = 5.4$.

Therefore the sample is drawn from normal population with $\mu = 5.4$.

13|3|25

Q A random sample of 400 members has mean 4.45 cm can it be regarded as a sample from the population, with mean 5 cm and variance 4cm.

$n = 400$

$\bar{x} = 4.45$

$\mu = 5$

$\sigma^2 = 4$

$H_0 : \mu = 5$

$H_a : \mu \neq 5$

$$Z_{cal} = \frac{|\bar{x} - \mu|}{\sigma/\sqrt{n}} = \frac{0.55}{2/20} = 5.5$$

at 5% LOS, $Z_\alpha = 1.96$,

$\therefore Z_{cal} > Z_\alpha$

The sample is not drawn from population whose mean is 5.

Two Population :

Let $\bar{x}_1$ and $\bar{x}_2$ be the samples taken from population 1 and 2 respectively whose mean is μ_1 and μ_2 and standard deviation is σ_1 and σ_2.

Then $\bar{x}_1 - \bar{x}_2$ follows a normal distribution with mean $\mu_1 - \mu_2$ and s.d. $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

Difference of mean

$$\boxed{\begin{gathered} \bar{x}_1 - \bar{x}_2 \sim N\left(\mu_1 - \mu_2, \underbrace{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}_{\text{standard error S.E.}}\right) \\ \text{then } z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S.E.} \end{gathered}}$$

① If $\sigma_1 = \sigma_2 = \sigma$ (known)

then $S.E. = \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

② If $\sigma_1 = \sigma_2 = \sigma$ (unknown)

then $S.E. = \sqrt{\frac{s_1^2}{n_2} + \frac{s_2^2}{n_1}}$ where s_1 and s_2 are called unbiased estimates of population standard deviation

③ If $\sigma_1 \neq \sigma_2$ (unknown)

then $S.E. = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Interval Estimation

$$\boxed{\mu_1 - \mu_2 \in \left((\bar{x}_1 - \bar{x}_2) - 1.96\ S.E.,\ (\bar{x}_1 - \bar{x}_2) + 1.96\ S.E.\right)}$$

95%

Q Find 95% confidence limit for the difference between population mean from given data.

	Size	Mean	S.D.
Sample 1	400 (n_1)	124 ($\bar{x}_1$)	14 (s_1)
Sample 2	250 (n_2)	120 ($\bar{x}_2$)	12 (s_2)

$$S.E. = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{196}{400} + \frac{144}{250}} = \sqrt{\frac{49}{100} + \frac{72}{125}} = \sqrt{1.066} = \underline{1.03247}$$

$$\mu_1 - \mu_2 \in (1.976, 6.023)$$

Q Means of 2 samples of sizes 1000 and 2000 are respectively 67.50 and 68 can the samples be regarded as drawn from the population (same) with s.d. 2.5

$n_1 = 1000$

$n_2 = 2000$

$\bar{x}_1 = 67.50$

$\bar{x}_2 = 68.00$

$\sigma = 2.5$

Let $H_0 : \mu_1 = \mu_2$

$H_a : \mu_1 \neq \mu_2$ (Implies two tailed test)

(mod cause two tailed test)

$$Z_{cal} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S.E.}$$

$\sigma_1 = \sigma_2 = \sigma$

$$S.E. = \sigma \cdot \sqrt{\frac{1}{n_2} + \frac{1}{n_1}} = 2.5 \sqrt{\frac{1}{1000} + \frac{1}{2000}} = 0.096$$

$$\therefore Z_{cal} = \frac{|(67.50 - 68.00) - (0)|}{0.096} = 5.163$$

because H_0 says they are equal

at 0.27% LOS

because closer to Zcal

smallest rejected value if Zcal is high

higher confidence level to be taken

check 5% then 1% then 0.27%

$z_\alpha = 3$

$\because$ $z_{cal} > z_\alpha$, we reject H_0.

$\therefore$ The samples are drawn from different populations.

Q Average of marks scored by 36 boys is 72 with s.d. 8, while that of 36 girls is 70 with s.d. 6. Using one tail test at 1% LOS, test whether boys perform better than girls.

↖ always if better, greater, lesser than.

$n_1 = 36$ $\quad$ $n_2 = 36$

$\bar{x}_1 = 72$ $\quad$ $\bar{x}_2 = 70$

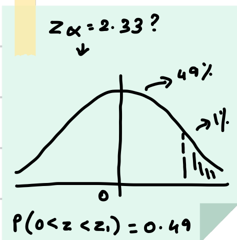
$s_1 = 8$ $\quad$ $s_2 = 6$

Let $H_0 : \mu_1 = \mu_2 \longrightarrow$ always equated

$H_a : \mu_1 > \mu_2 \longrightarrow$ if $1 > 2$ then right tailed +ve, $2 < 1$ then left tailed -ve.

$$z_{cal} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S.E.} \longrightarrow \text{no mod for one-tailed}$$

$$S.E. = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 1.66$$



$z_{cal} = 1.2$

Now at 1% LOS for one-tailed test

$z_\alpha = 2.33$

$\because$ $z_{cal} < z_\alpha$, accept H_0

$\therefore$ Boys do not perform better than girls

if by Pvalue

$P(z > 1.2) = 0.1151$

at 1% LOS $\alpha = 0.01$

$\Big\}$ $P > \alpha$

17/3/25

Small sampling

If we take large number of samples of size less than 30 then we use student's t-test distribution.

Sample in x	σ-known	σ-unknown
$n \geq 30$	z-test	z-test
$n < 30$	z-test	t-test

t - stat is defined as

$$\boxed{t_{cal} = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad \text{where } s \text{ is } \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}}}$$

called unbiased estimate of population standard deviation

OR

$$\boxed{t_{cal} = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} \quad \text{where } s = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n}}}$$

sample standard deviation

Degree of Freedom (dof)

It is the number of values that are free to use

eg: Interval estimation for population mean. Sample for size 10 has mean 40 and s.d. 10. Construct 99% confidence interval for the population mean.

