Referencer for Quick Revision

Foundation Course Paper-3: Business Mathematics, Logical Reasoning and Statistics

A compendium of subject-wise capsules published in the monthly journal "The Chartered Accountant Student"

Board of Studies
(Academic)
ICAI

INDEX

CHAPTER 4 : TIME VALUE OF MONEY

At the foundation level with regards to Business Mathematics the topic Time Value of money is very important for students not only to acquire professional knowledge but also for examination point of view. Here in this capsule an attempt is made for solving and understanding the concepts of time value of money with the help of following questions with solutions.

Problems on Simple, Compound and Effective rate of interest

Simple interest = P.T.I., Where P = Principal, T = Time, I = Rate of Interest

1. A sum of money amount to ₹ 6,200 in 2 years and ₹ 7,400 in 3 years. The principal and rate of interest is

Solution: A sum of money in 2 years = P + P.2.I = ₹ 6200
A sum of money in 3 years = P + P.3.I = ₹ 7400
Interest in 1 year = ₹ 1200; Interest in 2 years = ₹ 2400
Amount = ₹ 6200, P= Principal = 6200-2400= ₹ 3800.

$2400 = 3800\times2\times\frac{I}{100}$, I = rate of interest = 31.58% .

2. A sum of money doubles itself in 10 years. The number of years it would triple itself is

Solution: $2P = P + \frac{PTR}{100}$, $P = \frac{PTR}{100}$; T = 10 then R = 10%

$3P = P + \frac{PTR}{100}$ then $2P = \frac{PTR}{100}$ and R= 10%

Time (T) = 20 years.

Amount =P $(1+i)^n$
Compound rate of interest = P(1+i)^n-P= A-P
where P = principle i= interest n =conversion period

3. The population of a town increases every year by 2% of the population at the beginning of that year. The number of years by which the total increase of population be 40% is

Solution:1.4P = P(1+0.02)n
$(1.02)^n$=1.4, n = 17 years (app)

Depreciation (A)= $P(1-i)^n$
Where, A = Scrap Value, P = Original Cost, I =Depreciated at the rate, n = Number of years

4. A machine is depreciated at the rate of 20% on reducing balance. The original cost of the machine was ₹ 1,00,000 and its ultimate scrap value was ₹ 30,000. The effective life of the machine is

Solution: Here A = Scrap Value = 30,000 and
P = Original Cost = ₹ 1,00,000
30000 = 100000 $(1-0.2)^n$
3/10 = 0.3 = $(0.8)^n$, n = 5.4 years .

5. The useful life of a machine is estimated to be 10 years and cost ₹ 10,000. Rate of depreciation is 10% p.a. The scrap value at the end of its life is

Solution: Here A = Scrap Value =?,
P= Original cost = 10,000 n= 10 , I = 10%
A = 100000 $(1-0.1)^{10}$ = $10000(0.9)^{10}$
A = ₹ 3486.78

Effective rate of interest = $(1+i)^n$-1

6. The effective rate of interest corresponding a nominal rate of 7% p.a convertible quarterly is

Solution: Effective rate of interest = $(1+i)^n$-1, here n= 4,
i= 0.07/4 = 0.0175
= $(1+0.0175)^4$-1=1.07186-1=7.19%

The difference between simple and compound interest for 2 years = $P.i^2$, where P =Principal, i= interest

7. The difference between the S.I and the C.I on ₹ 2,400 for 2 years at 5% p.a is

Solution: The difference between simple and compound interest for 2 years =$2400(0.05)^2$= ₹ 6.

The difference between simple and compound interest for 3 years = $3P.i^2 + P.i^3$, where P = Principal i= Interest

8. The difference between the S.I and the C.I on a certain money invested for 3 years at 6% p.a is ₹ 110.16 the principle is

Solution: The differences between simple and compound interest for 3 years
= 110.16 =$P(3i^2+i^3)$= $P(3\times0.06^2+0.06^3)$
110.16 = P(0.0108+0.000216) =P(0.011016)

$P= \frac{110.16}{0.011016}$ = ₹ 10,000

9. The annual birth and death rates per 1,000 are 39.4 and 19.4 respectively. The number of years in which the population will be doubled assuming there is no immigration or emigration is

Solution: Here given, birth rate per 1,000 = 39.4 and death rates per 1,000 = 19.4
difference = 20 % per 1000 population

growth rate = $\frac{20}{1000}\times100=2\%$

Future population double
P =1000, A = 2000, r = 2%
2000= $1000(1+0.02)^n$
$(1.02)^n$ = 2
Number of years = n= 35

10. What annual rate of interest compounded annually doubles an investment in 7 years?
(Given that $2^{1/7}$= 1.104090)

Solution: If the principal be P, A_n = 2 p
Since A_n = $P(1+ i)^n$
2P = P (1+i)^7
$2^{1/7}$ = (1 + i)
1.104090 = 1 + i
I = 0.10409, Required rate of interest = 10.41% per annum

11. Vidya deposited ₹ 60000 in a bank for two years with the interest rate of 5.5% p.a. How much interest she would earn? what will be the final value of investment?

Solution: Required interest amount is given by,
I = P × it = ₹ 60,000 × $\frac{5.5}{100}$ ×2 = ₹ 6,600

The amount value of investment is given by, A = P + I = ₹ (60,000 + 6600) = ₹ 66,600

12. Rajiv invested ₹ 75,000 in a bank at the rate of 8% p.a. simple interest rate. He received ₹ 135,000 after the end of term. Find out the period for which sum was invested by Rajiv.

Solution: We know A = P + Pit = P (1+it)

i.e. 135000 = 75000 (1 + $\frac{8}{100}$ × t)

135000/75000 = $\frac{100+8t}{100}$

1.8×100 – 100=8t
80 = 8 t
t (Time) = 10 years

13. Which is a better investment, 3.6% per year compounded monthly or 3.2% per year simple interest? Given that (1+0.003)^12 =1.0366.

Effective rate of interest $E = (1 + i)^n - i$

Solution: i=3.6/12 = 0.3% =0.003, n = 12

$E = (1 + i)^n - 1$

$= (1 + 0.003)^{12} - 1 = 1.0366 - 1 = 0.0366$ or = 3.66%

Effective rate of interest (E) 3.66% is more than the simple interest so the Effective rate of interest (E) is better investment than the simple interest 3.2% per year.

14. The C.I on ₹ 4,000 for 6 months at 12% p.a payable quarterly is

Solution: Here P = ₹ 4000, n = 6/3= 2 , r= 0.12/4= 0.03

Compound Interest = $[P(1+i)^n - P] = [4000(1+0.03)^2 - 4000]$ =₹ 243.60

Annuity *applications*

F = Future value = C.F. $(1 + i)^n$ Where C.F = Cash flow
i= rate of interest, n = time period

15. Ravi invest ₹ 5000 in a two-year investment that pays you 12% per annum. Calculate the future value of the investment

Solution: We know, F= Future value = C.F.$(1 + i)^n$, Where C.F= Cash flow =₹ 5000, i = rate of interest=0.12,n = time period=2

F= ₹ $5000(1+0.12)^2$=₹ 5000×1.2544 =₹ 6272.

Annuity regular means :First payment at the end of the period.
Future value of the annuity regular =A(n,i)= A. $\left[\frac{(1+i)^n - 1}{i}\right]$

Annuity regular: In annuity regular first payment/receipt takes place at the end of first period.

16. Find the future value of an annuity of ₹ 5000 is made annually for 7 years at interest rate of 14% compounded annually. [Given that $(1.14)^7$= 2.5023]

Solution: Here annual payment A=₹ 5000, n=7, i=14% = 0.14

Future value of the annuity=A(7,0.14)=5000. $\left[\frac{(1+0.14)^7 - 1}{(0.14)}\right]$ = $5000\left[\frac{(2.5023 - 1)}{0.14}\right]$ = ₹ 53653.57

Future value of Annuity due or Annuity Immediate: When the first receipt or payment at the beginning of the annuity) it is called annuity due or annuity immediate.

17. ₹ 2000 is invested at the end of each month in an account paying interest 6% per year compounded monthly. What is the future value of this annuity after 10th payment? Given that $(1.005)^{10}$ = 1.0511

Solution: Here A= ₹ 2000, n= 10, i= 6% per annum= 6/12 % per month = 0.005

Future value of annuity after 10 months is given by

$$A(n,i) = A\left[\frac{(1+i)^n - 1}{i}\right]$$

$$A(10, 0.005) = 2000\left[\frac{(1+0.005)^{10} - 1}{0.005}\right] = 2000\left[\frac{(1.0511) - 1}{0.005}\right]$$

=2000×10.22 = ₹ 20440

Future value of the annuity regular or annuity due = $A\left[\frac{(1+i)^n - 1}{i}\right] \times (1 + i)$

18. Swati invests ₹ 20,000 every year starting from today for next 10 years. Suppose interest rate 8% per annum compounded annually. Calculate future value of the annuity. Given that $(1 + 0.08)^{10}$ = 2.158925)

Solution: Calculate future value as though it were an ordinary annuity.
Future value of the annuity as if it were an ordinary annuity

=₹ $20000\left[\frac{(1+0.08)^{10} - 1}{0.08}\right]$

=₹ 20000 × 14.486563 =₹ 289731.25

Multiply the result by (1 + i) =₹ 289731.25 × (1+0.08)

=₹ 312909.76

19. What is the present value of ₹ 100 to be received after two years compounded annually at 10%.

Solution: Here A_n = ₹ 100, i=10% = 0.1 , n= 2

Required present value= $\frac{A_n}{(1+i)^n} = \frac{100}{(1+0.1)^2} = \frac{100}{(1.21)}$ =82.64

Thus ₹ 82.64 shall grow to ₹ 100 after 2 years at 10% compounded annually.

20. Find the present value of ₹ 10000 to be required after 5 years if the interest rate be 9%. Given that $(1.09)^5$=1.5386.

Solution: Here i= 0.09 , n =5, An=10000

Required present value = $\frac{A_n}{(1+i)^n} = \frac{10000}{(1+0.09)^5} = \frac{10000}{(1.5386)}$=₹ 6499.42

Present Value of Annuity regular= A= P(n, i) =A. $\left[\frac{(1+i)^n - 1}{i(1+i)^n}\right]$,
$P(n,i) = \left[\frac{(1+i)^n - 1}{i(1+i)^n}\right]$

21. Soni borrows ₹ 5,00,000 to buy a car. If he pays equal instalments for 10 years and 10% interest on outstanding balance, what will be the equal annual instalment? Given [P(10,0.10) =6.14457]

Solution: We know, A = $\frac{V}{P(n,i)}$ Here V = ₹ 500000 , n = 10, I = 10% p.a.= 0.10

Annual Instalment= $\frac{V}{P(n,i)}$= ₹ $\frac{5,00,000}{P(10,0.10)}$= ₹ $\frac{5,00,000}{6,14457}$

= ₹ 8,1372.66

22. If ₹ 10,000 is paid every year for ten years to pay off a loan. What is the loan amount if interest rate be 14% per annum compounded annually? Given [P(10,0.14) =5.21611]

Solution: V = A.P(n,i) Here A = ₹ 10000 , n=10, i=0.14

V = 10000×P(10, 0.14)

= 10000×5.21611=₹ 52161.10

Therefore, the loan amount is ₹ 52161.10

23. Ram bought a Scooter costing ₹73000 by making a down payment of ₹ 3000 and agreeing to make equal annual payment for four years. How much would be each payment if the interest on unpaid amount be 14% compounded annually? Given [P(4 ,0.14) =2.91371]

Solution: In the present case we have present value of the annuity i.e. ₹ 70000 (73000-3000) and we have to calculate equal annual payment over the period of four years.

We know that, V = A. P (n , i) Here n = 4 and I = 0.14

$$A = \frac{V}{P(n,i)} = \frac{70000}{P(4,0.14)} = \frac{70000}{2.91371}$$

Therefore, each payment = ₹ 24024.35

24. Suppose your Father decides to gift you ₹ 20,000 every year starting from today for the next six years. You deposit this amount in a bank as and when you receive and get 10% per annum interest rate compounded annually. What is the present value of this annuity?

