Date ___/___/___
Page _____

Data
- Unstructured
- Structured
  - Categorical
  - Numerical

Scale of measurement

1. Nominal → object / name / Identities No order
2. Ordinal → ——— ——— ——— order & rank
3. Interval → ———, ———, ——— order & rank, no absolute zero diff. (−) is possible
4. Ratio → ———, ———, ——— has absolute zero, mult, division, zero has a meaning

Descriptive Statistics

- graphical
- numericals
  - Central tendency
  - Spread
  - Peak

Measure of Central Tendency / location

mean, median, mode

Mean = $\bar{x}$

Median = M

mode

	Raw Data	Frequency	Class Interval
Mean	$\frac{\Sigma x}{x}$	$\frac{\Sigma f x}{\Sigma f}$	$\frac{\Sigma f u}{\Sigma f}$ u = midpoint of each interval

Class Interval	Mid pt				u
$x_0 - x_1$	f_1	u_1	$x_0 - x_1$	$x_0 - x_1'$	
$x_1 - x_2$	f_2	u_2	$x_2 - x_3$	$x_1' - x_2'$	
$x_2 - x_3$	f_3	u_3	$x_4 - x_5$	$x_2' - x_3'$	

open Intervals, unequal intervals

0 − 10	10 − 20	20 − 40	40 − 81
f_1	f_2	f_3	f_4

Date ___/___/___
Page ___

	Raw Data	Frequency	Class Interval
Median	sort, odd $\left(\frac{n+1}{2}\right)^{th}$ term even → $\frac{n}{2}^{th}$ & $\left(\frac{n}{2}+1\right)^{th}$ avg	cf, $N/2$, [cf > $N/2$] → $\boxed{x}$	$l + \frac{h}{f}\left(\frac{N}{2} - c\right)$ l: lower band c: preceding cf of median class
Mode	highest repetition	highest frequency	$l + h\left(\frac{f_k - f_{k-1}}{2f_k - f_{k-1} - f_{k+1}}\right)$

Q Following is the distribution of the rent paid by 88 employees. Find the median, mean & mode of the distribution

Rent	No. of employees
1000	5
1200	15
1500	25
1700	20
1800	14
2000	7
2500	2

$$\text{Mean} = \frac{5000 + 18000 + \dots 5000}{88}$$

$$= 1576.136$$

Median

1000	5	5	$N = 88$ $\quad N/2 = 44$
1200	15	20	
1500	25	45	cf = 45 > 44
1700	20	65	
1800	14	79	
2000	7	86	
2500	2	88	Median = 1500

Mode : 1500

Date __/__/__
Page _____

Q:

Expenditure	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of smokers	23	44	35	12	9	3	2

A survey of 128 smokers revealed the following frequency distribution of their daily expenditure on smoking

		u
10-20	23	15
20-30	44	25
30-40	35	35
40-50	12	45
50-60	9	55
60-70	3	65
70-80	2	75

Expen	No. (f)	midpt (u)	uf	cf
10-20	23	15		23
20-30	44	25		67
30-40	35	35		102
40-50	12	45		114
50-60	9	55		123
60-70	3	65		126
70-80	2	75	~~4050~~	128
		-	4050	-

$$\bar{x} = \frac{\Sigma fu}{\Sigma f} = \frac{4050}{128} = 31.64$$

(Median) $\frac{N}{2} = \frac{128}{2} = 64$, $cf > N/2$

$67 > 64$

Median class : 20-30, $f = 44$

$c = 23$, $h = 10$, $l = 20$

$$= 20 + \frac{10}{44}(64 - 23)$$

$$= 29.31$$

$$\text{Mode} = 20 + 10\left(\frac{44 - \cancel{35}\,23}{2 \times 44 - \cancel{23} - 35}\right) = \cancel{20}\ 27$$

Date __/__/__
Page

Q

% marks	10-19	20-29	30-39	40-49	50-59	60-69	70-
no. of students	8	19	29	36	25	13	4

9.5 - 19.5 19.5 - 29.5 29.5 - 39.5 39.5 - 49.5
49.5 - 59.5 59.5 - 69.5 69.5 -

% marks	no. of students	cf	u	uf
9.5 - 19.5	8	8	14.5	
19.5 - 29.5	19	27	24.5	
29.5 - 39.5	29	56	34.5	
39.5 - 49.5	36	92	44.5	
49.5 - 59.5	25	117	54.5	
59.5 - 69.5	13	130	64.5	
69.5 - 79.5	4	134	74.5	

Mean = 42.410

$$\text{Median} = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$$

$N = 134/2 = 67$ $h = 10$

Median class = 39.5 - 49.5 $f = 36$ $l = 39.5$

$c = \cancel{29}\ 56$

$$= 39.5 + \frac{10}{134}(67 - 56)$$

$= 42.55$

$$\text{Mode} = 39.5 + 10\left(\frac{36 - 29}{72 - 29 - 25}\right)$$

$= 43.38$

Date __/__/__
Page _____

Q

1	1	1
2	16	17
3	8	7
4	14	15
5	7	5
6	4	4
7	5	5
8	1	1
9	2	2
10	1	1
11	3	3
12	1	1

The median & mode of the following distribution are known to 33.5 and 34. Find the values of f_1, f_2, f_3