$n = 10 \quad \bar{x} = 40 \quad s = 10$

$$\mu \in (\bar{x} - t_\alpha \hat{\sigma}_{\bar{x}} \ , \ \bar{x} - t_\alpha \hat{\sigma}_{\bar{x}})$$

where $\hat{\sigma}_{\bar{x}} = \dfrac{s}{\sqrt{n-1}} = 3.33$

at 1% LOS and $10 - 1 = 9$ degree of freedom

$t_\alpha = 3.25$ from the table

$\therefore \mu \in (40 - 3.25 * 3.33, \ 40 + 3.25 * 3.33)$

$\mu \in (29.17, 50.83)$

Testing of hypothesis

Q A soap manufacturing company distributes soaps through large numbers before the campaign, per week per shop was 140 dozen. After the campaign a sample of x shops was taken and then the mean sale was found to be 147 dozen with s.d. 16. Can you consider the advertisement effective?

$\mu = 140$

$n = 26$

$\bar{x} = 147$ (sample's mean)

$s = 16$

Let $H_0 : \mu = 140$

$H_a : \mu > 140$

$$t_{cal} = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{147 - 140}{16/\sqrt{25}} = \frac{7}{16/5} = 2.187$$

at 5% LOS and 26 - 1 = 25 df

$t_\alpha = 1.708$

$\therefore t_{cal} > t_\alpha$

H_a is accepted, rejecting H_0

Hence the advertisement was effective

Note:
t - test two tailed table will be given, learn 1-tail in terms of 2-tail

Q Consider the following values 45, 47, 50, 52, 48, 47, 49, 53, 51. Does the mean of these 9 items differ significantly from the assumed population mean 47.5.

$H_0 : \mu = 47.5$

$H_a : \mu \neq 47.5$

$n = 9$

$\bar{x} = \dfrac{\Sigma x}{n} = 49.11$

$s = \sqrt{\dfrac{\Sigma (x_i - \bar{x})^2}{n}} = \sqrt{\dfrac{54.888}{9}} = 2.47$

x_i	$(x_i - \bar{x})^2$
45	16.8921
47	4.4521
50	0.7921
52	8.3521
48	1.2321
47	4.4521
49	0.0121
53	15.1321
51	3.5721
	54.8889

$\therefore t_{cal} = \left| \dfrac{\bar{x} - \mu}{s/\sqrt{n-1}} \right| = 1.84$

at 5% LOS and 8 df

$t_\alpha = 2.306$

$\because t_{cal} < t_\alpha$

H_0 is accepted, $\mu = 47.5$

18|3|25 Two samples:

Testing Difference between means

① Independent Samples:

If sample size $(n_1 + n_2 - 2) < 30$ then unbiased estimate of common population standard deviation is given by:

$$\boxed{S_p = \sqrt{\frac{\Sigma (x_{i_1} - \bar{x}_1)^2 + \Sigma (x_{i_2} - \bar{x}_2)^2}{n_1 + n_2 - 2}}}$$

if $s_1 = \sqrt{\dfrac{\Sigma (x_{i_1} - \bar{x}_1)^2}{n_1 - 1}}$ & $s_2 = \sqrt{\dfrac{\Sigma (x_{i_2} - \bar{x}_2)^2}{n_2 - 1}}$ then: $\boxed{S_p = \sqrt{\dfrac{s_1^2 (n_1 - 1) + s_2^2 (n_2 - 1)}{n_1 + n_2 - 2}}}$

Now if the standard deviation of samples are

$$S_1 = \sqrt{\frac{\Sigma(x_{i_1} - \bar{x}_1)^2}{n_1}} \quad \& \quad S_2 = \sqrt{\frac{\Sigma(x_{i_2} - \bar{x}_2)^2}{n_2}}$$

then $$\boxed{S_p = \sqrt{\frac{S_1^2 n_1 + S_2^2 n_2}{n_1 + n_2 - 2}}}$$

then Standard Error $$\boxed{(S.E.) = S_p * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

and

$$\boxed{t_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{S.E.}}$$

Q A sample of 8 students of 16 years shows a mean systolic pressure 118.4 mm Hg with $S_1 = 12.17$ mm Hg while a sample of 10 students of 17 years shows a mean systolic pressure of 121 mm Hg with $S_2 = 12.88$ mm Hg. Investigator thinks that the blood pressure is related to age. Do you think the data provides enough reasons to support investigators thought at 5% LOS.

$n_1 = 8 \qquad n_2 = 10$

$\bar{x}_1 = 118.4$ mmHg $\qquad \bar{x}_2 = 121$ mm Hg

$S_1 = 12.17$ mm Hg $\qquad S_2 = 12.88$ Hg

$H_0 : \mu_1 = \mu_2$

$H_a : \mu_1 \neq \mu_2$

$$S_p = \sqrt{\frac{S_1^2 n_1 + S_2^2 n_2}{n_1 + n_2 - 2}} = \sqrt{\frac{2843.8152}{16}} = 13.331$$

$$S.E. = S_p * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 13.331 * \sqrt{0.225} = 6.323$$

$$t_{cal} = \left|\frac{\bar{x}_1 - \bar{x}_2}{S.E.}\right| = \left|\frac{-2.6}{6.323}\right| = 0.4119$$

at 5% LOS and $n_1 + n_2 - 2 = 16$ dof

$t_\alpha = 2.12$

$t_{cal} < t_\alpha$

We accept H_0, we reject H_a that blood pressure does not depend upon age.

Hence the data doesn't provide enough support for the investigator's thought

Q 6 pigs injected with 0.5 mg of medication took on an average 15.4 secs to fall asleep with an unbiased s.d 2.2 secs while 6 other pigs injected with 1.5mg of medication took an average 11.2 secs to fall asleep with an unbiased s.d 2.6 secs. Use 5% LOS to test that the difference in the doses have no effect.

$n_1 = 6 \quad n_2 = 6$

$\bar{x}_1 = 15.4 \quad \bar{x}_2 = 11.2$

$s_1 = 2.2 \quad s_2 = 2.6$

$H_0 : \mu_1 = \mu_2$

$H_a : \mu_1 \neq \mu_2$

$$S_p = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{24.2 + 33.8}{10}} = 2.408$$

$$S.E. = S_p * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 2.408 \times \sqrt{\frac{1}{3}} = 1.390$$

$$t_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{S.E.} = \frac{4.2}{1.390} = 3.021$$

at 5% LOS and 10 dof

$t_\alpha = 2.228$

$t_{cal} > t_\alpha$

We reject H_0, We accept H_a

Dose change does have an effect on the pig's sleeping time.