Solution: For calculating value of the annuity immediate following steps will be followed. Present value of the annuity as if it were a regular annuity for one year less i.e. for five years

$= ₹\ 20{,}000 \times P(5, 0.10)$

$= ₹\ 20{,}000 \times 3.79079 = ₹\ 75815.80$

Add initial cash deposit to the value, ₹ (75815.80+20,000) = ₹ 95815.80

Sinking Fund: Interest is computed at end of every period with specified interest rate.

25.How much amount is required to be invested every year so as to accumulate ₹ 5,00,000 at the end of 10 years if interest is compounded annually at 10%? Given A. (10, 0.1)= 15.9374248

Solution: Here A = 500000, n = 10, $A(n,i) = \left[\frac{(1+i)^n - 1}{i}\right] = \left[\frac{(1+01)^{10} - 1}{0.1}\right] = 15.9374248$

since A = P. A. (n ,i)

500000 =P A. (10, 0.1) = P × 15.9374248

$P = \left[\frac{500000}{15.9374248}\right] = ₹\ 31372.70$

26. ABC Ltd. wants to lease out an asset costing ₹ 360000 for a five year period. It has fixed a rental of Rs.105000 per annum payable annually starting from the end of first year. Suppose rate of interest is 14% per annum compounded annually on which money can be invested by the company. Is this agreement favorable to the company?

Solution: First, we have to compute the present value of the annuity of ₹ 105000 for five years at the interest rate of 14% p.a. compounded annually.

The present value V of the annuity is given by

V =A.P (n ,i) = 105000 × P (5, 0.14)

= 105000 ×3.43308 = ₹ 360473.40

which is greater than the initial cost of the asset and consequently leasing is favourable to the lessor.

27.A company is considering proposal of purchasing a machine either by making full payment of ₹ 4000 or by leasing it for four years at an annual rate of ₹ 1250. Which course of action is preferable if the company can borrow money at 14% compounded annually?

Solution: The present value V of annuity is given by

V =A.P (n ,i) = 1250 × P (4,0.14)

= 1250× 2.91371 = ₹ 3642.11

which is less than the purchase price, and consequently leasing is preferable.

28.A machine can be purchased for ₹ 50000. Machine will contribute ₹ 12000 per year for the next five years. Assume borrowing cost is 10% per annum compounded annually. Determine whether machine should be purchased or not.

Solution: The present value of annual contribution

V = A.P. (n ,i)

= 12000 P(5,0.10) = 12000 × 3.79079

= ₹ 45489.48

which is less than the initial cost of the machine. Therefore, machine must not be purchased.

29.A machine with useful life of seven years costs ₹ 10000 while another machine with useful life of five years costs ₹ 8000. The first machine saves labour expenses of ₹ 1900 annually and the second one saves labour expenses of ₹ 2200 annually. Determine the preferred course of action. Assume cost of borrowing as 10% compounded per annum.

Solution: The present value of annual cost savings for the first machine = ₹ 1900.P (7, 0.10)

= ₹ 1900 × 4.86842 = ₹ 9250

cost of machine being Rs.10000 it costs more by ₹ 750 than it saves in terms of labour cost.

The present value of annual cost savings of the second machine = ₹ 2200.P (5,0.10)= ₹ 2200 x 3.79079 = ₹ 8339.74

Cost of the second machine being ₹ 8000, effective savings in labour cost is ₹ 339.74. Hence the second machine is preferable.

30.An investor intends purchasing a three year ₹ 1000 par value bond having nominal interest rate of 10%. At what price the bond may be purchased now if it matures at par and the investor requires a rate of return of 14%?

Solution: Present value of the bond $= \frac{100}{(1+0.14)^1} + \frac{100}{(1+0.14)^2} + \frac{100}{(1+0.14)^3} + \frac{1000}{(1+0.14)^3}$

= 100 × 0.87719 + 100 × 0.769467 + 100 × 0.674 972

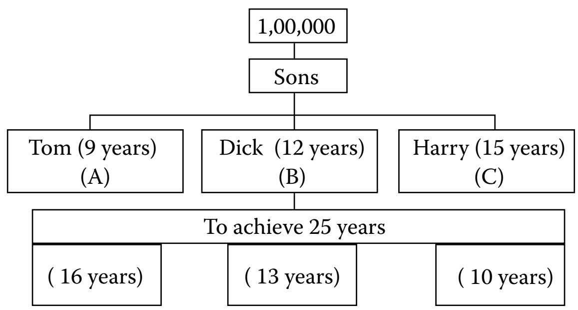
+ 1000 × 0.674972

= 87.719+ 76.947+ 67.497+ 674.972 = 907.125

Thus the purchase value of the bond is ₹ 907.125

31. Johnson left ₹ 1,00,000 with the direction that it should be divided in such a way that his minor sons Tom, Dick and Harry aged 9, 12 and 15 years should each receive equally after attaining the age 25 years. The rate of interest being 3.5%, how much each son receives after getting 25 years old?

Solution: Given problem can be explained as



$$A\left(1+\frac{3.5}{100}\right)^{16} = B\left(1+\frac{3.5}{100}\right)^{13} = +C\left(1+\frac{3.5}{100}\right)^{10}$$

$A(1.035)^{16}= B(1.035)^{13}= C(1.035)^{10}$

$A(1.035)^{6}= B(1.035)^{3}=$ C------(I)

$A : B : C = 1 : (1.035)^3 : (1.035)^6$

A + B + C= 100000

$x + x(1.035)^3 + x(1.035)^6 = 100000$

$x[1+(1.035)^3+(1.035)^6] = 10.0000$

x(3.337973) = 1,00,000 then x = ₹ 29958.30 (A' s share)

B' s share = 29958.30 (1.035)3= ₹ 33215.30

C' s share = 29958.30 (1.035)6= ₹ 36826.40

32. A machine costs ₹ 5,20,000 with an estimated life of 25 years. A sinking fund is created to replace it by a new model at 25% higher cost after 25 years with a scrap value realization of ₹ 25000. what amount should be set aside every year if the sinking fund investments accumulate at 3.5% compound interest p.a.?

Solution: Cost of new machine = $5{,}20{,}000 \times \frac{125}{100}$ = ₹ 6,50, 000, Scrap value = ₹ 25,000

For new machine = 650000-25000= ₹ 6,25, 000.

Here = ₹ 6,25, 000, n= 25 , i=3.5%=0.035

$$6{,}25{,}000 = P.\left[\frac{(1+i)^n - 1}{i}\right] = P.\left[\frac{(1+0.035)^{25} - 1}{0.035}\right]$$

$6{,}25{,}000 = P.[38.95]$ then $P = \frac{625000}{38.95} = ₹\ 16046.27$

Foundation Paper 3: Logical Reasoning Questions with explanations

At the Foundation level, students are expected to inculcate/evolve logical thinking and reasoning skills to further develop their analytical skills. This section attempts to capture basic techniques in sequential thinking as the underlying concept to solve problems. Here are a few Logical Reasoning Questions with explanations to get you psyched!

Chapter : 9 Number Series, Coding and decoding and Odd man out series

These questions deal in which series or letters in some orders and follows certain pattern throughout.

I. Find missing term of the series

(1) 101, 102, 106,115,131, 176, ?
(a)212 (b) 220 (c) 211 (d) 235
Explanation: Answer: (a)
The pattern of the series by adding **+1², +2², +3², +4²,+ 5², +6²,**
So missing term is 176+ 6²= 212.

(2) 3, 10, 29, 66, 127, ?
(a) 164 (b) 187 (c) 216 (d) 218
Explanation: Answer (d)
The pattern of the series is 1³+2 , 2³+2, 3³+2 , 4³+ 2, 5³+ 2, 6³+ 2
So missing number is, 6³+2= 216 + 2 = 218.

(3) 8, 13, 21, 32, 46, 63, 83 , ?
(a) 104 (b) 106 (c) 108 (d) 110
Explanation: Answer(b)
The pattern of the series is +5, +8, +11, +14, +17, +20, +23
So missing number is 83 + 23 = 106

4) 3, 4, 4, 6, 12, 15, 45, ?, 196
(a) 42 (b) 49 (c) 43 (d) 40
Explanation: Answer (b)
The pattern of the series is 3+1 ,4×1, 4+2, 6×2, 12+3, 15× 3, 45 + 4 , 49 × 4,
So missing term is = 45 + 4= 49

(5) 10, 12, 22, 34, 56, 90, ?
(a) 146 (b) 147 (c) 136 (d) 156
Explanation: Answer (a)
Each term in the series, except the first two terms, is the sum of preceding two terms
The right answer 56+90 = 146

(6) 4, 9, 19, 39, ?, 159, 319
(a) 40 (b) 41 (c) 78 (d) 79
Explanation: Answer (d)
Each number of the series is one more than the twice the preceding number
Therefore, missing term = 39×2+1= 79

(7) 7,15,29, 59, 117 ,?
(a) 238 (b) 235 (c) 120 (d) 155
Explanation: Answer (b)
The pattern is 7 ×2+1, 15 × 2-1, 29 × 2 + 1, 59 × 2 - 1, 117 × 2 + 1
So missing term is = 117 × 2 + 1 = 235

II. Find missing term of the letter series

(8) DBA, IDE, NFI, SHO, ?
(a) XJU (b) XYU(c) XUV (d) XUY
Explanation: Correct Option: Answer (a)
The first letter of the series is
D → +5 I → +5 N → +5 S → +5 X

The second letter of the series B →+2 D →+2 F →+2 H →+2 J

The third letters of the series are pattern
A – E – I – O – U › Vowels
So missing letter series is XJU

(9) cccaa_ bb_ cc_ aa_bb_
(a) abcab (b) babda (c) badna (d) bdanb
Explanation: Answer (a)
The pattern of the series is ccc, aaa, bbb, ccc, aaa follows.

(10) m _ n v _ n _a n _ a_ m a _
(a)aamvnn (b) aanvmm (c) vamaal (d) vanmak
Explanation: Answer (a)
The series man and van repeated

(11) In a certain language 'BISLERI' is written as 'CHTKFQJ' and 'AQUA' is written as 'BPVZ'. How is 'COMPUTER' written in the same Code?
(a) DNNOVSFQ (b) DNNVOSFX (c) DNNOVSXF
(d) DNONVSFX
Explanation: Answer (A)

B	I	S	L	E	R	I
•+1 c	•-1 H	•+1 T	•-1 K	•+1 F	•-1 Q	•+1 J

A	Q	U	A
•+1 B	•-1 P	•+1 V	•-1 Z

C	O	M	P	U	T	E	R
•+1 D	•-1 N	•+1 N	•-1 o	•+1 V	•-1 S	•+1 F	•-1 Q

So "COMPUTER' is coded as 'DNNOVSFQ

(12) If 'BROTHER, is coded as 2456784. 'SISTER' is coded as 919684 , what is the code for 'ROBBERS'?
(a) 4562684 (b) 9245784 (c) 4522849 (d) 4652684
Explanation: Answer (c)

B	R	O	T	H	E	R
•2	•4	•5	•6	•7	•8	•4

S	I	S	T	E	R
•9	•1	•9	•6	•8	•4

R	O	B	B	E	R	S
•4	•5	•2	•2	•8	•4	•9

Find odd one of the following series

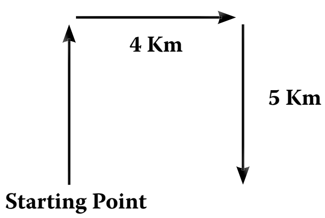
(13) (a) 144 (b) 169 (c) 288 (d) 324
Explanation: Answer (c), All others are perfect square numbers except 288

(14) (a) 73 (b) 53 (c) 87 (d) 23
Explanation: Answer (c), Except 87 all others are prime numbers

(15) (a) 4867 (b) 5555(c) 6243 (d) 6157
Explanation: Answer (d)
4867= 4+8+6+7 = 25, which is divisible by 5
5555= 5+5+5+5 = 20, which is divisible by 5
6243= 6+4+2+3 = 15, which is divisible by 5
6157= 6+1+5+7 =19, which is not divisible by 5

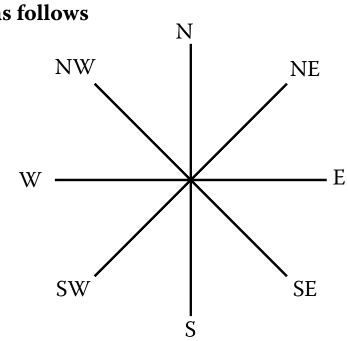
Chapter 10. Direction tests: In this test,the questions consist of a sort of direction puzzle. A successive follow-up of direction is formulated and the student is required to ascertain the final direction. The test is meant to judge the candidate's ability to trace and follow correctly and sense the direction correctly.

