Statistics

Numerical facts — Quantity

sample mean :- $\bar{x}$ → Statistics

population mean :- μ → Parameters

* difference between $\bar{x}$ and μ is called error.
* Branches of statistics

Descriptive Statistics
- deals with summarisation methods
  eg. average
  central tendency
  graphs
  variations

Inferential Statistics
- infer about the population based on sample
  infer :- draw conclusions
  1. hypothesis testing.
  2. Estimation of population parameters

generalization :- for those which cannot be approached.

* variables :- quantity which varies

quantitative
- can be expressed numerically

qualitative
- cannot be expressed numerically.

discrete variable
- countable

continuous variable
- measurable

Dalâ Types

1. Primary Dalā
• raw / initial dalā not under any statistical ~~data~~ method

2. Secondary Dalā
• organized dalā
• undergone any procedure.

Measurement Scales:-

1. Nominal - no numerical significance eg. gender
2. Ordinal - preferred relation Positions 1 > 2 > 3 > 4
3. Interval
4. Ratio

nominal and ordinal → Qualitative variable
Interval and ratio → Quantitative variable
nominal ← no preference / no ranking. used for group identification.

ratio → zero is treated as 0 → absence.
Interval → zero has a meaning, eg. temp. is 0 - (very cold)

Time Series → Dalā associated with time
eg. oil consumption in Pakistan in last 10 years. (time varies)

Cross Sectional → in one time, more no. of dalā categories collected.
eg. ~~in~~ ~~once~~ on Monday, a patients BP, sugar, etc. is all collected
more than one variable's dalā is collected

Multi variate Time Series.
i- * more ~~areas~~ years, and more variables

Ungrouped Data:-
- raw data, not classified.

Measure of Central Tendency
- mean $\Rightarrow \Sigma X / N \rightarrow$ sample size / pop size.
- median $\rightarrow$ middle value
- mode $\rightarrow$ most repeated value.

if $\frac{n}{2}$ / integer (even) $\rightarrow$ average.
\ (odd) $\rightarrow$ middle value $(\frac{n}{2}+1)^{th}$ value

Trimmed Mean:- Sorted Data.
- reduce / trim values from start/end.
- eliminates outliers.

Ex: 1.1
Drying Time

3.4	2.5	4.8	2.9	3.6
2.8	3.3	5.6	3.7	2.8
4.4	4.0	5.2	3.0	4.8

a) sample size :- 15
b) mean :- 3.787
c) median :- 3.6
d) trimmed mean :- 20% :- 3.678
e) conclude result :- the drying time of the particular brand is almost 4 hrs.
50% drying time is less than 3.6 hrs.
no extremes / outliers.

Scanned with CamScanner

Percentiles and Quartiles

Median => 2 parts of data
quartiles => 4 parts of data.
percentiles => 100 parts of data

i) Sort data in ascending order.
ii) calculate the index

$$i = \left(\frac{P}{100}\right) \times n$$

iii) a) index is an integer then p^{th} percentile is the average of the no's on i^{th} and $i+1^{th}$ index
b) if index is not an integer, round up the value it will be the position of the p^{th} percentile

Q4 monthly salaries => 12 graduates

a) calculate 85th percentile and 50th. Also evaluate lower and upper quartile.

1	3450	3
2	3550	9
3	3650	10
4	3480	5
5	3355	2
6	3310	1
7	3490	6
8	3730	11
9	3540	8
10	3925	12
11	3520	7
12	3480	4

- 85th total 10.2 position:- 3730
- 50th :- 3505
- 25% = ~~3480~~ 3465
- 75% = ~~3550~~ 3600

→ less than 85% graduates are earning 3730 → below.

%/ Percentile range => 1-99.
Quartile range => 1-3

Measures of Dispersion

- variance
- Standard deviation
- IQ range
- coefficient of variation

- $s^2 = \frac{\Sigma(X-\bar{x})^2}{n}$ (sum of differences)^2.

Sample variance => $s^2 = \frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2$ / $\frac{\Sigma(x-\bar{x})^2}{n-1}$

$\sigma^2 = \frac{\Sigma(X-\mu)^2}{N}$ population size.

- $SD = \sqrt{S}$
- IQ range = $Q_3 - Q_1$ (middle half is spread on IQR values)

IQR describes the dispersion of the middle half data

Range → complete data dispersion.

- $cv = \frac{s}{\bar{X}} \times 100$. → sample cv.

$cv = \frac{\sigma}{\mu} \times 100$ → pop. cv.