20/3/25 Q Heights of 6 sailors in inches are 63, 65, 68, 69, 71, 72 and heights of 10 soldiers are 61, 62, 65, 66, 63, 69, 70, 71, 72, 73 Discuss whether soldiers are taller on an average than sailors.

$n_1 = 6 \quad n_2 = 10$

$n_1 + n_2 - 2 = 14 < 30$

$\bar{x}_1 = 68$

$\bar{x}_2 = 67.8$

$$S_p = \sqrt{\frac{\Sigma(x_1 - \bar{x}_1)^2 + \Sigma(x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2}} \qquad S.E. = S_p * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$H_0 : \mu_1 = \mu_2$

$H_a : \mu_2 > \mu_1$

$(x_1 - \bar{x}_1)^2$	$(x_2 - \bar{x}_2)^2$
25	46.24
9	33.64
0	7.84
1	3.24
9	1.44
16	1.44
$\Sigma =$ 60	4.84
	10.24
	17.64
	27.04
	Σ 153.6

$$S_p = \sqrt{\frac{60 + 153.6}{14}} = \sqrt{15.257}$$

$$= 3.906$$

$$S.E. = 3.906 \times \sqrt{\frac{1}{10} + \frac{1}{6}} = 2.014$$

$$t_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{S.E.} = 0.09930$$

at 5% LOS and 14 df

$t_\alpha = 1.761$

$\because$ $t_{cal} < t_\alpha$

We accept H_0, soldiers are not taller than sailors.

② Not independent Samples

In this case we assume, $H_0 : \mu_1 - \mu_2 = \mu = 0$ and test the effectiveness of a particular process on the same sample

Q A certain injection given to 12 patients resulted in the following change in the blood pressure: (bp):
5, 2, 8, -1, 3, 0, 6, -2, 1, 5, 0, 4 can it be concluded that the injection in general will be accompanied by an increase in bp.

$H_0 : \mu = 0$

$H_a : \mu \neq 0 \Rightarrow \mu > 0$ (increase in bp)

$\bar{x} =$ (of diff) $= 2.583$

$$s = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n}} \qquad t_{cal} = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$$

$(x_i - \bar{x})^2$
5.840
0.339
29.349
12.83
0.173
6.671
11.675
21.003
2.505
5.841
6.671
2.007
$\Sigma =$ 104.898

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n}} = \sqrt{\frac{104.898}{12}} = 2.95$$

$$t_{cal} = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{2.583 - 0}{2.95/\sqrt{11}} = 2.9$$

at 5% LOS and 11 dof $(n-1)$

$t_\alpha = 1.796$

$\because t_{cal} > t_\alpha$

$\therefore$ We reject H_0.

This rejection is accompanied by a change in the bp.

Q 10 boys gave a test in statistics and the scores were recorded. They were given a month's coaching and a 2nd test was given. Test if the marks given below provide evidence to the fact that students benefited due to coaching.

Test 1 marks: 70, 68, 56, 75, 80, 90, 68, 75, 56, 58

Test 2 marks: 68, 70, 52, 73, 75, 78, 80, 92, 54, 55

$H_a: \mu > 0$
$(\mu_2 - \mu_1) > 0$
$\mu_2 > \mu_1$
$\Rightarrow$ Right tail test
{Here $t_2 - t_1$}
+ve t_{cal}

$H_a: \mu > 0$
$(\mu_1 - \mu_2) > 0$
$\mu_1 > \mu_2$
$\Rightarrow$ Left tail test
{Here: $t_1 - t_2$}
-ve t_{cal}
-ve sign added to t_α

$T_2 - T_1$	$(x_i - \bar{x})^2$
-2	4.41
2	3.61
-4	16.81
-2	4.41
-5	26.01
-12	146.41
12	141.61
17	285.61
-2	4.41
-3	4.61
10	$\Sigma = 650.1$

$\bar{x} = 0.1$

$H_0: \mu = 0$

$H_a: \mu_2 > \mu_1$

$$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}} = \sqrt{\frac{650.1}{10}} = 8.018$$

$$t_{calc} = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$$

$$= \frac{0.1 - 0}{8.018/\sqrt{9}}$$

$$= \frac{0.1}{8.018/3} = 0.037$$

at 5% LOS and 9 df

$t_\alpha = 1.833$

$\therefore t_{calc} < t_\alpha$

$\therefore$ We accept H_0

21/3/25

χ^2 (chi^2) distribution

→ If random variables $x_1, x_2, x_3, \ldots x_n$ follows normal distribution then $x_1^2 + x_2^2 + \ldots x_n^2$ will follow a distribution which is called χ^2(chi) distribution with n-degree of data.

Usage

1. It is used to test whether there is an association between two or more attributes.
It is also used to test if a characteristics is dependent upon another characteristic.

2. Using χ^2 distribution in this way to test the dependencies of one attributes or another is called test of independence.

3. χ^2 test is commonly known as χ^2 test of goodness of fit, it enables to a certain, how well the theoretical distribution like binomial, poisson's or normal fit the observed frequencies.

4. χ^2 is also used to test whether the proportions: P_1, P_2, P_3, P_4 in different population are equal.

Conditions for χ^2 test:

1. The total number of observations must be sufficiently large. i.e. $n > 50$
2. Frequency of every cell must be greater than 5. If frequency is less than or equal to 5, then combine that frequency with neighboring frequency so that it is greater than 5 so that degree of freedom is reduced accordingly.

Yates correction

In a 2×2 table, degree of freedom is $(row-1) \times (col-1) = (2-1)(2-1) = 1$
If any of the frequency cell is less than 5, we have to use pooling method (combining, 2nd condition)
It will result in χ^2 with 0 degree of freedom, which is meaningless.
Therefore we use yate's correction in this case.