(1) A man walks 6 km North, turns right and walks 4 km, again turns right and walks 5 km, in which direction is he facing now?
(a) South (b) North (c) East (d) West
Explanation:

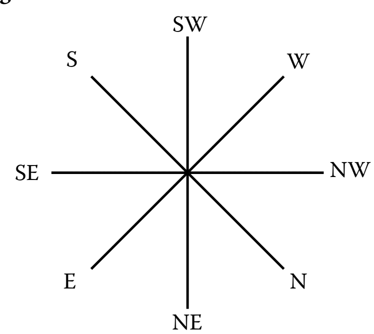


Answer: (a) South

(2) If South-East becomes North, North–East becomes West and all the rest of the directions are changed in same manner, what will be direction of the East?
(a) North-West (b) South (c) South-East (d) South-West
Explanation: According to question the direction of the diagram as follows



After changing the directions

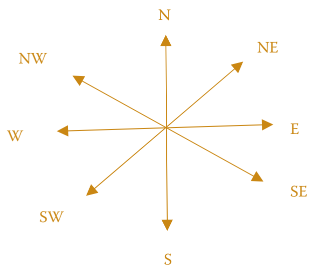


Now from the above diagrams North-West will be the direction for East.

Foundation Paper 3: Logical Reasoning Questions with explanations

At the Foundation level, students are expected to inculcate/evolve logical thinking and reasoning skills to further develop their analytical skills. This section attempts to capture basic techniques underlying concept of direction-related problems. Here are a few Logical Reasoning Questions with explanations to get you psyched!

CHAPTER 10: DIRECTION SENSE TESTS

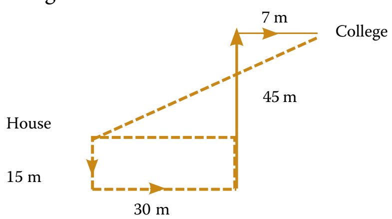


1. Gopal goes 15 m south from his house, turns left and walks 30 m , again turns left and walks 45 m , then turns right and walks 7 m to reach the college. In which direction is the college from his house?

   (a) North-East (b) West (c) East (d) North

   Explanation: Answer (a)

   According to the information stated in the question, direction diagram can be drawn as follows.

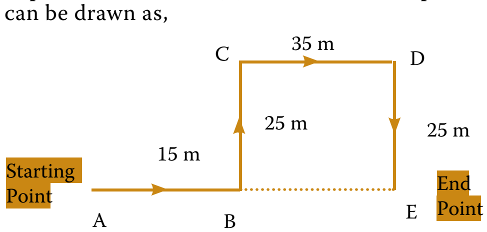


   So it's clear from the diagram that college is North -East direction from Gopal's house

2. Ram start moving from a point, facing in East direction. After walking 15 m, he turned to his left and walked 25m, before turning to his right. Then, he walked a distance of 35 m, then turned to his right and stop after walking further a distance of 25 m. Find how far Ram is from his starting point.

   (a) 20 m (b) 50 m (c) 15 m (d) 25 m

   Explanation: Answer (b), the direction map of Ram's walk can be drawn as,

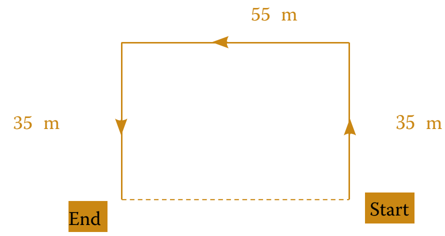


   The distance between the starting point and end point is AB + BE = 15 + 35 = 50 m.

3. Facing towards North, Ravi walks 35 m. He then turns left and walks 55 m. He again turns left and walks 35 m. How far is from original position and towards which direction.

   (a) 30 m, North (b) 20 m, East (c) 55 m, West (d) 20 m, South

   Explanation: Answer (c)



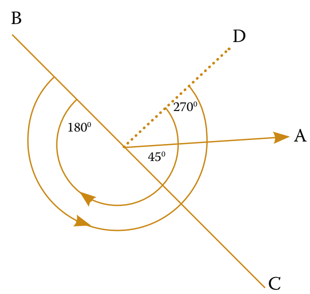
   From the figure it is clear that, Ravi is 55 m away in West direction from his original position.

4. A man is facing towards East and turns through 45° clock wise again 180° clock wise and then turns through 270° anti-clock wise. In which direction is he facing now?

   (a) West (b) North- East (c) South (d) South-West

   Explanation: Answer (b)

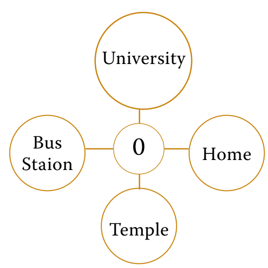
   As shown in figure , the man initially faces in the direction of OA. On moving 45° clockwise, the man faces the direction OB . On further moving 180° clockwise, he faces in the direction of OC. Finally on moving 270° anti-clock wise, he faces the direction OD, which is North-East.

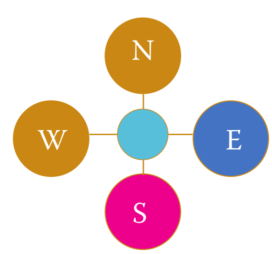


5. Kamal wants to go to university which is situated in a direction opposite to that of a temple. He starts from his house, which is in the East and comes at a four-ways place. His left side road goes to the temple and straight in front is the Bus Station. In which direction is university is located?

   (a) North (b) North- East (c) South (d) East

   Explanation: Answer (c)





   Kamal comes from East towards West. He reached O (four-way place). Now university will not in front or left. It will be towards the right, so it will be north direction.

CHAPTER 11. SEATING ARRANGEMENTS

The process of making group of people to sit as per a prefixed manner is called seating arrangement these questions, some conditions are given on the basis of which students are required to arrange objects, either in a row or on in circular order.

1. Six Children A, B, C, D, E and F are sitting in a row facing towards North. C is sitting between A and E, D is not at the end. B is sitting immediate right of E , F is not at the right of end, but D is sitting 3rd left of E . Which of the following is right of D.

   (a) A (b) F (c) E (d) C

   Explanation: Answer (a)

   According to the question A, B , C, D, E and F are sitting as follows.

F D A C E B

   Clearly A is sitting to the right of D.

2. Read the following information carefully and then answer the questions (i), (ii) and (iii).

   Six friends A, B, C, D, E and F are sitting on a bench, facing towards North.

   I. A is sitting next to B.
   II. C is sitting left to D.
   III. D is not sitting with E.
   IV. E is on the left end of the bench.
   V. C is third position from right.
   VI. A is on the right side of B and to the right side of E.
   VII. A and C are sitting together.
   VIII. F is sitting Right of D.

   (i) At what position A is sitting?
   (a) Between B and C (b) Between D and C
   (c) Between E and D (d) Between C and E
   (ii) What is position of B?
   (a) Second from right (b) Centre
   (c) Extreme left (d) Second from left
   (iii) What is position of D?
   (a) Extreme from left (b) Extreme right
   (c) Third from left (d) Second from right.

   Explanation: Arrangement according to the question is as follows.

Left E B A C D F Right

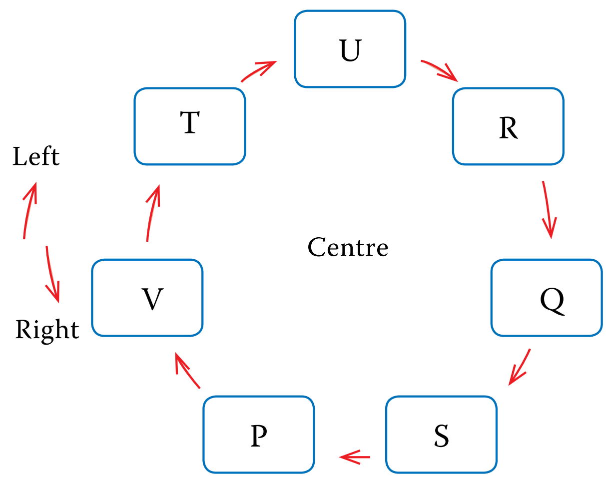
   2(i)Answer (a), A is sitting between B and C
   2(ii) Answer (d), B is sitting second from left.
   2 (iii) Answer (d), D is second from right

3. Read the following information carefully to answer the questions (i) and (ii)

I. P, Q, R, S, T, U, and V are sitting along a circle facing the centre
II. P is between V and S
III. R, who is second to the right of S , is between Q and U
IV. Q is not neighbour of T

3 (i) Which of the following statement is a correct statement?
(a) V is between T and P (b) S is second to left of V
(c) R is third to the left of P (d) P is to the immediate right of S
(ii) What is the position of P ?
(a) P is immediate left of S (b) to the immediate left of V
(c) 2nd to the left of R (d) 2nd to the right of Q

Explanation: Following seating arrangement is formed from the given information.



(i) Answer (a), based on diagram V is sitting T and P .
(ii) Answer (a), based on diagram P is immediate left of S.

4. Read the following information carefully to answer the questions given below:

   Seven boys A, B, C, D, E, F and G are standing in a line
   I . G is between A and D
   II. F and A have one boy between them
   III. D and C have two boys between them.
   IV E is immediate right of F.
   V. C and B have three boys between them
   (i) Who is second from right?
   (a) C (b) G (c) E (d) F
   (ii) Who is standing in the centre?
   (a) A (b) D (c) C (d) G

   Explanation: Arrangements according to the question.

Left B D G A C F E Right

   (i) Answer (d), Clearly F is second from right.
   (ii) Answer (a), Clearly A is standing in the centre.

FOUNDATION: PAPER 3 LOGICAL REASONING CHAPTER 12. BLOOD RELATIONS

Blood relations of a group of persons are given in jumbled form. In these tests, the questions which are asked depend on relation.

1. P is the sister of Q, Q is the brother of R, R is son of S. How S is related to P?

(a) Father

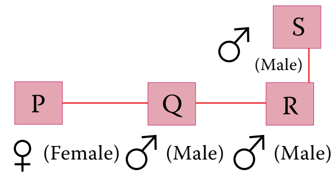
(b) Daughter

(c) Son

(d) Uncle

Explanation:

Answer: (a)



Based on the diagram Q and R brothers and P is their sister. Therefore, S is the father P.

2. A and B are brothers, C is A's mother, D is C's father, E is B's son. How is B is related to D.

(a) Son (b) Grandfather

(c) Grandson

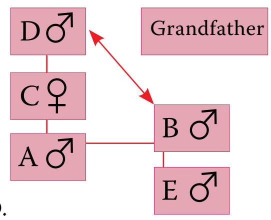
(d) Great Grandfather

Explanation: Answer(c)

D is father of C.

C is mother of A and B.

Therefore, B is Grandson of D.



3. A man showed a boy next to him and said – "he is the son of my wife's sister-in- law, but I am the only child of my parents ". How is my son is related to him?

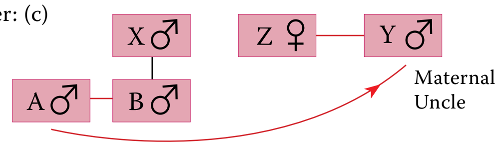
(a) Nephew (b) Cousin (c) Brother (d) Uncle

Explanation: Answer (b). The boy is the son of man's brother-in -law. Therefore, man's son is the cousin of that boy.

4. A and B are brothers. X is the father of B, Z is the only sister of Y and Y is maternal uncle of A, what is Z is related to X?

(a) Sister (b) Brother (c) Wife (d) Mother

Answer: (c)



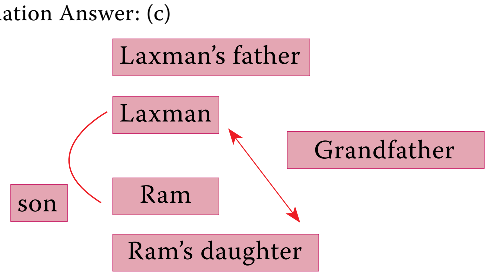
Based on the diagram A and B are bothers.