wages	0-10	10-20	20-30	30-40	40-50	50-60	60-70
frequency	4	16	f_1	f_2	f_3	6	4

Total 230

wages	frequency	cf	u	uf
0-10	4	4	5	
10-20	16	20	15	
20-30	f_1	$20+f_1$	25	
30-40	f_2	$20+f_1+f_2$	35	
40-50	f_3	$20+f_1+f_2+f_3$	45	
50-60	6	$26+f_1+f_2+f_3$	55	
60-70	4	$30+f_1+f_2+f_3$	65	$850+f_1+f_2+f_3$

$$\text{Mean} = \frac{850+f_1+f_2+f_3}{30+f_1+f_2+f_3}$$

Date ___/___/
Page ___

Median = 33.5

$$\text{Median} = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$$

Median class ⇒ 30 - 40

$l = 30$ $\quad cf = 20 + f_1$

$h = 10$ $\quad c = 20 + f_1$

$f = 30 + f_1 + f_2 + f_3$

$$33.5 = 30 + \frac{10}{30 + f_1 + f_2 + f_3}\left((30 + f_1 + f_2 + f_3)/2 - (20 + f_1)\right)$$

$$670 + 33.5f_1 + 33.5f_2 = 600 + 30f_1 + 30f_2 + 150 + 5f_1 + 5f_2 + 5f_3 - 200$$

$$70 + 33.5f_1 + 33.5f_2 = 30f_1 + 30f_2 - 50 + 5f_1 + 5f_2 + 5f_3 - 10f_1$$

$$120 + 33.5f_1 + 33.5f_2 = 25f_1 + 35f_2 + 5f_3$$

$$120 = -8.5f_1 + 1.5f_2 + 5f_3$$

$$24 = -1.7f_1 + 0.3f_2 + f_3$$

$$\text{Mode} = l + h\frac{(f_k - f_{k-1})}{(2f_k - f_{k-1} - f_{k+1})}$$

$$34 = 30 + 10\left(\frac{\cancel{20} + \cancel{f_1} + f_2 - \cancel{20} - \cancel{f_1}}{\cancel{40} + \cancel{2f_1} + \cancel{2}f_2 - \cancel{20f_1} - \cancel{20} - \cancel{f_1} - \cancel{f_2} - f_3}\right)$$

$$34 = 30 + \frac{10f_2}{f_2 - f_3}$$

$$34f_2 - 34f_3 = 30f_2 - 30f_3 + 10f_2$$

$$4f_2 - 4f_3 = 10f_2$$

$$-4f_3 = 6f_2$$

$$230 = 30 + f_1 + f_2 + f_3$$

$$200 = f_1 + f_2 + f_3$$

Date ___/___/___
Page _____

Measure of Spread

Absolute measure	Relative Measure
① Range	Coeff. of Range
② Quartile Deviation	Coeff of Q.D
③ Mean Deviation	Coeff of M.D
④ Variance or SD	Coeff of variation

① Range:-

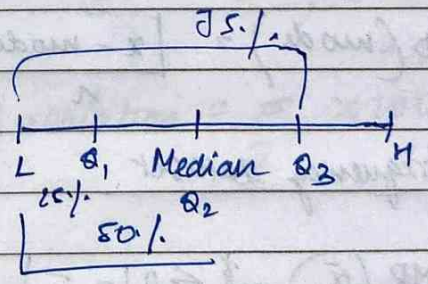
a - - - - - - - b
$\downarrow$ $\downarrow$
L H

Range : $H-L$

coeff of Range :- $\frac{H-L}{H+L}$

② Quartile Deviation

$QD = \frac{Q_3 - Q_1}{2}$



coeff of $QD = \frac{Q_3 - Q_1}{Q_3 + Q_1}$

Q_1 = First Quartile or lower Quartile
Q_3 = Third " or Upper Quartile
Q_2 = Second " or Median

$Q_1 = \left(\frac{n+1}{4}\right)^{th} = \frac{5}{4}^{th} = 1.25$

eg 1 2 3 4

Frequency Data

$Q_2 = \frac{(n+1)^{th}}{2} =$ $Q_r = \left[r\left(\frac{n+1}{4}\right)\right]^{th}$ term

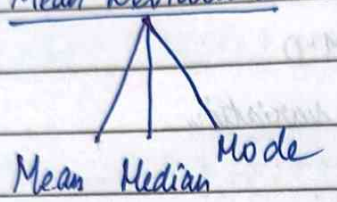
$Q_3 = 3\frac{(n+1)}{4} =$ $r = 1, 2, 3$

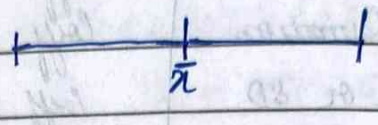
Date ___/___/___
Page ___

Class - Interval Data

$$Q_r = l + \frac{h}{f}\left[\frac{rN}{4} - c\right]$$

(3) Mean Deviation





$x \Rightarrow$

$$MD(\bar{x}) = \frac{\sum|x-\bar{x}|}{n}$$

Rawdata

$|-3|$
$|-1|$
0
3
1

$$MD(M) = \frac{\sum|x-M|}{n}$$

$$MD(\text{mode}) = \frac{|x-\text{mode}|}{n}$$

Frequency dataset

x	f
x_1	f_1
x_2	f_2
$\vdots$	$\vdots$
x_n	f_n

$$MD(\bar{x}) = \frac{\sum f|x-\bar{x}|}{N}$$

$$, \bar{x} = \frac{\sum fx}{\sum f = N}$$

$$MD(M) = \frac{\sum f|x-M|}{N}, \quad M = ?$$

$$MD(\text{Mode}) = \frac{\sum f|x-\text{Mode}|}{N}$$

Date __/__/__
Page _____

Class Interval (Same as frequency dataset provided $\bar{x}$, M, mode as per class interval definition)