Formula:

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

non-parametric test
O = observed frequency
E = expected frequency

Q Investigate the association between the darkness of eye color in father and son from the following data:

		Father		
		Dark	Not Dark	Total
Son	Dark	48 (x)	90	138
	Not Dark	80	782	862
	Total	128	872	1000

Let H_0 : no association between the darkness of eyecolor in father and son

H_a : There is association between the darkness of eyecolor in father and son

Expected frequency $\frac{x}{128} = \frac{138}{1000} \Rightarrow x = 18$ (17.664) (freq is whole no.)

$\frac{x}{872} = \frac{138}{1000} \Rightarrow x = 120$

$\frac{x}{872} = \frac{862}{1000} \Rightarrow x = 752$

$\frac{x}{128} = \frac{862}{1000} \Rightarrow x = 110$

O	E	$(O-E)^2$	$(O-E)^2/E$
48	18	900	50
90	120	900	7.5
80	110	900	8.18
782	752	900	1.19
		$\Sigma =$	66.87

$\chi^2_{cal} = 66.87$

at 5% LOS and $(2-1)(2-1) = 1$ dof

$\chi^2_{table} = 3.84$

$\because \chi^2_{cal} > \chi_{table}$

We reject H_0.

There is an association between father and son's darkness in eye color.

Q Two batches of 12 animals are given a test of inoculation. One was inoculated and the other was not. The number of dead and survived animals were given below. Can the inoculation be regarded as effective against the disease at 5% LOS.

	Dead	Survived	Total
Inoculated	2	10	12
Not Inoculated	8	4	12
Total	10	14	24

Yate's correction as 2×2 and less than 5.

Let H_0: No association between inoculation and survival

H_a: There is an association between inoculation and survival

O	E	$[\lvert O-E\rvert-0.5]^2/E$
2	5	1.25
10	7	0.892
8	5	1.25
4	7	0.892
		$\Sigma = 4.28$

$\chi^2_{cal} = 4.28$

at 5% LOS and $(2-1)(2-1) = 1$ dof

$\chi_{table} = 3.81$

$\chi^2_{cal} > \chi^2_{table}$

We accept H_A and reject H_0.

$\therefore$ There is an association between the inoculation and the survival of the animals.

24|3|25

Goodness of fit

Q Following table gives the no of accidents during a week, find whether the accidents are uniformly distributed over a week.

Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Total
No of accidents (O)	13	15	9	11	12	10	14	84
E	12	12	12	12	12	12	12	

Let H_0: Accidents are uniformly distributed over all the days.

H_a: Accidents are not uniformly distributed over all the days.

If H_0 is true, then there will be $84/7 = 12$ accidents per day.

$$\chi^2_{cal} = \frac{(13-12)^2}{12} + \frac{(15-12)^2}{12} + \cdots + \frac{(14-12)^2}{12}$$

$$= \frac{1}{12} + \frac{9}{12} + \frac{9}{12} + \frac{1}{12} + \frac{0}{12} + \frac{4}{12} + \frac{4}{12}$$

$$= \frac{28}{12} = 2.33$$

at 5% LOS and $n-1 = 7-1 = 6$ dof

$\chi^2_{table} = 12.59$

$\because \chi^2_{cal} < \chi^2_{table}$, therefore accept H_0.

Q Theory predicts that the proportion of beans in four groups A, B, C, D should be 9 : 3 : 3 : 1. In an experiment amongst 1600 beans, the no in the four groups were 882 : 313 : 287 : 118. Does the experimental result support the theory.

	A	B	C	D	Total
	9	3	3	1	16
O	882	313	287	118	1600
E	$\frac{9\times1600}{16}=900$	$\frac{3\times1600}{16}=300$	$=300$	$=100$	

Let H_0: The experiment result supports the theory.

H_a: The experiment result dont supports the theory.

$$\chi^2_{cal} = \frac{(882-900)^2}{900} + \frac{(313-300)^2}{300} + \frac{(287-300)^2}{300} + \frac{(118-100)^2}{100}$$

$$= 0.36 + 0.56 + 0.56 + 3.24$$

$$\chi^2_{cal} = 4.72$$

restriction due to fitting total

at 5% LOS and $n-1 = 4-1 = 3$ dof

$$\chi^2_{table} = 7.815$$

$$\chi^2_{cal} < \chi^2_{table}$$

We accept H_0, the experimental result supports the theory.

Q Figures given below are a) observed frequencies of the distribution. b) frequencies of normal distribution having the same mean, s.d. and total frequency as in A. Apply χ^2 test for goodness of fit.

H_0: Fit is good

H_a: Fit is not good

O (a)	E (b)	$(O-E)^2/E$
1	2	
12 (13 merge (1<5))	15 (17)	0.94
66	66	0
220	210	0.4761
495	484	0.25
792	799	0.061
924	943	0.382
792	799	0.061
495	489	0.25
220	210	0.4761
66	66	0
12	15	0.94
1 (Less than 5 so merge 13)	2 (17)	

$$= 3.84 = \chi^2_{cal} = \Sigma(O-E)^2/E$$

Since 2 merges occur dof decreases by 2.

at 5% LOS and $13-2-3=8$ dof (mean, s.d., total frequency)

$\chi^2_{table} = 15.507$

$\chi^2_{cal} < \chi^2_{table}$

We accept H_0

$\therefore$ The fit is good.

Q Following mistakes per page were observed in a book.