Y is bother of Z, therefore z is a female and Z is wife of X.

5. Introducing Ram to guests, Laxman said, "His father is the only son of my father". How is Ram's daughter is related Laxman?

(a) Nephew (b) Grandson (c) Grandfather (d) Son

Explanation Answer: (c)



Only son of Laxman's father is Laxman himself. Therefore, Ram' s father is Laxman.

Therefore, Laxman is Grandfather of Ram's daughter.

6. Pointing to a man in photograph, a woman said "His brother's father is the only son of my grandfather". Then How is women related to the man's son in the photograph?

(a) Daughter (b) Mother (c) Aunt (d) Sister

Explanation: Answer: (c) Only son of woman's grandfather means father of that woman.

Father of women is the father of man's brother and hence father of that man.

Therefore, the women is sister of the man and aunt to his son in photograph.

7. Read the following information carefully and answer the questions that follow

I: 'A + B' means A is the son of B.

II. 'A - B' means A is the wife of B.

III. 'A × B' means A is the sister of B.

IV. 'A ÷ B' means A is the mother of B.

V. 'A $ B' means A is the bother of B.

(i) What does P+ R - Q means

(a) Q is the father of P (b) Q is the son P

(c) Q is the uncle of P (d) Q is the bother of P.

(ii) What does P × R ÷ Q meaning

(a) P is the brother of R (b) P is the father of Q

(c) P is the aunt of Q (d) P is the nephew of Q

(iii) What does P $ R+ Q mean?

(a) P is the aunt of Q (b) P is the son of Q

(c) P is the niece of Q (d) P is the sister of Q

(iv) What does P $ R ÷ Q mean?

(a) P is the aunt of Q (b) P is the sister of Q

(c) Q is the niece of P (d) P is the uncle of Q

Explanation:

(i) Option (a), P+ R - Q, means P is the son of R, R is wife of Q, So Q is the father of P.

(ii) Option (c), P× R ÷ Q, means P is the sister of R, R is the mother of Q, So P is the aunt of Q.

(iii) Option (b), P $ R + Q, means P is the brother of R, R is the son of Q, So P is the son of Q.

(iv) Option (d), P $ R ÷ Q means P is the brother of R, R is the mother of Q. So P is uncle of Q.

8. On the basis of this information, you have to select the option which shows that A is the grandfather of T.

I. 'S x T' means that S is the mother of T,

II. 'S + T' means that S is the father of T,

III. 'S÷T' means that S is the brother of T.

(a) A + S + B ÷T

(b) A x B + C ÷ T

(c) A + C ÷T

(d) A + B ÷C x T

Explanation:

Option (a) represents that A is the grandfather of T

(i) B ÷T=> B is the brother of T.

(ii) S + B=> S is the father of B, hence S will be father of T [from information (i)).

(iii) A + S => A is the father of S, hence A will be grandfather of B and hence A is the grandfather of T.

CAPSULE: FOUNDATION PAPER 3: BUSINESS MATHEMATICS LOGICAL REASONING AND STATISTICS: CHAPTER 15 UNIT-I: MEASURES OF CENTRAL TENDENCY

At the foundation level with regards to Paper 3 Statistics part of the topic Measures of Central Tendency is very important for students not only to acquire professional knowledge but also for examination point of view. Here in this capsule an attempt is made for solving and understanding the concepts of Measures of Central Tendency with the help of following questions with solutions.

Definition of Central Tendency: Central tendency defined as the tendency of a given set of observations to cluster around a single central or middle value and the single value that represents the given set of observations is described as a measure of central tendency or, location, or average.

Following are the different measures of central tendency:

- Arithmetic Mean (AM)
- Median (Me)
- Mode (Mo)
- Geometric Mean (GM)
- Harmonic Mean (HM)

Criteria for an Idea l Measure of Central Tendency

- It should be properly and unambiguously defined.
- It should be easy to comprehend.
- It should be simple to compute.
- It should be based on all the observations.
- It should have certain desirable mathematical properties.
- It should be least affected by the presence of extreme observations.

Arithmetic Mean: defined as the sum of all the observations divided by the number of observations. Thus, if a variable x assumes n values $x_1, x_2, x_3, \ldots\ldots x_n$, then the AM of x, to be denoted by,

is given by, $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$, $\bar{X} = \frac{X_1 + X_2 + X_3 + \ldots\ldots + X_n}{n}$

In case of a simple frequency distribution relating to an attribute, we have $\bar{X} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$

$\bar{X} = \frac{\sum_{i=1}^{n} f_i x_i}{N}$ assuming the observation xi occurs fi times, i=1,2,3,……..n and N=≤fi.

In case of grouped frequency distribution also we may use formula with xi as the mid value of the i-th class interval, on the assumption that all the values belonging to the i-th class interval are equal to xi.

If classification is uniform, we consider the following formula for the computation of AM from grouped frequency the distribution:

$$\bar{x} = A + \frac{\sum f_i d_i}{N} \times C$$

Where, $d_i = \frac{x_i - A}{C}$ A = Assumed Mean C = Class Length

- **Properties of AM**
- **If all the observations assumed by a variable are constants, say k, then the AM is also k.**
- The algebraic sum of deviations of a set of observations from their AM is zero
- i.e. for unclassified data, ∑ (xi -) = 0 and for grouped frequency distribution, $\sum(f_i(xi -)) = 0$
- AM is affected due to a change of origin and/or scale which implies that if the original variable x is changed to another variable y by effecting a change of origin, say a, and scale say b, of x i.e. $\bar{y} = a + b\bar{x}$, then the AM of y is given by
- If there are two groups containing n1 and n2 observations and $\bar{x}_1$ and $\bar{x}_2$ as the respective arithmetic means, then combined AM is given by $\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

Question 1: Following are the daily wages in rupees of a sample of 10 workers: 58, 62, 48, 53, 70, 52, 60, 84, 75, 100. Compute the mean wage.

Solution: Let x denote the daily wage in rupees.

Applying the mean wage is given by,

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{(58+62+48+53+70+52+60+84+75+100)}{10} = \frac{₹662}{10} = ₹66.2$$

Question 2: Compute the mean weight of a group of B. Com students of Sri Ram College from the following data:

Weight in kgs	44-48	49-53	54-58	59-63	64-68	69-73
No. of students	3	4	5	7	9	12

Solution: Computation of mean weight of 40 B. Com students Applying formula, we get the average weight as

$$\bar{x} = \frac{\sum f_i x_i}{N} \frac{2495}{40} \text{ kgs.} = 62.38 \text{ kgs.}$$

Weight in kgs. (1)	No. of Student (f_i) (2)	Mid-Value (x_i) (3)	$f_i x_i$ (4) = (2) x (3)
44 – 48	3	46	138
49 – 53	4	51	204
54 – 58	5	56	280
59 – 63	7	61	427
64 – 68	9	66	594
69 – 73	12	71	852
Total	40	–	2495

Question 3: Find the AM for the following distribution:

Class Interval	5-14	15-24	25-34	35-44	45-54	55-64
Frequency	10	18	32	26	14	10

Solution: Any mid value can be taken as A. However, usually A is taken as the middle most mid-value for an odd number of class intervals and any one of the two middle most mid-values for an even number of class intervals. The class length is taken as C.

The required AM is given by

$$\bar{x} = A + \frac{\sum f_i d_i}{N} \times C = 39.5 + \frac{-64}{110} \times 10 = 39.5 - 5.82 = 33.68$$

Table: Computation of AM

Class Interval	Frequency (f_i)	Mid-Value (x_i)	$d_i = \frac{x_i - A}{C}$	$f_i d_i$
(1)	(2)	(3)	(4)	(5) = (2) x (4)
5-14	10	9.5	– 3	– 30
15-24	18	19,5	– 2	– 36
25-34	32	29.5	– 1	– 32
35-44	26	39.5(A)	0	0
45-54	14	49.5	1	14
55-64	10	59.5	2	20
Total	110	-	-	-64

Question 4: Given that the mean height of a group of students is 67.45 inches. Find the missing frequencies for the following incomplete distribution of height of 100 students.

Height in inches	60-62	63-65	66-68	69-71	72-74
No. of students	5	18	-	-	8

Solution: Let x denote the height and f_3 and f_4 as the two missing frequencies

Table: Estimation of missing frequencies

Class Interval	Frequency	Mid-Value (x_i)	$d_i = \frac{x_i - A}{C} = \frac{x_i - 67}{3}$	$F_i d_i$ (f_i)
(1)	(2)	(3)	(4)	(5) = (2) x (4)
60-62	5	61	-2	-10
63 – 65	18	64	– 1	– 18
66 – 68	f3	67 (A)	0	0
69 – 71	f4	70	1	f4
72 – 74	8	73	2	16
Total	$31+ f_3+ f_4$	–	–	$– 12+f_4$

As given, we have

$31+ f_3+ f_4= 100$, $f_3+ f_4= 69$.................(1)

$\overline{x} = 67.45$

and $A + \frac{\sum f_i d_i}{N} \times C = 67.45 = 67 + \frac{(12 + f_4)}{100} \times 3 = 67.45$

$(-12 + f_4) \times 3 = (67.45- 67) \times 100$

$-12 + f_4 = 15$, $f_4 = 27$

On substituting 27 for f_4 in (1), we get ,$f_3 + 27 = 69, f_3 = 42$, Thus, the missing frequencies would be 42 and 27.

Question 5: The mean salary for a group of 40 female workers is ₹5,400 per month and that for a group of 60 male workers is ₹7,800 per month. What is the combined mean salary?
Solution: As given n_1=40, n_2=60, $\overline{x}_1$=₹5,400 and $\overline{x}_2$=₹7,800 hence, the combined mean salary per month is

$$\overline{x} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2}{n_1 + n_2} = \frac{40\times ₹5,400+60\times ₹7,800}{40+60} = ₹6,840.$$

Question 6: The mean weight of 150 students (boys and girls) in a class is 60 kg. The mean weight of boy student is 70 kg and that of girl student is 55 kg . Find number of boys and girls in that class.
Solution: Le t the number of boy students be n1 and girl students be n_2 , as given $n_1+n_2 = 150$,
Then $n_2 = 150-n_1$, also $\overline{x} = 60, \overline{x}_1 = 70, \overline{x}_2 = 55$

$$\overline{x} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2}{n_1 + n_2}, 60 = \frac{n_1 \times 70 + (150 - n_1) \times 55}{150}$$

$$60 = \frac{70n_1 + 8250 - 55n_1}{150} = \frac{15n_1 + 8250}{150}$$

$= 9000 = 15n_1 + 8250$

$15n_1 = 750, n_1 = 50$

$n_2 = 150 - n_1 = 150 - 50 = 100$

Therefore, number of boys (n_1)= 50.
Number of girls (n_2)= 100

Question 7: The average salary of a group of unskilled workers is Rs. 10,000 and that of a group of skilled workers is Rs. 15000.If the combined salary is Rs.12000, then what is the percentage of skilled workers
Solution: Let x be unskilled and y be skilled
10000x+15000y= 12000(x+y) = 12000x+12000y
2000x=3000y then 2x=3y
skilled workers is 2x/3
total workers x+2x/3=5x/3
percentage of skilled=2x/3 divided by 5x/3=40%

Question 8: The average age of a group of 10 students was 20 years. The average age increased by two years when the two new students joined in the group. What is the average age of two new students joined who joined in the group?
Solution: Average age of 10 students = 20 years, then sum of ages of 10 students = 200 years
If the two boys are included, then total number of students = 10+2 = 12
And average increased by two years = 20 +2 =22
The average age of 12 students = 22, then sum of ages of 12 students = 22×12= 264
The Sum of ages of two boys = 264-200= 64
Average age of boys = 64/2 = 32

Median

- Partioned Values
- As compared to AM, median is a positional average which means that the value of the median is dependent upon the position of the given set of observations for which the median is wanted. Median, for a given set of observations, may be defined as the middle-most value when the observations are arranged either in an ascending order or a descending order of magnitude.