(relative measure of dispersion)

menahil
22L-7459

Q4 Price and life in hours

X	Price	Y	Life
8		130	
13		150	
18		180	
23		250	
30		345.	

a) calculate coefficient of variation

$CV = \frac{S}{\bar{X}} \times 100.$

$S^2_1 = \frac{1986}{5} - \frac{8464}{25}$

$S_1^2 = 58.64$

$S_1 = 7.66$

$\bar{X}_1 = 18.4$

$CV = 41.63$

$S_2^2 = 50665 - 211$

$= 50454$

$S_2 = 78.38$

$X_2 = 211$

$CV = 37.14$

→ on the basis of CV it was found that the relative ~~relation~~ variation in price in Rs data set is higher than the ~~life~~ second data set.

Practice Q4 ~~Ex~~ EX. 1.3
1.17, 1.21

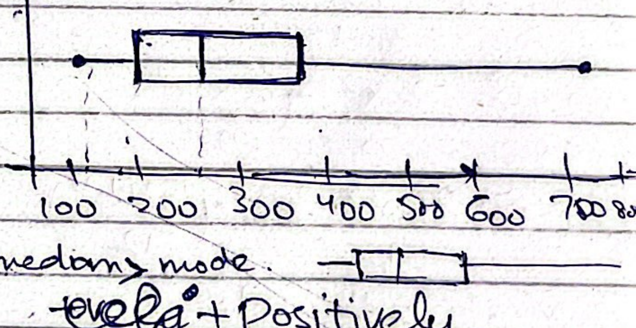
5. point summary
Min, Q_1, Median, Q_3, Max
→ collectively called 5 point summary of a data set.

Box Plot (Box and whisker diagram)
Q_4 Time in seconds → 12 vehicals.
t/s

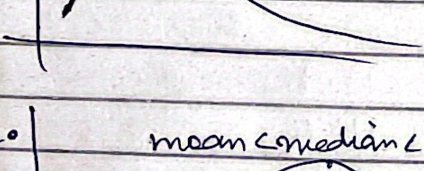
145	165	260	330	280	195
375	480	180	180	420	750

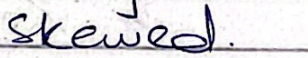
105	145	150	180	195	250	260.
330	375	420	480	750		

Min = 105
Max = 750
Q_1 = 165
Q_2 = 255.
Q_3 = 397.5



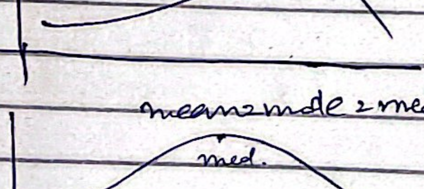
1. mean > median > mode.

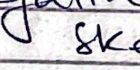




+ positively skewed.

2. mean < median < mode





• – negatively skewed.

3. mean = mode = median

• symmetrical

FASTRU

Limit / Fences Calculation
Lower limit :- $Q_1 - 1.5(IQR)$
Upper limit :- $Q_3 + 1.5(IQR)$

If above U/L or below L/L
then value is an outlier.

- outlier detection
- data distribution

Probability
Practice Qs

$\Sigma x = 1.3$

1.17

Smokers

a) mean :- $524.4 / 12$
$= 43.7$

b) $SD^2 = \frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2$

$= \frac{26068.32}{12} - \left(\frac{524.4}{12}\right)^2$

$v = 262.67$

$S.D = 16.21$

min = 13.8
Max = 69.3
Q_1 = 28.8
Q_2 = 47.85
Q_3 = 54.6

nonsmokers

$= 454.8 / 15$
$= 30.32$

$SD^2 = \frac{14500.82}{15} - \left(\frac{454.8}{15}\right)^2$

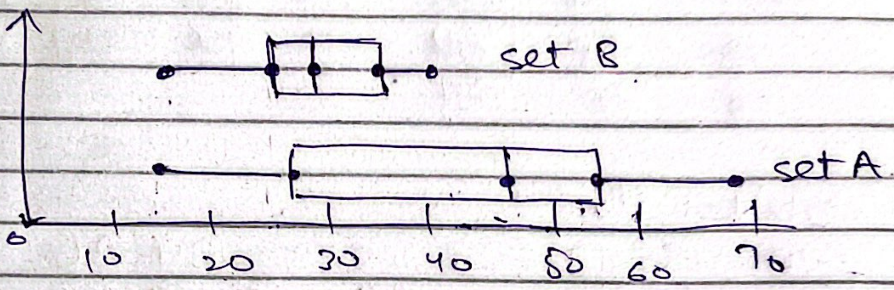
$v = 47.42$ 919.30

$SD = 7.13$

13.9
41.6
26.4
30.2
36

set A 13.8, 22.1, 23.2, 34.4, 43.8, 47.6, 48.1, 52.7, 53.2, 56, 60.2, 69.3

set B
13.9, 21.1, 25.1, 26.4, 28.4, 28.6, 29.8, 30.2, 30.6, 31.8, 34.9, 36, 37.9, 38.5, 41.6