Number of mistakes per page :	0	1	2	3	4	
No of pages :	211	40	19	5	0	325 (Σf_i)

$\dfrac{e^{-m} \cdot m^x}{x!} \longrightarrow$ Poisson's

$m = \dfrac{\Sigma f_i x_i}{\Sigma f_i} = \dfrac{143}{325} = 0.44$

$\therefore P(X=x) = \dfrac{e^{-m} \cdot m^x}{x!} = \dfrac{e^{-0.44} \cdot 0.44^x}{x!}$

Expected frequency : $325 \times P(X=x)$

O	P(X=x)	E = NP	$(O-E)^2/E$
211	0.649	209	0.019
40	0.283	92	0.043
19	0.0623	20	0
5	0.009	3	
0	0.001	1	

24 merge; 24; → 325 - total

$\chi^2_{cal} = \Sigma (O-E)^2/E = 0.0625$

at 5% LOS , $n = 5$ $\underset{\text{merge}}{(-2)}$ $\underset{\text{mean}}{(-2)}$ $= 1$ dof

$\boxed{f_i x_i}$ / $\boxed{f_i}$

$\chi^2_{table} = 3.841$

$\chi^2_{cal} < \chi^2_{table}$

We accept H_0, We reject H_a

Fit is good

25/3/25

Hypotheses concerning several proportions

Q A sample of 3 shipments A B C of defective items gave the following results

	A	B	C	Total
Defective	5	8	4	22
Non defective	35	42	51	118
Total	40	50	60	150

Test whether proportion of defective items is same at 5% LOS

Let $H_0 : P_1 = P_2 = P_3$

$H_a : P_1 \neq P_2 \neq P_3$

O	E		$(O-E)^2/E$
5	$\frac{x}{40} = \frac{22}{150} (5.8)$	6	0.166
8	$\frac{x}{50} = \frac{22}{150} = (7.3)$	7	0.142
4	$22-6-7 = 4$		0
35	$40-6 = 34$		0.029
42	$50-7 = 43$		0.023
51	$60-9 = 51$		0

$\Sigma = 0.3626$

$\chi_{cal}^2 = \Sigma (O-E)^2/E = 0.361$

at 5% LOS and $(r-1)(c-1) = (2-1)(3-1) = 2$ dof

$\chi_{table}^2 = 5.991$

$\chi_{cal}^2 < \chi_{table}^2$

$\therefore$ Accept H_0, proportions are same

Q 5 dice were thrown 192 times and the no of times 4, 5, and 6 were obtained are as follows.

No of dice showing 4, 5, 6	5	4	3	2	1	0
Frequency	6	46	70	48	20	

$^nC_x p^x q^{n-x} \implies n = 5 \quad x = 0-5 \quad p = q = \frac{1}{2}$

$^5C_0 (\frac{1}{2})^0 (\frac{1}{2})^5 = 0.03125$

$N \times P = 6$

$^5C_1 (\frac{1}{2})^1 (\frac{1}{2})^4 = 0.15625$

$N \times P = 30$

$^5C_2 (\frac{1}{2})^2 (\frac{1}{2})^3 = 0.3125$

$N \times P = 60$

$^5C_3 (\frac{1}{2})^3 (\frac{1}{2})^2 = 0.3125$

$N \times P = 60$

${}^5C_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right) = 0.156\ 25$ $\qquad$ ${}^5C_5\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^0 = 0.03125$

$N \times P = 30$ $\qquad$ $N \times P = 6$

	O	E		$(O-E)^2/E$
	6	6		0
	46	30		8.533
	70	60		1.667
	48	60		2.40
22 merge	20	30	36	5.444
	2	6		$\Sigma =$ 18.04

26/3/25

Module 4: Linear Programming Problem (LPP)

* Variables $x_1, x_2, \dots x_n$ which enter into the problem are called decision variables
* Function z which is to be optimized (either minimized or maximized) is called objective functions.
* Restrictions imposed on relationship between variables in the form of equalities or inequalities are called constraints.
* Any set of values $x_1, x_2 \dots x_n$ which satisfy the constraints are called solution of LPP.
* Any solution which satisfies non-negativity restrictions is called feasible solution.
* Region determined by the constraints and the axes in the first quadrant is called feasible region.
* Any feasible solution which optimises the objective function is called optimum feasible solution.
* Any solution in which one or more of the variables is called degenerate solution.

Canonical form and Standard form of LPP

① Maximise $z = \sum_{i=1}^{n} c_i x_i$, Subject to $\sum_{j=1}^{n} a_{ij} x_j \le b_i$ and $x_j \ge 0$, $\forall j$ } ⊛

Characteristics of this form are: (canonical form)

(Also for standard)

- Objective function must be of maximisation type. Must have less than or equal to type equality and if objective function is of minimisation type then must have greater than or equal to type equality

- If constraints are in the form of an equation, then express it as an inequality

(Also for standard)

- We should have $x_i \ge 0$, if any variable is unrestricted say x_j, then we write it down as $x_j = x_j' - x_j''$, x_j' & $x_j'' \ge 0$

Equality to inequality is canonical

Inequality to equality is standard

Always convert to Max
if Min z $z = -z$ and take as Maximise z

Standard form:

- In the standard form we introduce slack variables and express the objective function as well as constraints in the form of equalities.

$$\text{Maximise } z = \sum_{i=1}^{n} c_i x_i \text{ , Subject to } \sum_{j=1}^{n} a_{ij} x_j \le b_i$$
$$\text{and } x_j \ge 0 \text{ , } \forall j$$

eg Max $Z = c_1 x_1 + c_2 x_2 \dots + c_n x_n + 0s_1 + 0s_2 \dots + 0s_n$

Subject to $a_{11} x_1 + a_{12} x_2 \dots + a_{1n} x_n + s_1 + 0s_2 + 0s_3 \dots + 0s_n = b_1$

$a_{21} x_1 + a_{22} x_2 \dots a_{2n} x_n + 0s_1 + s_2 + 0s_3 + 0s_n = b_2$

$a_{m1} x_1 + a_{m2} x_2 + \dots a_{mn} x_n + 0s_1 + 0s_2 + 0s_3 \dots s_n = b_m$

$x_1, x_2 \dots x_n, s_1, s_2 \dots s_n \ge 0$

Characteristics of standard form are:

1. All the constraints are in the form of equations with the help of slack variables.

2. Right hand side of the constraints are non-negative

3. Objective function should be of maximisation type.

4. All decision variables and slack variables are non-negative.

*5. All should have restriction.

Q Convert the following LPP to standard form.

Min $z = -3x_1 + 2x_2 - x_3$

Subject to $x_1 - 3x_2 + 2x_3 \geq -6$

$3x_1 + 4x_3 \leq 3$

$-3x_1 + 5x_2 \leq 4$

$x_1, x_2 \geq 0$, x_3 is unrestricted

Since x_3 is unrestricted.