Question 9: The median of the data 13, 8, 11, 6, 4, 15, 2, 18, 20 is
Solution: Arranging the data in a ascending order, we get 2, 4, 6, 8, 11, 13, 15, 18, 20

Here n= 9, which is odd number of observations.

$$\text{Median} = \left(\frac{n+1}{2}\right)^{th} \text{item} = \left(\frac{9+1}{2}\right)^{th} \text{item} = 5^{th} \text{item} = 11$$

Question 10: What is the median for the observations 5, 8, 6, 9, 11 and 4
Solution: We write in ascending order 4, 5, 6, 8, 9 and 11

Here n = 6. So Median = Average of 3^{rd} and 4^{th} term = $\frac{6+8}{2} = 7$

In case of a grouped frequency distribution, we find median from the cumulative frequency distribution of the variableunder consideration.

$$M = l_1 + \left(\frac{\frac{N}{2} - N_l}{N_u - N_l}\right) \times C$$

Where,
l_1 = lower class boundary of the median class i.e. the class containing median.
N = total frequency.
N_l = less than cumulative frequency corresponding to l_1.(Pre median class)
N_u = less than cumulative frequency corresponding to l_2. (Post median class)
l_2 being the upper class boundary of the median class.
$C = l_2 – l_1$ = length of the median class.

Question 11: What is the Median for the following data?

Marks	5-14	15-24	25-34	35-44	45-5	50-59
No.of Students	10	18	32	26	14	10

Solution: First, we find the cumulative frequency distribution which is exhibited in the table

Marks	Frequency (f)	Marks	Less than Cumulative Frequency (CF)
5-14	10	14.5	10
15-24	18	24.5	28(N_l)
25-34	32	34.5	60(N_u)
35-44	26	44.5	86
45-54	14	54.5	100
55-64	10	64.5	110
Total (N)	110		

We find from table N/2 = 55

$\frac{N}{2} = \frac{110}{2}$ = 55 lies between the two cumulative frequencies 28 and 60 i.e. 28 < 55 < 60.

Thus, we have $N_l = 28$, $N_u = 60$, $l_1 = 24.5$ and $l_2 = 34.5$.

Hence C = 34.5 – 24.5 =10.

Substituting these values formula, we get,

$$M = 24.5 + \frac{55 - 28}{60 - 28} \times 10 = 24.5 + 8.44 = 32.94$$

Question 12: Find the missing frequency from the following data, given that the median mark is 23.

Marks	0-10	10-20	20-30	30-40	40-50
No.of Students	5	8	?	6	3

Solution: Let us denote the missing frequency by f_3. Following table shows the relevant computation.

Table (Estimation of Missing frequency)

Marks	Less than cumulative frequency
0	0
10	5
20 (l_1)	13(N_l)
30(l_2)	13+f_3(N_u)
40	19+f_3
50	22+f_3

Going through the mark column, we find that 20<23<30. Hence l_1=20, l_2 =30 and accordingly Nl=13, Nu=13+f_3. Also the total frequency i.e. N is 22+f_3. Thus,

$$M = l_1 + \left(\frac{\frac{N}{2} - N_l}{N_u - N_l}\right) \times C$$

$$23 = 20 + \frac{\left(\frac{22 + f_3}{2}\right) - 13}{(13 + f_3) - 13} \times 10$$

$$3 = + \frac{22 + f_3 - 26}{f_3} \times 5 \ , \ 3f_3 = 5f_3 - 20 \ , \ f_3 = 20$$

f_3 = 10 ,So, the missing frequency is 10.

Properties of median: We cannot treat median mathematically; the way we can do with arithmetic mean. We consider below two important features of median.

(i) If x and y are two variables, to be related by y=a+bx for any two constants a and b, then the median of y is given by

$y_{me} = a + bx_{me}$

For example, if the relationship between x and y is given by 2x – 5y = 10 and if x_{me} i.e. the median of x is known to be 16.

Then 2x – 5y = 10

⇨ y = –2 + 0.40x

⇨ $y_{me} = -2 + 0.40\, x_{me}$

⇨ $y_{me} = -2 + 0.40 \times 16$

⇨ $y_{me} = 4.40$.

(ii) For a set of observations, the sum of absolute deviations is minimum when the deviations are taken from the median. This property states that $\Sigma|x_i - A|$ is minimum if we choose A as the median.

PARTITION VALUES OR QUARTILES : These may be defined as values dividing a given set of observations into a number of equal parts. When we want to divide the given set of observations into two equal parts, we consider median. Similarly, quartiles are values dividing a given set of observations into four equal parts. So there are three quartiles – first quartile or lower quartile denoted by Q_1, second quartile or median to be denoted by Q_2 or Me and third quartile or upper quartile denoted by Q_3. First quartile is the value for which one fourth of the observations are less than or equal to Q_1 and the remaining three – fourths observations are more than or equal to Q_1. In a similar manner, we may define Q_2 and Q_3.

Deciles : Deciles are the values dividing a given set of observation into ten equal parts. Thus, there are nine deciles to be denoted by D_1, D_2, D_3,.....D_9. D_1 is the value for which one-tenth of the given observations are less than or equal to D_1 and the remaining nine-tenth observations are greater than or equal to D_1 when the observations are arranged in an ascending order of magnitude.

percentiles or centiles : Percentiles divide a given set of observations into 100 equal parts. The points of sub-divisions being P_1, P_2,...........P_{99}. P_1 is the value for which one hundredth of the observations are less than or equal to P_1 and the remaining ninety-nine hundredths observations are greater than or equal to P_1 once the observations are arranged in an ascending order of magnitude.

For unclassified data, the p^{th} quartile is given by the $(n+1)p^{th}$ value, where n denotes the total number of observations. p = 1/4, 2/4, 3/4 for Q_1, Q_2 and Q_3 respectively. p=1/10, 2/10,............9/10.

For D_1, D_2,......,D_9 respectively and l astly p=1/100, 2/100,....,99/100 for P_1, P_2, P_3....P_{99} respectively.

In case of a grouped frequency distribution, we consider the following formula for the computation of quartiles.

$$Q = l_1 + \left(\frac{N_p - N_l}{N_u - N_l}\right) \times C$$

The symbols, except p, have their usual interpretation which we have already discussed while computing median and just like the unclassified data, we assign different values to p depending on the quartile.

Another way to find quartiles for a grouped frequency distribution is to draw the ogive (less than type) for the given distribution. In order to find a particular quartile, we draw a line parallel to the horizontal axis through the point Np. We draw perpendicular from the point of intersection of this parallel line and the ogive. The x-value of this perpendicular line gives us the value of the quartile.

Question 13: Following are the wages of the labourers: ₹82, ₹56, ₹90, ₹50, ₹120, ₹75, ₹75, ₹80, ₹130, ₹65. Find Q_1, D_6 and P_{82}.

Solution: Arranging the wages in an ascending order, we get ₹50, ₹56, ₹65, ₹75, ₹75, ₹80, ₹82, ₹90, ₹120, ₹130.

Hence, we have

$Q_1 = \frac{(n+1)}{4}$th value $= \frac{(10+1)}{4}$th value = 2.75^{th} value

= 2^{nd} value + 0.75 × difference between the third and the 2^{nd} values.

= ₹[56 + 0.75 × (65 – 56)] = ₹62.75

D_6 = (10 + 1) × $\frac{6}{10}$ th value = 6.60^{th} value

= 6^{th} value + 0.60 × difference between the 7^{th} and the 6^{th} values.

= ₹(80 + 0.60 × 2) = ₹81.20

P_{82} = (10+1) x $\frac{82}{100}$ th value = 9.02^{th} value

= 9^{th} value + 0.02 × difference between the 10^{th} and the 9^{th} values

= ₹(120 + 0.02 ×10) = ₹120.20

Question 14: Compute the Third Quartile and 65^{th} percentile for the following data

Profits '000 Rs	Less than 10	10-19	20-29	30-39	40-49	50-59
No. of firms	5	18	38	20	9	2

Solution:

Profits'000(Rs.)	Frequency (f)	Cumulative Frequency (CF)
Less than 9.5	5	5
9.5-19.5	18	23
19.5-29.5	38	61
29.5-39.5	20	81
39.5-49.5	9	90
49.5-59.5	2	92
110		

Q_3 = Third Quartile = $\frac{3N}{4} = \frac{3 \times 92}{4} = 69$ Q_3 lies 29.5 and 39.5

$$Q_3 = 29.5 + \left(\frac{69-61}{20}\right) \times 10 = 33.5$$

Q_3 = ₹33,500

For 65^{th} percentile = $P_{65} = \frac{iN}{100} = \frac{65 \times 92}{100} = 59.8 = 65^{th}$

percentile lies in the class 19.5-29.5, here l= 19.5, c = 23, f = 38 and C = 10

$$P_{65} = 19.5 + \frac{(59.8-23)}{38} \times 10$$

$$P_{65} = 29.184$$

= ₹29,184

Question 15: Compute mode for the distribution for the following distribution

Class Interval	350-369	370-389	390-409	410-429	430-449	450-469
Frequency	15	27 (f-1)	31(fo)	19 (f1)	13	6

Solution: Going through the frequency column, we note that the highest frequency i.e., fo = 31 and f_{-1}= 27, f_1 = 19, LCB = 389.50, C = 409.5-389.50= 20

$$\text{Mode} = 389.5 + \frac{(31-27)}{2 \times 31 - (27+19)} \times 20$$

$$\text{Mode} = 389.5 + \frac{4}{16} \times 20 = 389.5 + 5 = 394.5$$

Question 16: For a moderately skewed distribution of marks in statistics for a group of 200 students, the mean mark and median mark were found to be 55.60 and 52.40. What is the modal mark?

Solution: Since in this case, mean = 55.60 and median = 52.40, applying, we get the modal mark as,

Mode = 3 × Median – 2 × Mean = 3 × 52.40 – 2 × 55.60 = 46.

Question 17: If x and y related by x-y-10=0 and mode of x is known to be 23, then the mode of y is :

Solution: Mode of x = 23 , x-y-10= 0 then y = x-10 , Mode of y = 23-10 = 13

Geometric Mean: For a given set of n positive observations, the geometric mean is defined as the n-th root of the product of the observations. Thus if a variable x assumes n values $x_1, x_2, x_3, \ldots\ldots\ldots, x_n$, all the values being positive, then the GM of x is given by G= $(x_1 \times x_2 \times x_3 \ldots\ldots\ldots \times x_n)^{1/n}$

For a grouped frequency distribution, the GM is given by

G= $(x_1^{f1} \times x_2^{f2} \times x_3^{f3} \ldots\ldots\ldots \times x_n^{fn})^{1/N}$, Where N = Σf_i

In connection with GM, we may note the following properties:

(i) Logarithm of G for a set of observations is the AM of the logarithm of the observations

(ii) if all the observations assumed by a variable are constants, say K > 0, then the GM of the observations is also K.

(iii) GM of the product of two variables is the product of their GM's i.e. if z = xy, then
GM of z = (GM of x) × (GM of y)

(iv) GM of the ratio of two variables is the ratio of the GM's of the two variables i.e. if z = x/y then

GM of z = $\frac{\text{GM of x}}{\text{GM of y}}$

Question 18: Find the GM of 8, 24 and 40.

Solution: As given x_1=8, x_2=24, x_3=40 and n=3.

Applying, we have G= $(8 \times 24 \times 40)^{1/3} = 8.\sqrt[3]{15}$

Question 19: If GM of x is 10, and GM of y is 15, then GM of xy

Solution: According to the GM of XY= GM of x * GM of y

= 10 × 15 = 150

Harmonic Mean: For a given set of non-zero observations, harmonic mean is defined as the reciprocal of the AM of the reciprocals of the observation. So, if a variable x assumes n non-zero values $x_1, x_2, x_3, \ldots\ldots\ldots, x_n$, then the HM of x is given by

$$H = \frac{n}{\Sigma(1/x_i)}$$

For a grouped frequency distribution, we have $H = \frac{N}{\Sigma\left[\frac{f_i}{x_i}\right]}$

Properties of HM

(i) If all the observations taken by a variable are constants, say k, then the HM of the observations is also k.