Coeff of MD

$$= \frac{MD(\bar{x})}{\bar{x}}$$

coeff of MD (M)

$$= \frac{MD(M)}{M}$$

coeff of MD (mode)

$$= \frac{MD(mode)}{mode}$$

Variance $(V) = \frac{\Sigma(x-\bar{x})^2}{n}$, Raw Data $= \frac{\Sigma x^2}{n} - \bar{x}^2$

= frequency $\Rightarrow \frac{\Sigma f(x-\bar{x})^2}{N} = \frac{\Sigma f x^2}{N} - \bar{x}^2$

$= \frac{\Sigma f(u-\bar{u})^2}{N} = \frac{\Sigma f u^2}{N} - \bar{u}^2$, class Interval u = mid point

$\sigma = \sqrt{V}$, coeff of variation $= \frac{\sigma}{\bar{x}} \times 100$

For the following data find the coefficient of range & coefficient of quartile deviation

36, 43, 30, 37, 38, 35, 29, 38, 35, 32, 35, 36

$\Rightarrow$ 29, 30, 32, 35, 35, 35, 36, 36, 37, 38, 38, 43

$L = 29$, $H = 43$

Coeff of Range $= \frac{H-L}{H+L} = \frac{43-29}{43+29} = 0.194$

$n = 12$

$Q_r = \left(\frac{n+1}{4}\right)^{th}$ term

$r = 1$, $Q_1 = \left(\frac{12+1}{4}\right)^{th} = (3.25)^{th}$ term

$= 3^{rd}$ term $+ 0.25\,[4^{th} - 3^{rd}]$

$= 32 + 0.25\,[35 - 32]$

$= 32 + 0.25\,(3)$

$= 32.75$

$r = 3,\ Q_3 = \left[3\left(\frac{12+1}{4}\right)\right]^{th} = (9.75)^{th}$ term

$= 9^{th}$ term $+ (0.75)\,[10^{th} - 9^{th}]$

$= 37 + (0.75)\,[38 - 37]$

$= 37.75$

$Q_2 = \underline{\underline{35.5}}$

coeff $QD = \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \dfrac{37.75 - 32.75}{37.75 + 32.75}$

$= 0.0709$

Q For the following distribution of number of students present in the class on different days, find the Quartile Deviation & coefficient of quartile deviation

Students present	no. of days f	cf
75	13	13
76	24	37
77	27	64
78	21	85
79	11	96
80	2	98

$N = \Sigma f = 98$

$Q_r = r\left(\frac{N+1}{4}\right)^{th}$ term

$Q_1 = \left(\frac{98+1}{4}\right)^{th}$ term $= 24.75$

Date ___/___/___
Page _____

$cf = 37 > 24.75$

$\boxed{Q_1 = 76}$

$Q_3 = \frac{3(N+1)}{4}^{th}$ term

$= 3 \times (24.75)$

$= (74.25)^{th}$ term

$cf = 85 > 74.25$

$\boxed{Q_3 = 78}$

$QD = \frac{Q_3 - Q_1}{2} = \frac{78 - 76}{2} = 1$

coefficient of Quartile Deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{1}{154}$

$= 0.0129$

Q Find the coefficient of quartile deviation, marks, students

exclusive data

marks	students
0 - 19	15
20 - 39	31
40 - 49	19
50 - 59	15
60 - 69	8
70 - 79	7
80 - 89	7

Inclusive	marks	students
	-0.5 - 19.5	15
	19.5 - 39.5	31
	39.5 - 49.5	19
	49.5 - 59.5	15
	59.5 - 69.5	8
	69.5 - 79.5	7
	79.5 - 89.5	7

marks	student	cf	
-0.5 - 9.5	~~15~~ 7	~~15~~ 7	
9.5 - 19.5	~~15~~ 8	~~30~~ 15	
19.5 - 29.5	~~31~~ 15	~~61~~ 30	
29.5 - 39.5	~~31~~ 16	~~92~~ 46	
39.5 - 49.5	19	65	Not needed
49.5 - 59.5	15	80	
59.5 - 69.5	8	88	
69.5 - 79.5	7	95	
79.5 - 89.5	7	102	

$$Q_r = l + \frac{h}{f}\left[r\left[\frac{N}{4}\right] - c\right]$$

$N = 102$

$\frac{N}{4} = \frac{102}{4} = 25.5$ $\qquad$ $cf = 80 > 25.5$

Q_1 class $\Rightarrow$ 19.5 - 29.5 $\qquad$ First quartile class

$h = 10$ $\qquad$ $l = 19.5$

$f = 15$

$c = 15$

$$Q_1 = 19.5 + \frac{10}{15}\left[25.5 - 15\right]$$

$Q_1 = 26.5$

Date __/__/__
Page _____

For Q_3 class

$\frac{3N}{4} = 3(25.5) = 76.5$

$cf = 80 > 76.5$

Third Quartile class = 49.5 – 59.5

$l = 49.5$

$f = 15$

$c = 65$

$h = 10$

$r = 3$

$$Q_3 = l + \frac{h}{f}\left[r\left[\frac{N}{4}\right] - c\right]$$

$$Q_3 = 49.5 + \frac{10}{15}\left[76.5 - 65\right]$$

$$= 57.1667$$

coeff of Quartile Deviation:

$$\frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{57.1667 - 26.5}{57.1667 + 26.5}$$

$$= 0.3666$$

Q

IQ	No. of children
less than 80	2
— 90	16
— 95	33
— 100	55
— 105	77
— 110	95
— 130	100