$x_3 = x_3' - x_3''$ where x_3' & $x_3'' \geq 0$ (not derivative, diff variables)

$\therefore$ The standard form is

Max $z' = -z : 3x_1 - 2x_2 + (x_3' - x_3'')$

Subject to (x - sign)

$-x_1 + 3x_2 - 2(x_3' - x_3'') + S_1 = +6$

$3x_1 + 4(x_3' - x_3'') + S_2 = 3$

$-3x_1 + 5x_2 + S_3 = 4$

$x_1, x_2, x_3', x_3'', S_1, S_2, S_3 \geq 0$

3/4/25

Q Min $z = -3x_1 + 2x_2 - x_3$

subject to $x_1 - 3x_2 + 2x_3 \geq 6$

$3x_1 + 4x_3 \leq 3$

$-3x_1 + 5x_2 \leq 4$

$x_1, x_2 \geq 0$, x_3 is unrestricted

Let $x_3 = x_3' - x_3''$

standard form is

max $z' = 3x_1 - 2x_2 + (x_3' - x_3'')$

such that / subject $x_1 - 3x_2 + 2(x_3' - x_3'') - S_1 = 6$

$3x_1 + 4(x_3' - x_3'') + S_2 = 3$

* Basic Solution is obtained by putting any n variables out of m+n variables as 0 and obtain values of remaining m variables.
* These m variables which can be 0 or non-zero are called basic variables and the other n, zero valued variables are called non-basic variables.
* Basic feasible solution is a solution which satisfies non-negativity restriction.
* If all m variables in basic feasible solutions (bfs) are positive then they are called non-degenerate basic feasible solution.
* If one or more m variables is 0 in bfs then it is called degenerate bfs.

eg Find which of them are basic feasible, non-degenerate, infeasible basic, optimum basic feasible solution.

Max $z = x_1 + 3x_2 + 3x_3$

subject to $x_1 + 2x_2 + 3x_3 = 4$

$2x_1 + 3x_2 + 5x_3 = 7$

Solution: 3 variables 2 constraints

no of n variables (non basic) : no of variables - no of constraints

$= 3 - 2 = 1$

no of basic solutions	1	2	3
non basic variable	$x_3 = 0$	$x_2 = 0$	$x_1 = 0$
basic variable	x_1, x_2	x_1, x_3	x_2, x_3
Equation and values of basic variable	$x_1 + 2x_2 = 4$ $2x_1 + 3x_2 = 7$ $x_1 = 2$ $x_2 = 1$	$x_1 + 3x_3 = 4$ $2x_1 + 5x_3 = 7$ $x_1 = 1 = x_3$	$2x_1 + 3x_3 = 4$ $3x_2 + 5x_3 = 7$ $x_2 = -1$, $x_3 = 2$
Is solution feasible	Yes	Yes	No
Is solution degenerate	No	No	No
Z- value	5	4	-
Is solution optimal	Yes	No	-

Simplex Form

1. Write the given LPP in standard form.
2. Write initial basic feasible solution
3. Make initial simplex table
4. Are all entries in z-row, non-negative?
   - ↳ if Yes: Current solution is optimal basic feasible solution.
   - if No: Select most non-negative value in z-row and the corresponding column becomes key column and corresponding variable enters the basis.
5. Obtain replacement ratio by dividing solution column by key column.
6. Are all the replacement ratios infinite and/or negative
   - ↳ if Yes : Then current solution is unbounded solution
   - if No : Select minimum finite non-negative ratio and corresponding row is called key row and corresponding variable will leave the basis.
7. Mark the key element as the intersection of key row and key column.
8. Make new simplex table or update it by using elementary row transformation
   1. Divide the key row by key element (to make key element 1)
   2. Make all other elements in key column '0' above and below key element as '0'.
9. Go to step 4 and continue till you reach step 4: Yes or step 6: Yes.

4/4/25

Q Min $z = x_1 - 3x_2 + 3x_3$

such that (subject to) $3x_1 - x_2 + 2x_3 \leq 7$

$2x_1 + 4x_2 \geq -12$

$-4x_1 + 3x_2 + 8x_3 \leq 10$

Standard form: RHS is positive, multiply by -2 and slack variable

Max $z' = -z = -x_1 + 3x_2 - 3x_3$

s.t. $3x_1 - x_2 + 2x_3 + S_1 = 7$

$-2x_1 - 4x_2 + S_2 = 12$

$-4x_1 + 3x_2 + 8x_3 + S_3 = 10$

$x_1, x_2, x_3, S_1, S_2, S_3 \geq 0$

non basic variable $= 6 - 3 = 3$

$x_1 = 0 = x_2 = x_3$

$S_1 = 7, \quad S_2 = 12, \quad S_3 = 10$

$z' = 0$

Step 3 :

Key column

Iteration No:	Basic Variable	Coefficient x_1	x_2	x_3	x_4	S_1	S_2	RHS Solution	Ratio
0	z	-107	-1	-2	0	0	0	0	-
x_1 enters,	x_4	$14/3$	$1/3$	-2	1	0	0	$7/3$	$1/2$
S_2 leaves	S_1	16	$1/2$	-6	0	1	0	5	$5/16$
	S_2	*3	-1	-1	0	0	1	0	$0/3$
1	Z	0	$-110/3$	$-113/3$	0	0	$107/3$	0	-
x_3 enters	x_4	0	$17/9$	$-4/9$	1	0	$-14/9$	$2/3$	$-3/2$ -ve
	S_1	0	$35/6$	$-2/3$	0	1	$-16/3$	5	$-15/2$ -ve
	x_1	1	$-1/3$	$-1/3$	0	0	$1/3$	0	-0

Key Row

Key Element

$\therefore$ From the table,

$x_1 = 31/5$, $x_2 = 174/15$, $x_3 = 0$

$S_1 = 0$, $S_2 = 1062/15$, $S_3 = 0$

Max $z' = 143/5$

Min $z = -143/5$

Q Solve following LPP by simplex.