(ii) If there are two groups with n1 and n2 observations and H1 and H2 as respective HM's than the combined HM is given by = $\frac{n_1 + n_2}{\frac{n_1}{H_1} + \frac{n_2}{H_2}}$

Question 20: A man travels at a speed of 20 km/ hr and then returns at a speed of 30 km/hr. His average of the whole journey is

Solution: Harmonic Mean is the method which is preferred for the computation of average speed

$$HM = \frac{2ab}{a+b} = \frac{2 \times 20 \times 30}{20+30} = 24\text{km/hr}$$

Question 21: Find the HM for 4, 6 and 10.

Solution: Applying formula, we have

$$H = \frac{3}{\frac{1}{4}+\frac{1}{6}+\frac{1}{10}} = \frac{3}{0.25+0.17+0.10} = 5.77$$

Question 22: An aeroplane flies from A to B at the rate of 500 km/ hr and comes back B to A at the rate of 700 km / hr . The average speed of the aeroplane is;

Solution: Required average speed of the aeroplane =

$$\frac{2}{\left(\frac{1}{500}+\frac{1}{700}\right)} = \frac{2 \times 3500}{7+5} = 583.33\text{km / hr}$$

Question 23: Find the HM for the following data:

x	2	4	8	16
f	2	3	3	2

Solution: Using formula, we get $H = \dfrac{10}{\frac{2}{2}+\frac{3}{4}+\frac{3}{8}+\frac{2}{16}} = 4.44$

Relation between AM, GM, and HM → For any set of positive observations, we have the following inequality: AM ≥ GM ≥HM → The equality sign occurs, as we have already seen, when all the observations are equal.

Question 24: compute AM, GM, and HM for the numbers 6, 8, 12, 36.

Solution: In accordance with the definition, we have

$$AM = \frac{6+8+12+36}{4} = 15.50$$

$$GM = (6 \times 8 \times 12 \times 36)^{1/4}$$

$$= (2^8 \times 3^4)^{1/4} = 12$$

$$HM = \frac{4}{\frac{1}{6}+\frac{1}{8}+\frac{1}{12}+\frac{1}{36}} = 9.93$$

The computed values of AM, GM, and HM establish AM ≥ GM ≥HM

Question 25: If there are two groups with 75 and 65 as harmonic means and containing 15 and 13 observations, then the combined HM is given by

Solution: Combined HM is given by $= \dfrac{n_1+n_2}{\left(\frac{n_1}{H_1}+\frac{n_2}{H_2}\right)} = \dfrac{15+13}{\left(\frac{15}{75}+\frac{13}{65}\right)} = 70$

Weighted average

When the observations under consideration have a hierarchical order of importance, we take recourse to computing weighted average, which could be either weighted AM or weighted GM or weighted HM.

$$\text{Weighted AM} = \frac{\sum w_i x_i}{\sum w_i}$$

$$\text{Weighted GM} = \text{Ante log}\left(\frac{\sum w_i \log x_i}{\sum w_i}\right)$$

$$\text{Weighted HM} = \frac{\sum w_i}{\sum\left(\frac{w_i}{x_i}\right)}$$

FOUNDATION: PAPER 3 - BUSINESS MATHEMATICS LOGICAL REASONING AND STATISTICS

At the foundation level with regards to Paper 3 Statistics part of the topic Measures of Dispersion is very important for students not only to acquire professional knowledge but also for examination point of view. Here in this capsule an attempt is made for solving and understanding the concepts of Measures of Dispersion with the help of following questions with solutions

CHAPTER 15 UNIT-II: MEASURES OF DISPERSION

The second important characteristic of a distribution is given by dispersion. Two distributions may be identical in respect of its first important characteristic i.e. central tendency and yet they may differ on account of scatterness. The following figure shows a number of distributions having identical measure of central tendency and yet varying measure of scatterness. Obviously, distribution is having the maximum amount of dispersion.

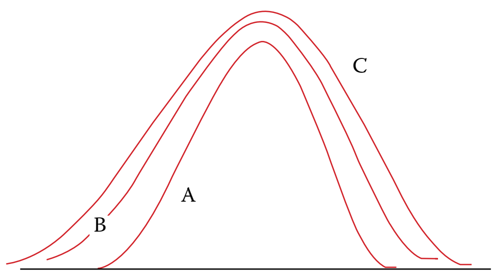


Figure: Showing distributions with identical measure of central tendency and varying amount of dispersion.

Dispersion

for a given set of observations may be defined as the amount of deviation of the observations, usually, from an appropriate measure of central tendency. Measures of dispersion may be broadly classified into the following:

Absolute measures of dispersion

(i) Range
(ii) Mean Deviation
(iii) Standard Deviation
(iv) Quartile Deviation

Relative measures of dispersion

(i) Coefficient of range.
(ii) Coefficient of Mean Deviation
(iii) Coefficient of Variation
(iv) Coefficient of Quartile Deviation

Distinction between the absolute and relative measures of dispersion :

- Absolute measures are dependent on the unit of the variable under consideration whereas the relative measures of dispersion are unit free.
- For comparing two or more distributions, relative measures and not absolute measures of dispersion are considered.
- Compared to absolute measures of dispersion, relative measures of dispersion are difficult to compute and comprehend.

Characteristics for an ideal measure of dispersion

An ideal measure of dispersion should be properly defined, easy to comprehend, simple to compute, based on all the observations, unaffected by sampling fluctuations and amenable to some desirable mathematical treatment

Range :

- For a given set of observations, range may be defined as the difference between the largest and smallest of observations.
- Thus if L and S denote the largest and smallest observations respectively thenRange = L – S

Coefficient of Range = $\frac{L-S}{L+S}\times 100$

For a grouped frequency distribution: Range is defined as the difference between the two extreme class boundaries. The corresponding relative measure of dispersion is given by the ratio of the difference between the two extreme class boundaries to the total of these class boundaries, expressed as a percentage.

Range remains unaffected due to a change of origin but affected in the same ratio due to a change in scale i.e., if for any two constants a and b, two variables x and y are related by y = a + bx, Then the range of y is given by
$R_y = |b| \times R_x$

Example 1: Following are the wages of 8 workers expressed in rupees: 80, 65, 90, 60, 75, 70, 72, 85. Find the range and find its coefficient.

Solution: The largest and the smallest wages are L = ₹90 and S= ₹60
Thus range = ₹90 – ₹60 = ₹30

Coefficient of range = $\frac{90-60}{90+60}\times 100 = 20$

Example 2: What is the range and its coefficient for the following distribution of weights?

Weight in kgs	10-19	20-29	30-39	40-49	50-59
No. of Students	11	25	16	7	3

Solution: The lowest class boundary is 9.50 kgs. and the highest-class boundary is 59.50 kgs. Thus, we have
Range = 59.50 kgs. – 9.50 kgs. = 50 kgs.

Also, coefficient of range = $\frac{59.50-9.50}{59.50+9.50}\times 100 = \frac{50}{69}\times 100 = 72.46$

Example 3: If the relationship between x and y is given by 2x+3y=10 and the range of x is ₹15, what would be the range of y?

Solution: Since $2x+3y=10$

Therefore, $y = \frac{10}{3} - \frac{2}{3}x$, Applying the range of y is given by

$R_y = |b| \times R_x$ = 2/3 × ₹15 = ₹10.

Mean Deviation : Since range is based on only two observations, it is not regarded as an ideal measure of dispersion. A better measure of dispersion is provided by mean deviation which, unlike range, is based on all the observations. For a given set of observation, mean deviation is defined as the arithmetic mean of the absolute deviations of the observations from an appropriate measure of central tendency. Hence if a variable x assumes n values $x_1, x_2, x_3 \ldots x_n$, then the mean deviation of x about an average A is given by

$MD_A = \frac{1}{n}\sum|x_i - A|$ For a grouped frequency distribution, mean deviation about A is given by $MD_A = \frac{1}{n}\sum|x_i - A|.f_i$

Where x_i and f_i denote the mid value and frequency of the i^{th} class interval and $N = \sum f_i$

In most cases we take A as mean or median and accordingly, we get mean deviation about mean or mean deviation about median or mode.

A relative measure of dispersion applying mean deviation is given by

$$\text{Coefficient of Mean Deviation} = \frac{\text{Mean deviation about A}}{\text{A}} \times 100$$

Mean deviation takes its minimum value when the deviations are taken from the median. Also mean deviation remains unchanged due to a change of origin but changes in the same ratio due to a change in scale i.e. if y = a + bx, a and b being constants,

then MD of y = |b| × MD

Example 4: What is the mean deviation about mean for the following numbers?

50,60,50,50,60,60,60,50,50,50,60,60,60,50.

Solution: The mean is given by

$$\bar{X} = \frac{50+60+50+50+60+60+60+50+50+50+60+60+60+50}{14} = \frac{770}{14} = 55$$

x_i	50	60	50	50	60	60	60	50	50	50	60	60	60	50	Total
$\lvert x_i - \bar{x}\rvert$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	70

Thus, mean deviation about mean is given by $\frac{\sum|x_i - \bar{x}|}{n} = \frac{70}{14} = 5$

Example. 5 : The coefficient of Mean Deviation about the first 9 natural numbers ?

Solution: The Mean of first 9 natural numbers = $\frac{n+1}{2} = \frac{9+1}{2} = 5$

coefficient of Mean Deviation about the first 9 natural numbers =

$$\frac{\text{Mean deviation about A}}{\text{A}} \times 100 = \frac{\frac{20}{9}}{5} = \frac{4}{9} \times 100 = \frac{400}{9}$$

Example. 6 : . The mean deviation about the mode for the following observations 4/11, 6/11, 8/11, 9/11, ,12/11, 8/11 is

Solution: For the 4/11, 6/11, 8/11, 9/11, 12/11, 8/11 Mode is 8/11

$$\text{Mean deviation from Mode} = \frac{\sum|x_i - \text{Mode}|}{n}$$

$$= \frac{\left|\frac{4}{11}-\frac{8}{11}\right| + \left|\frac{6}{11}-\frac{8}{11}\right| + \left|\frac{8}{11}-\frac{8}{11}\right| + \left|\frac{9}{11}-\frac{8}{11}\right| + \left|\frac{12}{11}-\frac{8}{11}\right| + \left|\frac{8}{11}-\frac{8}{11}\right|}{6}$$

$$= \frac{\frac{4}{11}+\frac{2}{11}+0+\frac{1}{11}+\frac{4}{11}+0}{6} = \frac{\frac{11}{11}}{6} = \frac{1}{6}$$

Example 7: Find mean deviations about median and the corresponding coefficient for the following profits ('000₹) of a firm during a week. 82, 56, 75, 70, 52, 80, 68.

Solution: The profits in thousand rupees is denoted by x. Arranging the values of x in an ascending order, we get 52, 56, 68, 70, 75, 80, 82.

Therefore, Median = $\left(\frac{n+1}{2}\right)^{th} = \left(\frac{7+1}{2}\right)^{th}$ item = 4^{th} item = 70, thus,

Median profit = ₹70,000.

Computation of Mean deviation about median

x_i	52	56	68	70	75	80	82	Total
$\lvert x_i - Me\rvert$	18	14	2	0	5	10	12	61

Thus mean deviation about median $\frac{\sum|x_i - \text{Median}|}{n} = ₹\frac{61}{7} \times 1000 =$ ₹8714.29

$$\text{Coefficient of mean deviation} = \frac{\text{MD about median}}{\text{Median}} \times 100$$

$$= \frac{8714.29}{70000} \times 100 = 12.45$$

Example 8: Compute the mean deviation about the arithmetic mean for the following data:

Variable (x)	5	10	15	20	25	30
Frequency (f)	3	4	6	5	3	2

Solution: We are to apply formula as these data refer to a grouped frequency distribution the AM is given by

$$\bar{x} = \frac{\sum f_i x_i}{N} = \frac{5\times3+10\times4+15\times6+20\times5+25\times3+30\times2}{3+4+6+5+3+2} = 16.52$$

$$\text{Mean deviation from Mean} = \frac{\sum f_i.|x_i - \bar{x}|}{n} = \frac{139.56}{23} = 6.07$$

Coefficient of MD about its AM = $\frac{\text{MD about AM}}{\text{AM}} \times 100 = \frac{6.07}{16.52} \times 100 =$ 36.73

Example 9: The mean and SD for a, b and 2 are 3 and $\frac{2}{\sqrt{3}}$ respectively, The value of ab would be

Solution: Here the mean a, b and 2 ($\bar{x}$) = 3 , $\bar{x} = \frac{a+b+2}{3}$, 9 = a+b+2

Then a +b = 9-2= 7

$$\text{Standard deviation} = \sqrt{\frac{\sum x^2}{n} - \left(\bar{x}\right)^2}$$

$$\frac{2}{\sqrt{3}} = \sqrt{\frac{\sum x^2}{n} - (3)^2}$$

$$\Rightarrow \frac{4}{3} = \frac{\sum x^2}{3} - (3)^2$$

$$\frac{4}{3} = \frac{\sum x^2}{3} - 9$$

$$\sum x^2 = 27 + 4 = 31 \Rightarrow \sum x^2 = 31$$

$$a^2 + b^2 + 2^2 = 31$$

$$a^2 + b^2 = 31 - 4 = 27$$

$$(a+b)^2 - 2ab = a^2 + b^2$$

$(7)^2 - 2ab = 27$, 2ab = 49-27

2ab = 22

ab = 11

Example 10: If x and y are related as 4x+3y+11 = 0 and mean deviation of x is 5.40, what is the mean deviation of y?