c)


d) set A => negatively skewed, shows that smokers tend to take longer to sleep than non-smokers.

SCORES

Q4 - ~~23~~ ~~60~~ ~~79~~ ~~32~~ ~~57~~ ~~74~~ ~~52~~
~~36~~ ~~80~~ ~~77~~ ~~81~~ ~~95~~ ~~41~~ ~~65~~
~~55~~ ~~76~~ ~~52~~ ~~10~~ ~~64~~ ~~75~~ ~~78~~
~~98~~ ~~81~~ ~~67~~ ~~41~~ ~~71~~ ~~83~~ ~~54~~
~~88~~ ~~62~~ ~~74~~ ~~43~~ ~~60~~ ~~78~~ ~~89~~
~~48~~ ~~84~~ ~~90~~ ~~15~~ ~~79~~ ~~34~~ ~~62~~
~~69~~ ~~74~~ ~~63~~ ~~80~~ ~~85~~ ~~61~~ ~~70~~
~~72~~ ~~25~~ ~~64~~ ~~76~~ ~~17~~ ~~82~~ ~~85~~
~~80~~ ~~72~~ ~~84~~ ~~82~~

Sorted Data:

10 15 17 23 25 32 34 36 41
41 43 48 52 52 54 55 57 60 60
61 62 63 64 64 65 67 67 69
70 71 72 74 74 74 75 76 76
77 78 78 79 79 80 80 80 81
81 82 82 83 84 84 85 85 88
89 90 92 95 98

construction of frequency distribution

1) Range :- 98 − 10 = 88

2) $K = 1 + 3.3\log(N)$ ≈ 6.87 ≈ 7
no. of classes

3) Interval / class size
h = Range/K ~ 88/7 = 12.57 ≈ 13

Scanned with CamScanner

class limits

limits	tally	f	Rf	%f	cf
10 − 22	|||	3	0.05	5	3
23 − 35	||||	4	0.067	6.7	7
36 − 48	𝍸	5	0.0833	8.33	12
49 − 61	𝍸 |||	8	0.133	13.3	20
62 − 74	𝍸 𝍸 ||||	14	0.233	23.3	34
75 − 87	𝍸 ×4	20	0.33	33.3	54
88 − 100	𝍸 |	6	0.1	10	60

ungrouped data

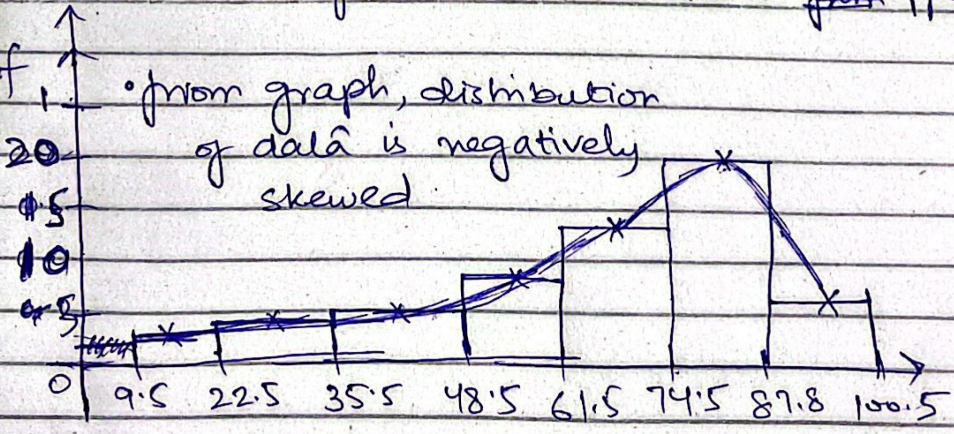
mean :- $\frac{\Sigma fx}{\Sigma f}$ $\quad$ x => midpoint of class limits

eg :- $\frac{10+22}{2} = \frac{32}{2} = 16$

$= \frac{(16\times3)+(4\times29)+ \ldots}{60} = \ldots \frac{3950}{60} = 65.83$

standard deviation $= \sqrt{\frac{\Sigma f(x)^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2} = 20.86$

Histogram :- define class boundaries.
−0.5 from lower and +0.5 in upper.