Following is the distribution of IQ of 100 children. Find the coeff. of Quartile Deviation

Soln

IQ	No. of children	f	Interval
less than 80	2	2	0 – 80
– 90	~~14~~ 16	14	80 – 90
– 95	~~19~~ 33	17	90 – 95
– 100	~~36~~ 55	22	95 – 100
– 105	~~41~~ 77	22	100 – 105
– 110	~~54~~ 95	18	105 – 110
– 130	~~46~~ 100	5	110 – 130

Interval	f	cf
0-80	2	2
80-90	14	16
90-100	39	55
100-110	40	95
110-120	2	97
120-130	3	100

$$Q_r = l + \frac{h}{f}\left(r\left(\frac{N}{4}\right) - c\right)$$

$\frac{N}{4} = \frac{100}{4} = 25$ $\quad$ CF = 55 > 25

Q_1 class = 90 - 100 $\quad$ $l = 90$

$f = 39$ $\quad$ $h = 10$

$c = 16$

$$Q_1 = 90 + \frac{10}{39}\left(25 - 16\right)$$

$Q_1 = 92.307$

$\frac{3N}{4} = 75$

Q_3 class 100 - 110

$l = 100$

$f = 40$ $\quad$ $100 + \frac{10}{40}\left[75 - 55\right]$

$h = 10$

$c = 55$

$Q_3 = 105$

$$\text{Coeff of QD} = \frac{105 - 92.307}{105 + 92.307}$$

$= 0.06433$

marks, students

Q Compute the coefficient of MD from median.

given in question

Marks	students	cf	u	$\lvert x-M \rvert$	$f\lvert x-M \rvert$
0-10	2	2	5	24	48
10-20	10	12	15	14	140
20-30	20	32	25	4	80
30-40	15	47	35	6	90
40-50	10	57	45	16	160
50-60	3	60	55	26	78
	$\Sigma f = 60$				$\Sigma = 596$

$$MD(M) = \frac{\Sigma f \lvert x - M \rvert}{N}$$

$$=$$

$$\frac{N}{2} = \frac{60}{2} = 30$$

class ⇒ 20 - 30

$f = 20$

$c = 12$

$$\text{Median} = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$$

$$= 20 + \frac{10}{20}(30 - 12)$$

$$= 20 + \frac{\cancel{18}\,9}{\cancel{2}}$$

$$= 29$$

Median

$$MD(M) = \frac{\Sigma f \lvert x - M \rvert}{N} = \frac{596}{60} = 9.933$$

$$\text{Coeff of Median} = \frac{MD(M)}{M} = \frac{9.933}{\cancel{60}\,29} = 0.3425$$

Date __/__/__
Page ____

Q

wage	workers
40-60	12
60-80	47
80-100	31
100-120	8
120-140	2
140-160	3

Find coeff of MD from mean

	f	u	uf	$\lvert x-\bar{x}\rvert$	$f\lvert x-\bar{x}\rvert$
40-60	12	50	600	30.3	363.6
60-80	47	70	3290	10.3	484.1
80-100	31	90	2790	9.7	300.7
100-120	8	110	880	29.7	237.6
120-140	2	130	260	49.7	99.4
140-160	3	150	450	69.7	$\Sigma fu = 8270$
	$f = 103$				

~~Σ 50xf~~

$$\text{Mean} = \frac{\Sigma fu}{\Sigma f}$$

$$= \frac{50\times12 + 70\times47 + 90\times31 + 110\times8 + 130\times2 + 3\times150}{103}$$

$$= \frac{8270}{103}$$

$$= 80.291$$

$$\bar{x} = 80.3$$

$$\Rightarrow \frac{\Sigma f\lvert x-\bar{x}\rvert}{\Sigma N} = \frac{1694.5}{103}$$

$$= 16.451$$

$$\text{coeff of MD}(\bar{x}) = \frac{MD(\bar{x})}{\bar{x}} = \frac{16.451}{80.3}$$

$$= 0.204$$

Date __/__/__
Page _____

Q The following are the runs scored by 2 batsmen A & B in 10 innings.

(i) Who is the better run scorer (mean)

(ii) Who is more consistent in scoring (coefficient of variance)

A:	101	27	0	36	82	45	7	13	65	14
B:	97	12	40	96	13	8	85	8	56	15

A	B
101	97
27	12
0	40
36	96
82	13
45	8
7	85
13	56
65	15
14	8

$\bar{x} = \frac{390}{10} = 39$ $\quad$ $\bar{x} \Rightarrow \frac{430}{10} = \underline{43}$

(i) B is the better scorer

$\sigma_A = \sqrt{\frac{\Sigma x_A^2}{n} - (\bar{x}_A)^2}$

Variance $= \frac{\Sigma (x-\bar{x})^2}{n}$

$\sigma_B = \sqrt{\frac{\Sigma x_B^2}{n} - (\bar{x}_B)^2}$

A ⇒ $\sigma = 32.723$

coeff of var $= \frac{\sigma}{\bar{x}} \times 100 = \frac{32.723}{39} \times 100$
$= 82.9$

B ⇒ $\sigma = 35.75$

coeff of var $= \frac{\sigma}{\bar{x}} \times 100 = \frac{35.75}{43} \times 100$
$= 83.07$

Since coeff of variance for A is lesser hence he is more consistent

Date __/__/__
Page ____

pay	A	B
400-600	4	10
600-800	18	20
800-1000	25	42
1000-1200	2	18
1200-1400	1	10

The following are the distribution of monthly pay of workers of 2 factories

(i) In which factory is total salary bill higher.