Solution: Since 4x + 3y + 11 = 0

Therefore, $y = \left(\frac{-11}{3}\right) + \left(\frac{-4}{3}\right)x$

Hence MD of y= |b| × MD of x = $\frac{4}{3}$x5.40 = 7.20

Standard Deviation: Although mean deviation is an improvement over range so far as a measure of dispersion is concerned, mean deviation is difficult to compute and further more, it cannot be treated mathematically. The best measure of dispersion is, usually, standard deviation which does not possess the demerits of range and mean deviation.

Standard deviation for a given set of observations is defined as the root mean square deviation when the deviations are taken from the AM of the observations. If a variable x assumes n values x_1, x_2, x_3 ....x_n then its standard deviation(s) is given by

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

For a grouped frequency distribution, the standard deviation is given by

$$S = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{N}}$$

can be simplified to the following forms for unclassified data

$$S = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}$$

$$= \sqrt{\frac{\sum f_i x_i^2}{N} - \bar{x}^2} \text{ for a grouped frequency distribution.}$$

Variance: The square of standard deviation, known as variance

$$\text{Variance} = s^2 = \sqrt{\frac{\sum x_i - \bar{x}^2}{n}} \text{ for unclassified data}$$

$$= \sqrt{\frac{\sum f_i (x_i - \bar{x}^2)}{N}} \text{ for a grouped frequency distribution}$$

Coefficient of variation (CV) = $\frac{SD}{AM} \times 100$ (A relative measure of dispersion using standard deviation is given by Coefficient of Variation (CV) which is defined as the ratio of standard deviation to the corresponding arithmetic mean, expressed as a percentage.)

Example 11: Find the standard deviation and the coefficient of variation for the following numbers: 5, 8, 9, 2, 6

Solution: We present the computation in the following table:

Computation of standard deviation

x_i	5	8	9	2	6	$\Sigma x_i = 30$
x_i^2	25	64	81	4	36	$\Sigma x_i^2 = 210$

Applying, we get the standard deviation as

$$= \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2} = \sqrt{\frac{210}{5} - \left(\frac{30}{5}\right)^2} \qquad \left(\text{since } \bar{x} = \frac{\sum x_i}{n}\right)$$

$= \sqrt{42-36} = \sqrt{6} = 2.45$

The coefficient of variation is $CV = 100 \times \frac{SD}{AM} = 100 \times \frac{2.45}{6} = 40.83$

We consider the following formula for computing standard deviation from grouped frequency distribution with a view to saving time and computational labour:

$$S = \sqrt{\frac{\sum f d_i^2}{N} - \left(\frac{\sum f d_i}{N}\right)^2} \times C, \text{ Where } d_i = \frac{x_i - A}{C}$$

Properties of standard deviation

1. If all the observations assumed by a variable are constant i.e. equal, then the SD is zero. This means that if all the values taken by a variable x is k, say , then s = 0. This result applies to range as well as mean deviation.

2. SD remains unaffected due to a change of origin but is affected in the same ratio due to a change of scale i.e., if there are two variables x and y related as y = a+bx for any two constants a and b, then SD of y is given by $s_y = |b|\, s_x$

3. If there are two groups containing n_1 and n_2 observations, 1 and 2 as respective AM's, s_1 and s_2 as respective SD's, then the combined SD is given by

$$s = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}} \text{ where, } d_1 = \bar{x}_1 - \bar{x},\ d_2 = \bar{x}_2 - \bar{x} \text{ and}$$

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2} = \text{combined AM}$$

This result can be extended to more than 2 groups,we have

$$s = \sqrt{\frac{\sum n_i s_i^2 + \sum n_i d_i^2}{\sum n_i}} \text{ With } d = x_i - \bar{x} \text{ and } \bar{x} = \frac{\sum n_i \bar{x}_i}{\sum n_i}$$

Where $\bar{x}_1 = \bar{x}_2$ is reduced to $s = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}}$

4. For any two numbers a and b, standard deviation is given by $\frac{|a-b|}{2}$

5.SD of first n natural numbers is SD = $\frac{\sqrt{n^2 - 1}}{12}$

Example 12: If the S.D. of x is 3, what is the variance of (5 – 2x)?

Solution: If y = a + bx, then $\sigma_y = |b|\ \sigma_x$

Let y = 5 – 2x

$\therefore\ \sigma_y = |-2|\ \sigma x$

= 2 × 3 = 6

∴ Variance (5 – 2x) = $(2)^2$ × 9 = 36

Example 13: The coefficient of variation of the following numbers 53, 52, 61, 60 64, is

Solution: $\bar{x} = \frac{(53+52+61+60+64)}{5} = 58$

$$\therefore\ \sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

$$\therefore\ \sigma = \sqrt{\frac{(-5)^2 + (-6)^2 + 3^2 + 2^2 + 6^2}{5}} = 4.69.$$

Coefficient of variation = $\frac{\text{S.D.}}{\text{A.M.}} \times 100$

$= \frac{4.69}{58} \times 100 = 8.09.$

Example 14: What is the standard deviation of 5,5,9,9,9,10,5,10,10?

Solution: Mean = $\frac{3\times5+3\times9+3\times10}{9} = \frac{72}{9} = 8.$

Standard Deviation = $\sqrt{\frac{\sum f(x_i-\bar{x})^2}{n}} = \sqrt{\frac{3(9)+3(1)+3(4)}{9}} = \sqrt{42/9} = 2.16.$

Example 15: If x and y are related by 2x + 3y + 4 = 0 and S.D. of x is 6, then S.D. of y is :

Solution: $y = \frac{-2}{3}x - \frac{4}{3};$

$\sigma_y = |-2|\ 3|\ \sigma_x = (2/3)\ 6 = 4.$

Example 16: If x and y are related by y = 2x + 5 and the S.D. and A.M. of x are known to be 5 and 10 respectively, then the coefficient of variation of y is:

Solution: Y = 2x+5

$\sigma_y = |-2|\ \sigma_x = 2 \times 5 = 10.$

Also $\bar{y} = 2\bar{x} + 5 = 20+5 = 25$

Coefficient of variation of y = $\frac{\sigma_y}{\bar{y}}\times100 = \frac{10}{25}\times100 = 40.$

Example 17: If the mean and S.D. of x are a and b respectively, then the S.D. of $\frac{x-a}{b}$ is

Solution: Let $y = \frac{(x-a)}{b} = \frac{1}{b}.x - \frac{a}{b}$

$\sigma_y = \left|\frac{1}{b}\right| \sigma_x$

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

Example 18: If x and y are related by 3y = 7x – 9 and the S.D. of y is 7, then what is he variance of x?

Solution: 3y = 7x – 9

$X = \frac{3}{7}y + 9$

Also $\sigma_x = \left|\frac{3}{7}\right| \sigma_y = \frac{3}{7}\times7 = 3.$

$\therefore$ Variance : $\sigma_x^2 = 3^2 = 9$

Example 19: Which of the following companies A and B is more consistent so far as the payment of dividend is concerned ?

Dividend paid by A	5	9	6	12	15	10	8	10
Dividend paid by B:	4	8	7	15	18	9	6	6

Solution: Here $\sum x_A = 75$

$\therefore \bar{x}_A = 75/8 = 9.375$

$\sum x_A^2 = 775$

$\sigma_A^2 = \frac{\sum x_A^2}{N} - \left(\frac{\sum x_A}{N}\right)^2$

$= \frac{775}{8} - \left(\frac{75}{8}\right)^2 = 9$

$\sigma_A = 3.$

$C.V_A = \frac{\sigma_A}{\bar{x}}\times100 = \frac{3}{9.375}\times100 = 32.$

Also $\Sigma x_B = 73$, $\therefore\ \bar{x}_B = \frac{73}{8} = 9.125$

$\Sigma x_B^2 = 831,$

$\sigma_B^2 = \frac{831}{8} - \left(\frac{73}{8}\right)^2 = 20.61$

$\therefore\ \sigma = \sqrt{20.61} = 4.54$

$C.V_B = \frac{4.54}{9.125}\times100 = 49.75$

C.VA < C.VB

Company A is more consistent

Example 20: Find the SD of the following distribution:

Weight(kgs)	50-52	52-54	54-56	56-58	58-60
No.of Students	17	35	28	15	5

Solution:

Weight	No .of Students	Mid – Value	$d_i=x_i-55$	f_id_i	$f_id_i^2$
50-52	17	51	-2	-34	68
52-54	35	53	-1	-35	35
54-56	28	55	0	0	0
56-58	15	57	1	15	15
58-60	5	59	2	10	20
Total	100			-44	138

Applying, we get the SD of weight as

$= \sqrt{\frac{\sum f_id_i^2}{N} - \left(\frac{\sum f_id_i}{N}\right)^2} \times C = \sqrt{\frac{138}{100} - \frac{(-44)^2}{100}} \times 2\text{ kgs.} = \sqrt{1.38 - 0.1936} \times 2\text{ kgs.}$

=2.18 kgs

Example 21: The mean and variance of the 10 observations are found to be 17 and 33 respectively. Later it is found that one observation (i.e.26) is inaccurate and is removed. What is mean and standard deviation of remaining?

Solution: Mean of 10 observations = 17 then Total of the observations = 17 ×10 = 170

Total of the 9 observations = 170-26 = 144

Changed Mean = 144/9 = 16

Variance (σ^2) = 33

$\frac{\sum x^2}{n} - (17)^2 = 33 \Rightarrow \frac{\sum x^2}{10} = 33 + 289 = 322$

$\frac{\sum x^2}{10} = 322$

$\sum x^2 = 3220 - (26)^2 = 3220 - 676 = 2544$

Changed Variance = $\frac{\text{Changed} \sum x^2}{n} - (\text{Changed } \bar{x})^2$

$= \frac{2544}{9} - (16)^2 = 26.67$

SD of remaining observations = $\sqrt{26.67} = 5.16$

Example 22: If AM and coefficient of variation of x are 10 and 40 respectively, what is the variance of (15–2x)?

Solution: let y = 15 – 2x; AM of x= 10

Then applying formula, we get,

$s_y = 2 \times s_x$

As given cv_x = coefficient of variation of x = 40 and= 10

Thus $cv_x = \frac{S_x}{\bar{x}}\times100 \Rightarrow 40 = \frac{S_x}{10}\times100$

$\Rightarrow s_x = 4$

Then, $S_y = 2 \times 4 = 8$ Therefore, variance of (15-2x) = $S_y^2 = 64$

Example 23: Compute the SD of 9, 5, 8, 6, 2. Without any more computation, obtain the SD of

Sample I	-1	-5	-2	-4	-8
Sample II	90	50	80	60	20
Sample III	23	15	21	17	9

Solution:

x_i	9	5	8	6	2	30
x_i^2	81	25	64	36	4	210

The SD of the original set of observations is given by

$$s = \sqrt{\frac{210}{5} - \left(\frac{30}{5}\right)^2} = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2} = \sqrt{42-36} = \sqrt{6} = 2.45$$

If we denote the original observations by x and the observations of sample I by y, then we have

$y = -10 + x$

$y = (-10) + (1)\,x$

$\therefore S_y = |1| \times S_x = 1 \times 2.45 = 2.45$

In case of sample II, x and y are related as

$Y = 10x = 0 + (10)x$

$\therefore S_y = |10| \times S_x$

$= 10 \times 2.45 = 24.50$

And lastly, $y = (5) + (2)x$

$\Rightarrow S_y = 2 \times 2.45 = 4.90$

Example 24: For a group of 60 boy students, the mean and SD of stats. marks are 45 and 2 respectively. The same figures for a group of 40 girl students are 55 and 3 respectively. What is the mean and SD of marks if the two groups are pooled together?