• from graph, distribution of data is negatively skewed

Graphical Representation

* Bar chart
* Pie chart
* Dot Plot

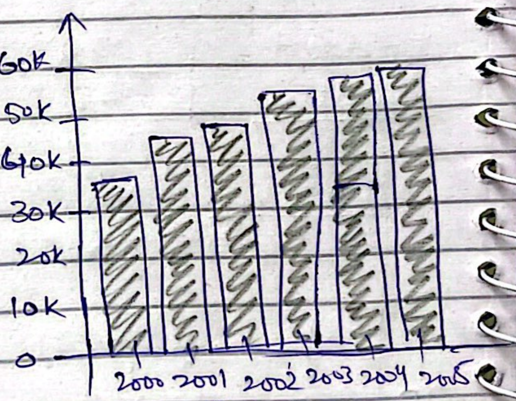
Q. eg. no of children → discrete
0–5 (50 values)
range is very less
⇒ no need of classes

x	tally	f
0		
1		
2		
3		
4		
5		

Q. Draw a bar diagram to represent the turnover of a company for 6 years.

Years	Turnover
2000	38,000
2001	45,000
2002	48,000
2003	52,500
2004	55,000
2005	58,000.

Bar chart



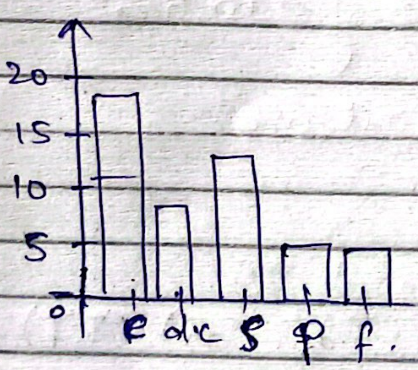
⇒ gradual increase
⇒ discuss very high / low values

component barchart ⇒ parts of bars

Q. Table shows the frequency distribution of soft drink purchases

	Soft Drink	frequency
1.	Coke	19
2.	Diet coke	8
3.	pepsi	5
4.	sprite	13
5.	Fanta	5
		50

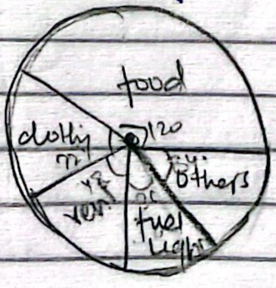
⇒ Show by an appropriate graph, which brand is most preffered.



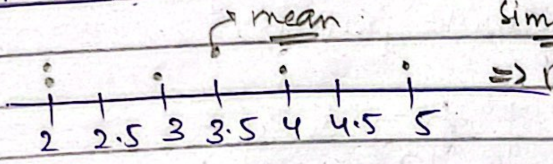
Pie chart / Sector / pizza slice diagram
component barchart = piechart

Q. Total expenditure and expense

	items	expense
120°	food	50,000
72'	clothing	30,000
48'	houserent	20,000
36'	fuel+light	15,000
84'	others	35,000
		150,000

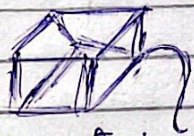


Dot Plot



simplest
=> plot data directly

Probability



Subjective → prior knowledge
objective → based on recorded observations

Sample Space — S
set of possible outcomes of statistical experiment
element :- each possible outcome.
sample point / member.

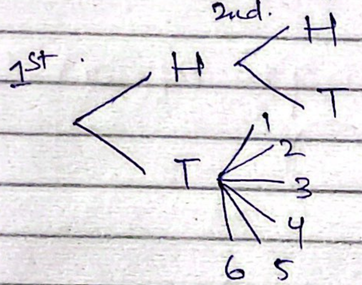
~~Possible~~ if 2 ~~dice~~ coins tossed.
$S = (2)^2 = 4$ (H,H), (H,T), (T,H), (T,T)

Dice Rolled.
$S_1 = \{1, 2, 3, 4, 5, 6\}$
if in even odd
$S_2 = \{\text{even}, \text{odd}\}$

Tree Diagram

Q4 Coin toss — head - coin toss again / tail - roll a dice.