(ii) In which factory is average salary higher

(iii) Which factory has salary variation more

pay	u	A	B	$f_A x$ A ⇒ fu	$f_B x$ B ⇒ fu
400-600	500	4	10	2000	5000
600-800	700	18	20	12600	1400
800-1000	900	25	42	22500	37800
1000-1200	1100	2	18	2200	19800
1200-1400	1300	1	10	1300	13000
		50	100	Σ 40600	Σ 89600

$\sigma_A = \sqrt{\frac{\Sigma f_A x^2}{N} - (\bar{x}_A)^2}$ $\bar{x}_A = 812$ $\sigma_A = 155.74$

$\sigma_B = \sqrt{\frac{\Sigma f_B x^2}{N} - (\bar{x}_B)^2}$ $\bar{x}_B = 896$ $\sigma_B = 217.218$

$f_B x^2 = 85000000$

$f_A x^2 = 39180000$

Date ___/___/___
Page _____

coff of variation (A) $= \frac{\sigma_A}{\bar{x}_A} \times 100$

$= \frac{155.79}{812} \times 100$

$= 19.18$

coff of variation (B) $= \frac{\sigma_B}{\bar{x}_B} \times 100$

$= \frac{217.218}{896} \times 100$

$= 24.241$

(i) Total Bill (A) = $\sum x = 90600$
$\sum x = 89600$ > B has more salary bill

(ii) Average Bill

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

Date ___/___/___
Page ___

Probability Distributions

R.V ⇒ real no. assigned S.P →

Range

Discrete — continuous (Infinite countable)

(finite | Infinite countable)

Probability Function

	Discrete	continuous
① Range	finite | Infinite countable	Infinite uncountable.
② P.F	Probability mass function $P(x)$ $P(x) \geq 0 \ \forall x$ $\sum P(x) = 1$	Probability density function $f(x)$ $f(x) \geq 0 \ \forall x$ $\int_{-\infty}^{+\infty} f(x)\,dx = 1$
③ C.F	$F(x) = P(X \leq x) = \sum_{0}^{x} P(x)$	$F(x) = \int_{-x}^{x} f(x)\,dx$
④ mean	$\mu = \sum x P(x)$	$\mu = \int_{-\infty}^{\infty} x f(x)\,dx$
⑤ variance	$V = \sigma^2 = \sum (x-\mu)^2 P(x)$	$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$ $= \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$

Date ___/___/___
Page _____

1 2 3 . . . 10

odd = 5

T = 10

$p = \frac{5}{10} = \frac{1}{2}$

2 tickets | 3 draws

Binomial

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

$= {}^3C_2 \, p^2 q^{3-2}$

$= {}^3C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^1 =$

$\Rightarrow 3/8 = 0.375$

success or failure

conditional probability

(heavily dependencies)

$A_1 A_2 A_3$, $\bar{A}_1 \bar{A}_2 \bar{A}_3$

$P(A_1 A_2 \bar{A}_3) + P(A_1 \bar{A}_2 A_3) + P(\bar{A}_1 A_2 A_3)$

$P(A_1 \cap A_2 \cap \bar{A}_3) + P(A_1 \cap \bar{A}_2 \cap A_3) + P(\bar{A}_1 \cap A_2 \cap A_3)$

$= P(A_1) P(A_2 | A_1) P(\bar{A}_3 | A_1 \cap A_2) + \cancel{P(A_1)}$

$\cancel{+ P(A_1}$

$+ P(A_1) P(\bar{A}_2 | A_1) P(A_3 | A_1 \cap \bar{A}_2) +$

$P(\bar{A}_1) P(A_2 | \bar{A}_1) P(A_3 | \bar{A}_1 \cap A_2)$

$= \frac{1}{2} \times \frac{4}{9} \times \frac{5}{8} + \frac{1}{2} \times \frac{5}{9} \times \frac{4}{8} + \frac{1}{2} \times \frac{5}{9} \times \frac{4}{8}$

$= 0.41667$

Or

$= \frac{3 \cdot 5 \cdot 5 \cdot 4 \cdot 2}{10 \cdot 9 \cdot 8 \cdot 2} = \frac{5 \cdot 4}{1 \cdot 2} \times \frac{5}{1} \times \frac{1}{\frac{(10 \cdot 9 \cdot 8)}{(1 \cdot 2 \cdot 3)}}$

$= {}^5C_2 \times {}^5C_1 \times 1 / {}^{10}C_3$

$$= \frac{(5C_2)(5C_1)}{10C_3}$$

$$\left.\frac{MC_x \cdot (N-M)C_{n-x}}{NC_x}\right\} \text{ hypergeometric distribution}$$