Max $z = 107x_1 + x_2 + 2x_2$

st $14x_1 + x_2 - 6x_3 + 3x_4 = 7$

$16x_1 + (1/2)x_2 - 6x_3 \leq 5$

$3x_1 - x_2 - x_3 \leq 0$

The standard form is

Max $z = 107x_1 + x_2 + 2x_3$

s.t. $\frac{14}{3}x_1 + \frac{1}{3}x_2 - 2x_3 + x_4 = 7/3$

$16x_1 + \frac{1}{2}x_2 - 6x_3 + S_1 = 5$

$3x_1 - x_2 - x_3 + S_2 = 0$

$x_1, x_2, x_3, x_4, S_1, S_2 \geq 0$

Let $x_1 = x_2 = x_3 = 0$

$x_4 = 7/3$

$S_1 = 5, \quad S_2 = 0$

Coefficient ↓ most negative

Iteration No:	Basic Variable	x_1	x_2	x_3	S_1	S_2	S_3	RHS Soln	Ratio
0	z	1	$\boxed{-3}$	3	0	0	0	0	
x_2 enters,	S_1	3	-1	2	1	0	0	7	-7 $(7/-1)$
S_3 leaves	S_2	-2	-4	0	0	1	0	12	-3 $(12/-4)$
(most non-neg value of column)	S_3	-4	$\boxed{3}$ *	8	0	0	1	10	$10/3$ $(10/3)$
1	z	$\boxed{-3}$	0	11	0	0	1	10	
x_1 enters,	S_1	$\boxed{5/3}$	0	$14/3$	1	0	$1/3$	$31/3$	$31/5$
S_1 leaves	S_2	$-22/3$	0	$32/3$	0	1	$4/3$	$76/3$	$-76/22$
	x_2	$-4/3$	1	$2/3$	0	0	$1/3$	$10/3$	$-5/2$
2	z	0	0	$37/5$	$3/5$	0	$8/5$	$143/5$	
	x_1	1	0	$14/5$	$3/5$	0	$1/5$	$31/5$	
	S_2	0	0	$468/15$	$66/15$	1	$42/15$	$1062/15$	
	x_2	0	1	$96/15$	$12/15$	0	$5/15$	$174/15$	

→ Key Column

→ Key Row

Key Element (make 1)

Since all entries in z now is non-negative, the current solution is optimal BFS.

Here $-113/3$ is least no., ∴ x_3 is incoming but all the ratios are -ve, hence no variable leaves basis, ∴ x_3 cannot enter, ∴ Soln is unbounded.

7/4/25

Max $z = 3x_1 + 2x_2$

Subject to $= 3x_1 + 2x_2 \leq 18$

$0 \leq x_1 \leq 4$

$0 \leq x_2 \leq 4$

Sol: Standard form

Step 1: Max $z = 3x_1 + 2x_2$

$z - 3x_1 - 2x_2 = 0$

subject to $3x_1 + 2x_2 + S_1 = 18$

$x_1 + S_2 = 4$

$x_2 + S_3 = 6$

$x_1, x_2, S_1, S_2, S_3 \geq 0$

Step 2: The initial basic feasible solution is $x_1 = x_2 = 0$

$\therefore \; S_1 = 18, \; S_2 = 4, \; S_3 = 6$

$\therefore \; z = 0$

Step 3:

Iteration No	Basic Variable	Coefficient x_1	x_2	S_1	S_2	S_3	RHS Solution	Ratio
0	z	-3	-2	0	0	0	0	
x_1 enters,	S_1	3	2	1	0	0	18	6
S_2 leaves	S_2	* 1	0	0	1	0	4	4 →
	S_3	0	1	0	0	1	6	∞
1	z	0	-2	0	3	0	12	
x_2 enters,	S_1	0	* 2	1	-3	0	6	3 →
S_1 leaves	x_1	1	0	0	1	0	4	∞
	S_3	0	1	0	0	1	6	6
2	z	0	0	1	0	0	18	
S_2 enters,	x_2	0	1	$1/3$	$-3/2$	0	3	-2
S_3 leaves	x_1	1	0	0	1	0	4	4
	S_3	0	0	$-1/4$	$3/2$	1	3	2 →
3	z	0	0	1	0	0	18	
	x_1	0	1	0	0	1	6	
	x_2	1	0	$1/3$	0	$-2/3$	2	
	S_2	0	0	$-1/3$	1	$2/3$	2	

(non basic, surplus variable is 0, so extra step)

* Since all entries in z row are non-negative the current solution is optimal basic feasible solution.

$x_1 = 4, \; x_2 = 3, \; S_1 = 0, \; S_2 = 0, \; S_3 = 3, \; Z_{max} = 18$

Entry for non-basic variable S_2 is 0 in z-row this indicates that the LPP has an alternate solution.
Alternate optimal solution is obtained by taking S_2 as the entering variable and corresponding column is key column.

Final Answer: Since all entries in z-row are non-negative. Therefore alternate optimal basic feasible solution is

$x_1 = 2, \; x_2 = 6, \; S_1 = 0, \; S_2 = 2, \; S_3 = 0, \; Z_{max} = 18$

$\therefore$ Infinite number of non basic optimal feasible solutions are $X = \lambda x_1 + (1-\lambda) x_2 \qquad 0 < \lambda < 1$

where $x_1 = \begin{bmatrix} 4 \\ 3 \\ 0 \\ 0 \\ 3 \end{bmatrix} \quad x_2 = \begin{bmatrix} 2 \\ 6 \\ 0 \\ 2 \\ 0 \end{bmatrix} \quad \therefore \; X = \begin{bmatrix} 2+2\lambda \\ 6-3\lambda \\ 0 \\ 2-2\lambda \\ 3\lambda \end{bmatrix}$

Penalty and Duality

The simplex method we have studied can not be used if the contraints are of $\geq$ type.

$\therefore$ we apply the following method.

The Big M- method

The method is due to Charmes and is based on the following considesation.