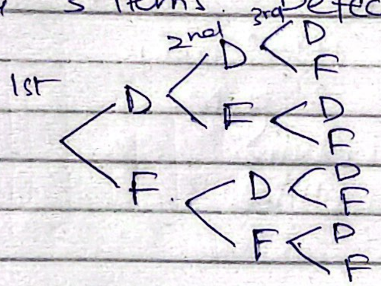
$S = \{H, H, HT, T1, T2, T3, T4, T5, T6\}$



HH
HT
T1
T2
T3
T4
T5
T6

Points collected

Q4 3 items Defected / Not defected



DDD, DDF,
DFD, DFF,
FDD, FDF,
~~FDDF~~ FFD, FFF

Sample space.

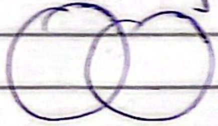
Events :- (Subset)
eg. create event which has >1 Defective set.
$B = \{DDD, DDF, DFD, FDD\}$
=> occurence of certain events
=> subset of sample set.

Compliment :- Shows absense (negation)
S = all cards (52) R = red cards (26)
R' = black card (26)

Intersection:- two events A & B, event containing all elements common in A & B

no intersection (not occur together)
→ multually exclusive / disjoint.

Union:- all elements belong to A/B or both.
Venn diagram



etc.

Counting Sample Points:
→ multiplication rule

total ways = $n1 \times n2$
total ways of n1   total ways of n_2

Q. ext style = 4 possibilities.
floor plan = 3 possiblities.

total ways = $4 \times 3 = 12$ ways

Q. 22 members — chair person (22 possibilities)
— treasurer (21 possiblities)

$= 22 \times 21 = 462$ ways

Permutations :- find possible arrangements.
arrange of all / part of set of objects
Form ①
* no. of permutations of n objects is $n!$

a, b, c, d.
if two pair = 4×3 | 4 | 3 |
=> 12 combo's possibilities

Form ② selection.

$nPr = \frac{n!}{(n-r)!}$ (arranges selected objects)

→ distinct objects.

Combinations :- ~~combina~~ distinction not possible.

Q. 25 candidates → 3 awards given. distinguishable
possible selections?
=> $25P_3$ = 13,800 or $25 \times 24 \times 23$ = 13,800

③ form.
Arrangements in a circle :- $(n-1)!$

Form ④
Distinct permutation in 'n' things
but grouped (eg. gpa ≥ 1) G1 $n1$
(gpa ≥ 2) G2 $n2$
⋮
nK

$= \frac{n!}{n1!\,n2!\,n3! \ldots nK!}$

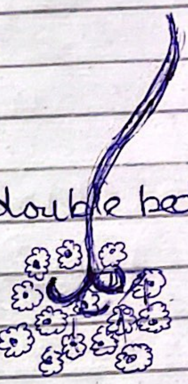
eg. $n=10$ — 1 freshman, 2 soph., 4 jun., 3 sen.

in a row arrangements?

$$= \frac{10!}{1!\,2!\,3!\,4!} = 12,600$$

Q. 7 students: 1 triple bed 2 double bed.

$$= \frac{7!}{3! \times 2! \times 2!} = 210$$



Combination :- order doesn't matter

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

Q. Take from 5 cards from 52 cards

$\Rightarrow$ order doesn't matter $\underline{^{52}C_5}$

Q. 10 arcade, 5 sport $\rightarrow$ how can she get total 5: 3 arcade and 2 sports?

$^{10}C_3 = 120$ & $^{5}C_2 = 10$

$120 \times 10 = \underline{1200}$ possibilites

Q. STATISTICS $\rightarrow$ how many arrangements made?

$$= \frac{10!}{3! \times 3! \times 2!} = \underline{50,400}$$

(object of similar kind case) = S, I, T groups.

Q4 4 married couples have bought 8 seats in the same row for a concert. In how many different ways can they be seated?

a) no restrictions
$= 8! = 40320$

b) if each couple sits together
$= 4! = 24 \times 2 \times 2 \times 2 \times 2$ 2 2 2 2
$= 48 \times 8$ ⇒ external and internal arrangements
$= 384$

Q4 $\underline{2.37}$ $\underline{2.32}$ $\underline{2.41 - 2.47}$ $\underline{2.21 - 2.25}$
$\underline{2.37}$ $\underline{2.30}$ $\underline{2.27}$ (eg)

Q4 4 Boys 5 girls
G B G B G B G B G
| 5 | 4 | 4 | 3 | 3 | 2 | 2 | 1 | 1 | $= 2880$