A random variable X is said to follow hypergeometric Distribution with parameters $N, M, n : x$

$$X \sim H(N, M, n : x)$$

It is assumes only positive integer (non negative integer values) & its probability mass function is given by

$$\frac{MC_x \, (N-M)C_{n-x}}{NC_n}$$

where $t \in T$ $\quad$ Range $(X) = T : \{\max(0, n-(N-M)) \ldots \min(n, M)\}$

where N is the number of elements in the population
M is the no. of elements in the population labelled success.
m is the no. of trials ~~per~~ or sample

eg $N = 100$
$M = 50$
$n = 20$ $\quad$ Find range:

$$\max(0, 20-(100-50)) - \min(20, 50)$$
$$0 - 20$$

Date __/__/__
Page _____

eg $N = 100$

$M = 90$

$n = 20$

$$\text{Max}(0, 20 - (100 - 90)) - \text{Min}(20, 90)$$
$$= 10 - 20$$

Conditions:

(i) There are finite number of dependent trials

(ii) Every trial results in 2 possible outcomes success and failure

(iii) Probability of success & failure is not same in each trial because the sampling is done without replacement

Mean & variance of hypergeometric

$\mu = np$

$V = npq$

$$\text{mean } (\mu) = np = n\left[\frac{M}{N}\right]$$

$$\text{Variance } (V \text{ or } \sigma^2) = n\left[\frac{M}{N}\right]\left[1 - \frac{M}{N}\right]\left[\frac{N-n}{N-1}\right]$$

finite population correction

(1) $n << N$, $N \to \infty$

$$\frac{N-x}{N-1} \underset{N \to \infty}{=} \frac{1 - n/N}{1 - 1/N} = 1$$

$$V = n\left(1 - \frac{M}{N}\right)\left(\frac{M}{N}\right)$$

(i) Note: If population size $N \to \infty$ in such a way that the proportion of success $(M/N) \to p$ then the hypergeometric probability mass function approaches binomial mass function

$$H(N, M, n : x) \to B(n, p : x)$$

$$N >> 20x$$

(ii) If the sampling were done with replacement then x would be binomial random variable & its variance is npq

Q. A ~~jury~~ jury of 5 members is drawn at random from a ~~ballots~~ voter's list of 100 persons out of which 60 are non graduates and 40 are graduates. What is the probability that the jury will consists of 3 graduates

Sol.

$X \to$ no. of graduates in a jury

$N = 100$

$M = 40$ $\qquad X \sim H(N, M, n : x)$

$n = 5$

$N - M = 60$

$$H(100, 40, 5 : x) = \frac{{}^{40}C_x \times {}^{60}C_{5-x}}{{}^{100}C_5}$$

$$H(100, 40, 5 : 3) = \frac{{}^{40}C_3 \times {}^{60}C_2}{{}^{100}C_5}$$

$$= 0.23227$$

Date ___/___/___
Page _____

Q A committee of 2 is chosen from 5 faculty members out of which 3 are doctrates & 2 are post graduates. If X is the number of doctrates in the committee. Obtain the probability

X = no. of ~~candidate~~ doctrates

$X \sim H(N, M, n : x) = H(5, 3, 2 : x)$

$H(N, M, n : x) = \dfrac{{}^{3}C_{x} \; {}^{2}C_{2-x}}{{}^{5}C_{2}}$

x	0	1	2
$p(x)$	0.1	0.6	0.3

Q TV's are shipped in a lots of 50. A shipment is accepted if a sample of 5 TVs inspected from a lot which does not contain any defective TV. If 1 or more are found defective the entire shipment is rejected. If lot of 50 contains 3 defective TV's. Determine the probability that 100% inspection is required

X → no. of defective TVs

N → 50 M = 3

n → 5 $x = 1, 2, 3 \ldots 5$

≥ 1

$X \sim H(N, M, n : x) = H(50, 3, 5 : x)$

$P(100\% \text{ inspection}) = P(X \geq 1)$

$= 1 - P(X = 0)$

$= 1 - \left[\dfrac{{}^{3}C_{0} \; {}^{47}C_{5-x}}{{}^{50}C_{5}}\right]_{x=0}$

$= 0.2760$

Date __/__/__
Page ____

Q A box contains 12 red marbles and 8 black marbles. If 5 marbles are drawn successively from the box. Find the probability

(i) 3 red 2 black

(ii) Atleast 3 red

(iii) All 5 marbles are of same color

Sol (i) $N = 20$ $\quad$ X - either red marble or black marble

red $\quad$ black

$x = 3$ $\quad$ $x = 2$

$n = 5$ $\quad$ $n = 5$

$M = 12$ $\quad$ $M = 8$

$$\frac{{}^{12}C_3 \times {}^{8}C_2}{{}^{20}C_5} \text{ or } \frac{{}^{8}C_2 \times {}^{12}C_3}{{}^{20}C_5}$$

$= \cancel{0.1578}\ 0.3973$

(ii) $P(x \geq 3)$

$= 1 - P(X=0) - P(X=1) - P(X=2)$

$= 0.70381$

(iii) P(all same color)

$= P(5R) + P(5B)$

$$= \frac{{}^{12}C_5\ {}^{8}C_0}{{}^{20}C_5} + \frac{{}^{12}C_0\ {}^{8}C_5}{{}^{20}C_5}$$

$= 0.054$

Date __/__/__
Page _____

Q (i) Determine the probability that exactly three computers are defective out of 10 computers purchased from a lot of 5000 computer containing 1000 defective computers using hypergeometric distribution. (ii) Use binomial approximation & explain the accuracy (iii) What is the finite population correction factor (iv) calculate the mean and standard deviation for the sample.