1. If the constraint is of $\geq$ Type then add $-s_1, -s_2$ ....-etc to convert $\geq$ type to equality but we would not get a unit matrix, to overcome this difficulty.
2. we introduce in addition to surplus variables $-s_1, -s_2$ ... artificial variables A_1, A_2 ... with +ve sign in these contraints.
3. An artificial variable is introduced even when the constraint is of equality type.
4. In the objective function we assign big penalty by subtracting $MA_1, MA_2, \ldots$ if the objective functn is of max type.

consider The following model problem.

Max $z = c_1x_1 + c_2x_2 + c_3x_3$

subject to
$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \geq b_1$$
$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \leq b_2$$
$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \leq b_3$$
$$x_1, x_2, x_3 \geq 0$$

subtract s_1 and add an artificial variable A_1 and assign big penalty in the object function ie subtract MA_1 from it

$\therefore$ we have

Max $z = c_1x_1 + c_2x_2 + c_3x_3 - MA_1$

subject to
$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 - s_1 + A_1 = b_1$$
$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + s_2 = b_2$$
$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + s_3 = b_3$$
$$x_1, x_2, x_3, s_1, s_2, s_3, A_1 \geq 0$$

we now write the objective function free from artificial variable by adding M times, the first constraint to the object functⁿ.

$\therefore$ we get

$$z = (c_1 + Ma_{11})x_1 + (c_2 + Ma_{12})x_2 + (c_3 + Ma_{13})x_3 - MS_1 - Mb_1$$

$$\therefore z - (c_1 + Ma_{11})x_1 - (c_2 + Ma_{12})x_2 - (c_3 + Ma_{13})x_3 + MS_1 + Mb_1 = 0.$$

Now follow simplex steps as before.

After the required # of iteration we will find one of the following situations.

(1) The artificial variables leave the process and the optimality condition is satisfied by the basic variables. This is then the optimal basic feasible solⁿ.

(2) Atleast one of the artificial variable remains in the basis with zero value and the optimality condition is satisfied. this is the optimal basic feasible solⁿ (though degenerate).

(3) Atleast one of the artificial variables remains in the basis with non-zero value and the optimality condition is satisfied. This solⁿ though satisfies optimality conditions is not an optimal solⁿ since it contains large penalty M. This is not a solⁿ but a pseudo-solⁿ.

Big M method

Use penalty method to solve the following LPP

Max $z = 6x_1 + 4x_2$

subject to $2x_1 + 3x_2 \leq 30$

$3x_1 + 2x_2 \leq 24$

$x_1 + x_2 \geq 3$, $x_1, x_2 \geq 0$

Standard form Max $z = 6x_1 + 4x_2 - MA_3$

subject to $2x_1 + 3x_2 + S_1 = 30$

$3x_1 + 2x_2 + S_2 = 24$

$x_1 + x_2 - S_3 + A_3 = 3$

We eliminate term $-MA_3$ from objective function by adding M times 3rd constraint to objective function.

$\therefore z = 6x_1 + 4x_2 - MA_3 + Mx_1 + Mx_2 - MS_3 + MA_3$

8/4/25

$z = 6x_1 + 4x_2$

subject to $2x_1 + 3x_2 \leq 30$

$3x_1 + 2x_2 \leq 24$

$x_1 + x_2 \geq 3$

$x_1, x_2 \geq 0$

Solution: Standard form is:

$z - (6+M)x_1 - (4+M)x_2 + MS_3 = -3M$

subject to $2x_1 + 3x_2 + S_1 = 30$

$3x_1 + 2x_2 + S_2 = 24$

$x_1 + x_2 - S_3 + A_3 = 3$

$x_1, x_2, S_1, S_2, S_3, A_3 \geq 0$

IBFS: $x_1 = x_2 = S_3 = 0$

$\Rightarrow S_1 = 30$, $S_2 = 24$, $A_3 = 3$

Iteration No	Basic Variable	x_1	x_2	S_1	S_2	S_3	A_3	RHS Soln	Ratio	
		Coefficient								
0	z	-6-M	-4-M	0	0	M	0	-3M		$R_1 + (6+M)R_4$
										$R_2 - 2R_4$
x_1 enters,	S_1	2	3	1	0	0	0	30	15	$R_3 - 3R_4$
A_3 leaves	S_2	3	2	0	1	0	0	24	8	
	A_3	* 1	1	0	0	-1	1	0	0	

1	Z	0	2	0	0	-6	18	
S_3 enters,	S_1	0	1	1	0	2	24	12
S_2 leaves	S_2	0	-1	0	1	*3	15	5
	x_1	1	1	0	0	-1	6	-3
2	Z	0	0	0	2	0	48	
x_2 enters,	S_1	0	*$5/3$	1	$-2/3$	0	14	$42/5$
S_1 leaves	S_3	0	$-1/3$	0	$1/3$	1	5	-15
	x_1	1	$2/3$	0	$1/3$	0	8	12
3	Z	0	0	0	2	0	48	
	x_2	0	1	$3/5$	$-2/5$	0	$42/5$	
	S_3	0	0	$1/5$	$1/5$	1	$39/5$	
	x_1	1	0	$-2/5$	$3/5$	0	$12/5$	

Since all the entries in z-row is non negative the current solution is optimal basic feasible solution

$x_1 = 8, \quad x_2 = 0, \quad S_1 = 14, \quad S_2 = 0, \quad S_3 = 5$

$Z_{max} = 48$

The value of non basic variable x_2 is 0 in z-row, therefore LPP alternate solution.

Alternate Solution:

$x_1 = 12/5 \qquad x_2 = 42/5, \qquad S_1 = 0$

$S_2 = 0, \quad S_3 = 39/5$

$Z_{max} = 48$

$\therefore$ Infinitely many solutions

$$X = \lambda x_1 + (1-\lambda) x_2 \qquad 0 < \lambda < 1$$

$$x_1 = \begin{bmatrix} 8 \\ 0 \\ 14 \\ 0 \\ 5 \end{bmatrix} \qquad x_2 = \begin{bmatrix} 12/5 \\ 42/5 \\ 0 \\ 0 \\ 39/5 \end{bmatrix}$$

Q Using M method minimise $z = x_1 + 2x_2 + x_3$

subject to $x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_3 \leq 1$

$\frac{3}{2}x_1 + 2x_2 + x_3 \geq 8$