$1! \times 5! \times 4! = 2880$
(no external arrangement)

Menahil

Probability of an Event

* Classical Approach
* Relative Frequency Approach
* Axiomatic Approach

classical:- $\frac{n}{N} = \frac{\text{favourable outcomes}}{\text{total outcomes}}$

⇒ should be equally likely outcomes
(eg H/T), (eg. dice all 1/6 chance)

Rf approach:- repeat experiment and ratio is answer.

$$\lim_{n \to \infty} \left(\frac{n}{N}\right)$$

Axiomatic approach:-
Rules. (assign weights and add)

$P(\text{sure}) = 1$
$P(\text{impossible}) = 0$
$\therefore 0 < P < 1$

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

Q. A coin is tossed twice. What is the P, that at least one head occurs.

→ $S = \{TT, TH, HT, HH\} = \frac{3}{4} = \frac{n(A)}{n(S)}$
$A = \{HT, TH, HH\}$

→ through axiomatic. $4 \times w = 1$
$w = 1/4$

$3 \times \left(\frac{1}{4}\right) = \frac{3}{4}$

Q4 A dice is ~~loaded~~ rolled in such a way that that an even no. is twice as likely to occur as an odd no. If E is the event that a no. less than 4 occurs on a single toss of the die, then find probability of E.

$6/3=2$

~~$4w+2w=1$~~

~~$6w=1$~~

$2w$

$S = \{1, 2, 3, 4, 5, 6\}$

$w \quad 2w \quad w \quad 2w \quad w \quad 2w$

$3(2w) + 3(w) = 1$

$9w = 1$

$w = 1/9$

$E = \{1, 2, 3\}$

$E = \frac{1}{9} + \frac{2}{9} + \frac{1}{9} = \frac{4}{9}$

Q4 Cards

52

13 13 13 13

red: Diamond OR heart

black: spade OR club

13:- ace, 2, 3 —— 10, king, queen, jack

Q4 In a poker hand, consisting of 5 cards. find the probability of holding 2 aces and 3 jacks.

favourable $= {}^4C_2 \times {}^4C_3 = 24$

any 5 total $= {}^{52}C_5 = 2598960$

$$P = \frac{1}{108290}$$

Additive Rules

If A and B are two events then
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

mutually exclusive = A & B cannot occur together
$P(A \cup B) = P(A) + P(B)$

Theorems

Menahil Amir
22L-7459

Q4 Company A = 0.8 } not mutually
Company B = 0.6 } exclusive.
Both A & B = 0.5 }
P that he will get at least one offer?
⇒ $(0.8+0.6)-0.5=0.9$

Q4 What is the P of getting a total of 7 or 11 when a pair of dice is rolled
$7 \Rightarrow (1,6)(2,5)(3,4)(4,3)(5,2)(6,1)$
$P(7)=\frac{6}{36}=\frac{1}{6}$
$11 \Rightarrow (5,6)(6,5)=\frac{2}{36}=\frac{1}{18}$ } mutually exclusive

$P(A \cup B)=P(A)+P(B)$
$=\frac{1}{6}+\frac{1}{18}=\frac{4}{18}=\frac{2}{9}$

Q4 $P(G)=0.09$ $P(W)=0.15$ $P(R)=0.21$ $P(B)=0.23$
What is P that buyer will buy this colour automobile.
$P=0.09+0.15+0.21+0.23=0.68$

Scanned with CamScanner

Complimentation

$P(A) + P(A') = 1$

Q4. P cars # 3 4 5 6 7 8

0.12 0.19 0.28 0.24 0.10 0.07.

(P) that mechanic services atleast 5 cars.

$P = 0.28 + 0.24 + 0.1 + 0.07 = 0.69$

or $P = 1 - (0.12 + 0.19) = 0.69$

2. 53, 54, 56, 69, 65.

Conditional Probability

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

Formula :

Given A → already occured

find conditional P of B

$P(B|A) \Rightarrow$ given that

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

in presence of other.

Q.

	empl.	unempl.	Total
male	460	40	500
female	140	260	400
	600	300	900

⇒ out of 900, what are chances he is male but given that he is employed

$P(M|E) = \frac{460}{600} = \frac{23}{30}$

Combined Probability Table

460/900	40/900	500/900
140/900	260/900	400/900
600/900	300/900	900/900

Q. $P(D) = 0.83$ → departs on time.

$P(A) = 0.82$ → arrives on time.