Ans (i) $X \rightarrow$ no. of defective computers

$x = 3$

$M = 1000$ $\quad n = 10$

$N = 5000$

$$= \frac{{}^{1000}C_3 \times {}^{4000}C_7}{{}^{5000}C_{10}}$$

$$= 0.20147$$

(ii) $${}^{10}C_3 \times \left(\frac{1000}{5000}\right)^3 \times \left(\frac{4000}{5000}\right)^7$$

$$= \cancel{0.6144} \; 0.20132$$

(iii) $$\frac{N-n}{N-1} = 0.9981$$

(iv) mean

$$\mu = n\left[\frac{M}{N}\right] = 2$$

Std deviation $$= \sqrt{n\left[\frac{M}{N}\right]\left[1-\frac{M}{N}\right]\left[\frac{N-n}{N-1}\right]}$$

$$= \sqrt{\cancel{10}\; 2 \times 0.9981 \times \frac{4}{5}}$$

$$= 1.263$$

Discrete Uniform distribution

$X \Rightarrow$ finite distinct outcomes

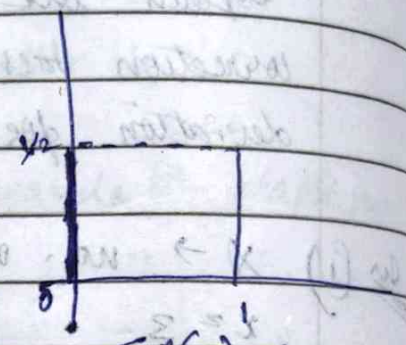
$\hookrightarrow$ probability $\rightarrow$ equally likely events

Tossing a coin $= \{H, T\}$

$\rightarrow P(H) = 1/2$

$P(T) = 1/2$

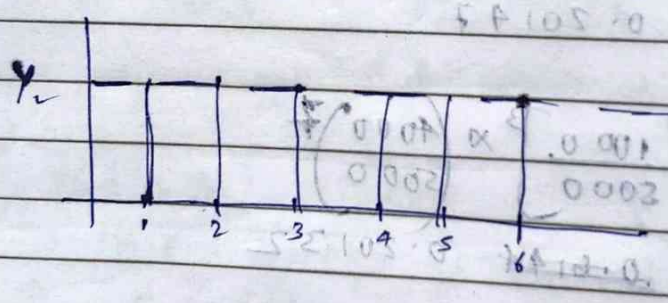
$$P(x) = \frac{1}{n}$$



$\sum P(x_i) = 1$

Throwing a dice $= \{1, 2, 3, 4, 5, 6\}$

$P(x_i) = 1/6 \quad i = 1, \ldots, 6$



A discrete uniform distribution is a discrete probability distribution. A random variable X follows discrete uniform distribution if there are number of distinct outcomes & each outcome is equally likely.

Date ___/___/___
Page _____

	Range	Pmf $(P(x))$
(1)	$1, 2, \ldots, 12$	$1/n$
(2)	$a, a+1, a+2, \ldots, b$	$1/(b-a+1)$ $\quad (a, a+1, \ldots, b = a+n)$

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

$b = a+(n-1)d$

$b = a+(n-1)$

$b-a+1 = n$

(3) $a, a+d, a+2d, \ldots, b$ $= \dfrac{1}{\left(\dfrac{b-a+d}{d}\right)} = \dfrac{d}{b-a+d}$

---

mean & variance of uniform distribution

case 1

$$\text{Mean}(\mu) = \sum_{1}^{n} x P(x)$$

$$= \sum_{1}^{n} x \frac{1}{n}$$

$$= \frac{1}{n} \sum_{1}^{n} x$$

$$= \frac{1}{n}\left[1+2+3+\ldots+n\right]$$

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

$$\boxed{\mu = \frac{n+1}{2}}$$

case 2:

$\cancel{\mu = \frac{(b-a+1)+1}{2}}$ $\quad \mu = ?$

$$\sum_{a}^{b} x P(x)$$

$$= \sum_{a}^{a+n} x \frac{1}{b-a+1}$$

$$= \frac{1}{b-a+1} \sum_{a}^{a+n} = \frac{1}{b-a+1}\left[\sum_{1}^{b} x - \sum_{1}^{a-1} x\right]$$

$$= \frac{1}{b-a+1}\left[\frac{b(b+1)}{2} - \frac{(a-1)(a-1+1)}{2}\right]$$

Date ___/___/___
Page ___

$$= \frac{b^2 + b - a^2 + a}{2(b-a+1)}$$

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

$$= \frac{(b+a)\cancel{(b-a+1)}}{2\cancel{(b-a+1)}}$$

case 2:

$$\boxed{\mu = \frac{b+a}{2}}$$

case 1

variance $(v) = \sum_{1}^{n} (x-\mu)^2 p(x)$

$$= \sum_{1}^{n} x^2 p(x) - \mu^2$$

$$= \sum_{1}^{n} x^2 \left(\frac{1}{n}\right) - \mu^2$$

$$= \frac{1}{n} \sum_{1}^{n} x^2 - \mu^2$$

$$= \frac{1}{n}\left[1^2 + 2^2 + \cdots + n^2\right] - \mu^2$$

$$= \frac{1}{n}\left[\frac{n(n+1)(2n+1)}{6}\right] - \frac{(n+1)^2}{4}$$

$$= \frac{n+1}{2}\left[\frac{2n+1}{3} - \frac{n+1}{2}\right]$$

$$= \frac{n+1}{2}\left[\frac{2(2n+1) - 3(n+1)}{6}\right]$$

$$= \frac{n+1}{2}\left[\frac{4n+2-3n-3}{6}\right]$$

$$= \frac{(n+1)(n-1)}{12}$$

Date ___/___/___
Page _____

$$= \frac{n^2-1}{12}$$

Case 2:

$$V = \sum_{a}^{b} (x-\mu)^2 p(x)$$

$$= \sum_{1}^{b} (x-\mu)^2 p(x) - \sum_{1}^{a-1} (x-\mu)^2 p(x)$$

$$= \sum_{1}^{b} x^2 p(x) - \mu^2 - \left(\sum_{1}^{a-1} x^2 p(x) - \mu^2 \right)$$

$$= \sum_{1}^{b} x^2 p(x) - \sum_{1}^{a-1} x^2 p(x)$$

$$\sum_{1}^{b} x^2 \frac{1}{(b-a+1)} - \sum_{1}^{a-1} \frac{x^2}{b-a+1}$$

$$= \frac{1}{b-a+1} \left(\sum_{1}^{b} x^2 - \sum_{1}^{a-1} x^2 \right)$$

$$\frac{1}{b-a+1} \left(\left[\frac{b(b+1)(2b+1)}{6} \right] - \left[\frac{(a-1)(a)(2a-1)}{6} \right] \right)$$

$$= \frac{1}{6(b-a+1)} (b^2+b)(2b+1) - (a^2-a)(2a-1)$$

$$\frac{1}{6(b-a+1)} \left[2b^3 + b^2 + 2b^2 + b - (2a^3 - a^2 - 2a^2 + a) \right]$$

$$\frac{1}{6(b-a+1)} \quad 2b^3 + b^2 + 2b^2 + b - 2a^3 + 3a^2 - a$$

$$\frac{1}{6(b-a+1)} \quad 2b^3 + 3b^2 + b - 2a^3 + 3a^2 - a$$

$$\frac{1}{6(b-a+1)} \quad 2(b^3 - a^3) + (b-a) + 3(b^2 + a^2)$$

$$\frac{1}{6(b-a+1)} \quad b-a$$

Date ___/___/___
Page _____

case 2: $V = \sum_{a}^{b} (x-\mu)^2 p(x)$

$$= \sum_{1}^{b} (x-\mu)^2 p(x) - \sum_{1}^{a-1} (x-\mu)^2 p(x)$$

$$= \frac{b^2-1}{12} - \left(\frac{(a-1)^2-1}{12}\right)$$

$$= \frac{1}{12}\left[b^2-a^2+2a-1+\cancel{1}-\cancel{1}\right]$$

$$= \frac{1}{12}\left(b^2+2a-1-a^2\right)$$

$$= \boxed{\frac{1}{12}\left[(b-a+1)^2-1\right]}$$

case 3: $V = \frac{d^2-1}{12}\left[\left(\frac{b-a+d}{d}\right)^2 - 1\right]$

$$\boxed{V = \frac{(b-a+d)^2-d^2}{12}}$$

$V(ax) = a^2 V(x)$ - $\sigma(ax) = a\,\sigma(x)$

$E(ax) = aE(x)$

Date __/__/__
Page _____

Q An employee is selected from a staff of 10 to supervise a certain project by selecting a tag at random from a box containing 10 tags numbered from 1 to 10. What is the probability that the number drawn is less than 4.

Ans $X = \{1, 2, \ldots, 10\}$

$X \sim U(1, 10)$, $P(X) = \frac{1}{n} = \frac{1}{10}$

$P(\text{less than } 4) = P(X < 4)$

$= P(X \leq 3)$

$= P(X=1) + P(X=2) + P(X=3)$

$= \frac{1}{10} + \frac{1}{10} + \frac{1}{10} = 0.3$

Q Find the mean & variance for the same problem

$\text{mean} = \frac{n+1}{2}$

$= \frac{11}{2} = 5.5$

$\text{Variance} = \frac{n^2-1}{12} = \frac{100-1}{12}$

$= \frac{99}{12} = \cancel{48.5}\ 8.25$

Q Suppose X has a discrete uniform distribution ~~from~~ on the integers 0 to 9. Determine the mean & std. deviation ~~from~~ of ~~X~~ random variable 5X

$X \sim U(0, 9)$ $X = \{0, 1, 2, 3 \ldots 9\}$

$P(X) = \frac{1}{b-a+1} = \frac{1}{9-0+1} = \frac{1}{10} = 0.1$

$E(5X) = ?$ $SD(5X) = ?$

Page

$E(5X) = \mu(5X) = 22.5$

$E(5X) = 5(E(X))$ $\qquad \mu = \frac{a+b}{2}$

$= 5\left[\frac{0+9}{2}\right]$

$= 22.5$

$SD(X) = \sqrt{V(5X)} = \sqrt{5^2\, V(X)}$

$= \sqrt{25 \times \frac{1}{12}\left((9-0+1)^2 - 1\right)}$

$= \sqrt{25 \times \frac{99}{12}}$

$= 14.36$

Q The thickness measurement of coating are made to the nearest hundred of a millimeter. The thickness measurements are uniformly distributed from the values $\{0.15, 0.16, \ldots, 0.19\}$. Determine the mean and standard deviation used for this process.