$P(D \cap A) = 0.78$ → depart and arrival on time

a) $P(x)$ ⇒ that a plane arrives on time given that it departed on time.

$P(A|D)$

$P(x) = \frac{0.78}{0.83} = 0.94$

b) $P(x)$ ⇒ departed on time, given that it has arrived on time

$P(D|A)$

$P(x) = \frac{0.78}{0.82} = 0.95$

$$P(A \mid D') = \frac{P(A \cap D')}{P(D')} = \frac{0.82 - 0.78}{0.17} = 0.24$$

Q3. Sold :- 657,000 in 2009
2008 → volume down by 37%.

US sold :- 280,500.
2008 → volume down by 48%.

Data in 1000's ⇒ Crosstab

	Car	Light Truck	
US	87.4	193.1	280.5
Non US	228.5	148	376.5
	315.9	341.1	657

a) Construct joint probability table
b) marginal prob?
c) If ~~its~~ made in US. P that it was car. P it was light truck?
d) If not in US. P(car)? P(light veh)
e) If light truck. P(us)
f) what does the probability information tell ~~a~~ you about sales?

a)

	Car	L/T	
US	0.1330	0.2939	0.4269
non US	0.3470	0.2253	0.5731
	0.48	0.5192	1

- divide eug value by 657,000.
→ marginals

c) $P(C/US) = \frac{0.1330}{0.4269}$ $P(LT/US) = \frac{0.2939}{0.4269}$

e) P(US/C)
f) LT preffered more than cars.
US has less sales than non US.

Scanned with CamScanner

Independent Events:

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

=> occurence of B doesn't affect occurence of A

=> if $P(A|B) \neq P(A)$, occurence of B, has affect on occurence of A.

without replacement
=> total effected
=> total sample space effected

with ~~rep~~ replacement
=> ~~dono~~ doesn't affect total eachothers P.

① Ace. $P(A) = 4/52$
② Spade $P(S) = 13/52$

with rep.
$P(S/A) = \frac{13}{52} = P(S)$
doesn't effect.

w/o rep.
$P(S/A) = \frac{13}{51} \neq P(S)$.
P(A) effects P(S)

Multiplicative Law / Product Rule

① $P(A \cap B) = P(B) * P(A|B)$
$\therefore$ dependent events

② $P(A \cap B) = P(B) * P(A)$
$\therefore$ Independent events

③ $P(A_1 \cap A_2 \cap A_3 \cap \cdots A_K)$
$= P(A_1) * P(A_2/A_1) * P(A_3/A_2 \cap A_1) * \cdots P(A_K/A_1 \cap A_2 \cdots A_{K-1})$

$\therefore$ ② extended uptill K values.

Q. Suppose a fuse box has 20 fuses of which 5 are defective. 2 selected and removed from box. w/o replacing first. P that both fuses are defective.

through ① $\Rightarrow \frac{5}{20} \times \frac{4}{19} = \frac{20}{380} = \frac{1}{19}$

Q. 2.38

Q4 3 cards drawn in succession w/o replacement from 52 cards. Find P that the ~~vec~~ event:- $A_1 \cap A_2 \cap A_3$ occurs where A_1 is the event that the first card is a red ace. A_2 is event that second card is 10/Jack. A_3 is the event that third card is greater >3 and <7.

$A_1 = 2/52$

$A_2 = 8/51$

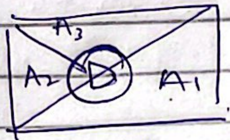
$A_3 = 12/50$

$$P(A_1 \cap A_2 \cap A_3) = \frac{\cancel{8}2}{52} \times \frac{8}{51} \times \frac{12}{50} =$$

Ex. 2.75, 2.77, 2.80, 2.83, 2.88, 2.89

Bayes' Theorum.

eg. semple :



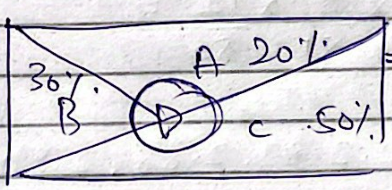
D is common event.

$P(D) = P(A_1)P(D|A_1) + P(A_2)P(D/A_2) + P(A_3)P(D/A_3)$

$P(A_i|D) = \dfrac{P(A_i \cap D)}{P(D)}$

$= \dfrac{P(A_i)P(D/A_i)}{P(D)}$

Q4 events



⇒ mutually exclusive

D = defective rate.

$P(D/A) = 0.01$, $P(D/B) = 0.02$, $P(D/C) = 0.04$ — conditional P.

$P(A) = 0.2$, $P(B) = 0.3$, $P(C) = 0.5$ — prior P

$P(A/D) = ?$ → Posterior P.

$= \dfrac{P(A \cap D)}{P(D)} = \dfrac{P(A) \times P(D/A)}{\cancel{0.01 + 0.02 + 0.04}}$

$= \dfrac{(0.2)(0.01)}{(0.2)(0.01) + (0.3)(0.02) + (0.5)(0.04)}$

pg #72 ⇒.

Scanned with CamScanner

Generalized

$$P(A_i/D) = \frac{P(A_i) \cdot P(D/A_i)}{\sum P(A_i) \cdot P(D/A_i)}$$

For Bayes Σ of individual events must be 100%.

$P(D/A) = 0.01$

$P(A/D) \Rightarrow$ posterior probability

Bayes Theorem

Events must be mutually exclusive and collectively exhaustive.

2.41 and 2.42

Pg # 72 → Baye's Theorum

2.101

Q4 A paint store purchases and sales latex and semi gloss paint. Based on the sales the P that a customer will purchase latex paint is 0.75 of those who purchase latex paint 60% also purchase rollers, but only 30% of the semi gloss paint buyers purchase rollers. A randomly selected buyer purchases a roller and a can of paint. what is the P that paint is latex

latex = 0.75 roller = 0.6 P(r/L)

semigloss = 0.25 roller = 0.3 P(r/S)

$$P(L|R) = \frac{P(L \cap R)}{P(R)}$$

$$= \frac{0.75 \times 0.6}{(0.75 \times 0.6) + (0.25 \times 0.3)} = \frac{0.45}{0.525}$$

$= 0.857$

Q 2.41, 2.42

Q4 2.99.

Q: Expiry ⇒ (4 members)

A = 20% ⇒ once in 200 packs fails

B = 60% ⇒ once in 100 packs fails

C = 15% ⇒ once in every 90 fails

D = 5% ⇒ once in every 200 miss

$P(A/M)$

$P(A) = 0.2$ — $P(M/A) = 0.005$

$P(B) = 0.6$ — $P(M/B) = 0.01$

$P(C) = 0.15$ — $P(M/C) = 0.011$

$P(D) = 0.05$ — $P(M/D) = 0.005$

$$P(A/M) = \frac{0.2 \times 0.005}{0.0077 + (1/600) + 2.5\times10^{-4}} = \underline{\underline{0.112}}$$

Ex 2.95, 2.96, 2.97, 2.98, 2.100
pg #76

Spam Filter

eg email → spam / not spam

Mechanism :- word search

criteria :- 0.9

threshhold

0.9 > spam

< 0.9 not spam

E=word

$$P(S/E) = \frac{P(S)P(E|S)}{[P(S)P(E|S)] + [P(NS)P(E/NS)]}$$

eg 2words search

$$P(S/E_1 \cap E_2) = \frac{P(S)P(E_1/S)P(E_2|S)}{P(S)P(E_1/S)P(E_2|S) + P(NS)P(E_1/NS)P(E_2/NS)}$$

Qu Found word Rolex. occurs in 250 of 2000 messages known to be spam. 5/1000 messages a known not to be spam. Estimate the P. that incoming message containing word Rolex is spam, assuming that it is equally likely that an incoming message is spam or not spam. threshold rejection - 0.9

Ans spam = 0.5 $\quad$ P(Rolex/S) = 0.125

not spam = 0.5 $\quad$ P(Rolex/NS) = 0.005

$$= \frac{0.5 \times 0.125}{(0.5 \times 0.125) + (0.5 \times 0.005)} = \frac{0.0625}{0.065}$$

$= 0.9615$

=> spam message

Q_4 2000 spam
1000 not spam
Stock ⇒ 400 spam
60 notspam
undervalued ⇒ 200 spam
25 notspam
P ⇒ containing both words
No prior knowledge of spam/Not spam

Spam = 0.5 — $P(E_1/S) = 0.2$
$P(E_2/S) = 0.1$
not spam = 0.5 — $P(E_1/NS) = 0.06$
$P(E_2/NS) = 0.025$

$$P(S/E_1 \cap E_2) = \frac{0.5 \times 0.2 \times 0.1}{(0.5 \times 0.2 \times 0.1) + (0.5 \times 0.06 \times 0.025)}$$
$$= \frac{0.01}{0.01075} = 0.93$$

⇒ Spam message