Solution: As given $n_1 = 60$, $\bar{x}_1 = 45$, $s_1 = 2$ $n_2 = 40$, $\bar{x}_1 = \bar{x}_2 = 55$, $s_2 = 3$

Thus the combined mean is given by

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} = \frac{60\times45 + 40\times55}{60+40} = 49$$

Thus $d_1 = \bar{x}_1 = \bar{x} = 45 - 49 = -4$

$d_2 = \bar{x}_2 - \bar{x} = 55 - 49 = 6$

Applying formula, we get the combined SD as

$$s = \sqrt{\frac{n_1s_1^2 + n_2s_2^2 + n_1d_1^2 + n_2d_2^2}{n_1 + n_2}}$$

$$s = \sqrt{\frac{60\times2^2 + 40\times3^2 + 60\times(-4)^2 + 40\times6^2}{60+40}}$$

$= \sqrt{30}$

$= 5.48$

Example 25: The mean and standard deviation of the salaries of the two factories are provided below:

Factory	No. of Employees	Mean Salary	SD of Salary
A	30	₹4800	₹10
B	20	₹5000	₹12

(i) Find the combined mean salary and standard deviation of salary.

(ii) Examine which factory has more consistent structure so far as satisfying its employees are concerned.

Solution: Here we are given

$n_1 = 30$, $\bar{x}_1 = ₹4800$, $s_1 = ₹10$,

$n_2 = 20$, $\bar{x}_2 = ₹5000$, $s_2 = ₹12$

(i) Combined mean $= \dfrac{30\times ₹4800 + 20\times5000}{30+20}$ ₹4880

$d_1 = \bar{x}_1 = \bar{x} = ₹4{,}800 - ₹4880 = -₹80$

$d_2 = \bar{x}_2 - \bar{x} = ₹5{,}000 - ₹4880 = ₹120$

hence, the combined SD in rupees is given by

$$s = \sqrt{\frac{30\times10^2 + 20\times12^2 + 30\times(-80)^2 + 20\times120^2}{30+20}} = \sqrt{9717.60} = 98.58$$

thus the combined mean salary and the combined standard deviation of salary are ₹4880 and ₹98.58 respectively.

(ii) In order to find the more consistent structure, we compare the coefficients of variation of the two factories.

Letting $CV_A = 100\times\dfrac{S_A}{\bar{x}_A}$ and $CV_B = 100\times\dfrac{S_B}{\bar{x}_B}$

We would say factory A is more consistent

if $CV_A < CV_B$. Otherwise factory B would be more consistent.

Now $CV_A = 100\times\dfrac{S_1}{\bar{x}_1} = \dfrac{100\times10}{4800} = 0.21$

and $CV_B = 100\times\dfrac{S_2}{\bar{x}_2} = \dfrac{100\times12}{5000} = 0.24$

Thus we conclude that factory A has more consistent structure.

Quartile Deviation: Another measure of dispersion is provided by quartile deviation or semi - inter –quartile range which is given by

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

A relative measure of dispersion using quartiles is given by coefficient of quartile deviation which is

$$\text{Coefficient of quartile deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}\times100$$

Merits

1) Quartile deviation provides the best measure of dispersion for open-end classification.
2) It is also less affected due to sampling fluctuations.
3) Like other measures of dispersion, quartile deviation remains unaffected due to a change of origin but is affected in the same ratio due to change in scale.

Example 26: The quartiles of a variable are 45, 52, and 65 respectively. Its quartile deviation is:

Solution: Quartile Deviation $= \dfrac{Q_3 - Q_1}{2} = \dfrac{65 - 45}{2} = 10$

Example 27: If x and y are related as 3x + 4y = 20 and the quartile deviation of x is 12, then the quartile deviation of y is

Solution: If $y = ax + b$

Q.D of y = a × (Q.D. of x)

$3x + 4y = 20$

then $y = \dfrac{-3}{4}x + 5$

Q.D. of y = (3/4) (Q.D. of x)

$= |(-3/4)|\ 12 = 9.$

Example 28: Following are the marks of the 10 students : 56, 48, 65, 35, 42, 75, 82, 60, 55, 50. Find quartile deviation and also its coefficient.

Solution: After arranging the marks in an ascending order of magnitude, we get 35, 42, 48, 50, 55, 56, 60, 65, 75, 82

First quartile observation $Q_1 = \frac{(n+1)}{4}$th observation$=\frac{(10+1)}{4}$th observation

= 2.75^{th} observation

= 2^{nd} observation + 0.75 × difference between the third and the 2^{nd} observation.

= 42 + 0.75 × (48 – 42)

= 46.50

Third quartile (Q_3) = $\frac{3(n+1)}{4}$th observation

= 8.25^{th} observation

= 65 + 0.25 × 10

= 67.50

Thus applying, we get the quartile deviation as

$$\frac{Q_3 - Q_1}{2} = \frac{67.50 - 46.50}{2} = 10.50$$

Also, using the coefficient of quartile deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$

$$= \frac{67.50 - 46.50}{67.50 + 46.50} = 18.42$$

Example 29: If the quartile deviation of x is 6 and 3x + 6y = 20, what is the quartile deviation of y?

Solution: 3x + 6y = 20

$$y = \left(\frac{20}{6}\right) + \left(\frac{-3}{6}\right)x$$

Therefore, quartile deviation of y = $\frac{|-3|}{6}$ X quartile deviation of X

$$= \frac{1}{2} \times 6 = 3$$

Example 30: Find an appropriate measures of dispersion from the following data:

Daily wages (₹)	upto 20	20-40	40-60	60-80	80-100
No. of workers	5	11	14	7	3

Solution: Since this is an open-end classification, the appropriate measure of dispersion would be quartile deviation as quartile deviation does not taken into account the first twenty five percent and the last twenty five per cent of the observations.

Here a denotes the first Class Boundary

Table Computation of Quartile

Daily wages in ₹ (Class boundary)	No. of workers (less than cumulative frequency)
a	0
20	5
40	16
60	30
80	37
100	40

$$Q_1 = \left[20 + \frac{10-5}{16-5} \times 20\right] = ₹29.09$$

$$Q_3 = \left[40 + \frac{30-16}{30-16} \times 20\right] = 60$$

Thus quartile deviation of wages is given by = $\frac{Q_3 - Q_1}{2} = \frac{₹60 - ₹29.09}{2}$

= ₹ 15.46

Example 31: The mean and variance of 5 observations are 4.80 and 6.16 respectively. If three of the observations are 2, 3 and 6, what are the remaining observations?

Solution: Let the remaining two observations be a and b, then as given

$$\frac{2+3+6+a+b}{5}$$

⇨ 11+a+b =24 ⇨ a+b =13 .... (1)

and $\frac{2^2 + a^2 + b^2 + 3^2 + 6^2}{5} - (4.80)^2$ ⇨ $\frac{49^2 + a^2 + b^2}{5} - 23.04 = 6.16$

⇨ 49 + a^2 + b^2 =146

⇨ a^2 + b^2 =97 .... (2)

From (1), we get a = 13 – b .... (3)

Eliminating a from (2) and (3), we get

$(13 – b)^2 + b^2$ =97⇨ $169 – 26b + 2b^2$ =97

⇨ b^2 – 13 b + 36 = 0

⇨ (b–4)(b–9) =0

⇨ b = 4 or 9

From (3), a= 9 or 4

Thus the remaining observations are 4 and 9.

Example 32: If Standard deviation of x is σ , then standard deviation of $\frac{ax+b}{c}$, where a, b and c are constants, will be , then SD of y will be

Solution: SD of X = σ.

Let y = $\frac{ax+b}{c} = \frac{ax}{c} + \frac{b}{c}$

$y = \frac{b}{c} + \frac{ax}{c}$

SD of y = $\left|\frac{a}{c}\right|$SD of x = $\left|\frac{a}{c}\right|.\sigma$

Example 33: Find at the variance given arithmetic mean = $\frac{(8+4)}{2}$

Solution: Here Largest Value (L) = 8

Smallest Value (S) = 4

Range = Largest Value – Smallest Value = 8-4 = 4

We know that SD = $\frac{\text{Range}}{2} = \frac{4}{2} = 2$

Variance = $(SD)^2$= (2) =4

Example 34: If Mean and coefficient of variation of the marks of 10 students is 20 and 80, respectively. What will be the variance of them?

Solution: Given No. of observations (N) = 10

Mean ($\bar{x}$) = 20

CV = 80

$$CV = \frac{SD}{AM} \times 100$$

$$80 = \frac{SD}{20} \times 100$$

$$SD = \frac{80 \times 20}{100} = 16$$

Variance = $(SD)^2 = (16)^2 = 256$

Example 35: If arithmetic mean and coefficient of variation x are 10 and 40 respectively then variance of 15- $\frac{3x}{2}$ will be

Solution: Given Mean of x = 10 , Coefficient of Variation of (x) = 40

$$C\,V \text{ of } X = \frac{\text{SD of X}}{\text{Mean of X}} \times 100$$

400 = SD of x × 100

SD of x = $\frac{400}{100}$ = 4

Now y = 15 - $\frac{3x}{2}$

2y = -30+3x ∴ 2y= 3x-30

$\therefore y = \frac{3x}{2} - \frac{30}{2} \quad \therefore y = \frac{3x}{2} - 15$

∴3x-2y-30 = 0

S.D of y = $|b|$S.D of X

S.D of y = $\left|\frac{3}{2}\right| \times 4 = 6$

Variance of y $= (6)^2 = 36$

Example 36: Coefficient of Quartile deviation is ¼ then $\frac{Q_3}{Q_1}$ is

Solution: Coefficient of QD = $\frac{1}{4}$

$$\frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{1}{4}$$

$4Q_3 - 4Q_1 = Q_3 + Q_1$

$$\frac{Q_3}{Q_1} = \frac{5}{3}$$

Example 37: SD from numbers 1, 4, 5, 7, 8 is 2.45 . If 10 is added to each then SD will be

Solution: We know a change in origin of SD have no change in SD .
So , New SD = Original Sd when 10 will be added,
So , SD will not change .

Example 38: A student computes the AM and SD for a set of 100 observations as 50 and 5 respectively. Later on, she discovers that she has made a mistake in taking one observation as 60 instead of 50. What would be the correct mean and SD if

i) The wrong observation is left out?

ii) The wrong observation is replaced by the correct observation?

Solution: As given, n = 100, $\bar{x}$ = 50, S = 5
Wrong observation = 60(x), correct observation = 50(V)

$$\bar{x} = \frac{\sum x_i}{n}$$

$$\therefore \sum x_i = n\bar{x} = 100 \times 50 = 5000$$

$$\text{and } s^2 = \frac{\sum x_i^2}{n} - \bar{x}^2$$

$$\therefore \sum x_i^2 = n(\bar{x}^2 + s^2) = 100(50^2 + 5^2) = 252500$$

i) Sum of the 99 observations = 5000 – 60 = 4940
AM after leaving the wrong observation = 4940/99 = 49.90
Sum of squares of the observation after leaving the wrong observation
= 252500 – 60^2 = 248900
Variance of the 99 observations = 248900/99 – $(49.90)^2$
= 2514.14 – 2490.01
= 24.13
∴ SD of 99 observations = 4.91

ii) Sum of the 100 observations after replacing the wrong observation by the correct observation = 5000 – 60 + 50 =4990

AM = $\frac{4990}{100}$ = 49.90

Corrected sum of squares = 252500 + 50^2 – 60^2 = 251400

Corrected SD = $\sqrt{\frac{251400}{100} - (49.90)^2}$

= $\sqrt{23.94}$ = 4.90

Example 39: Compute coefficient of variation from the following data:

Age	under 10	under 20	under 30	under 40	under 50	under 60
No. of persons Dying	10	18	30	45	60	80

Solution: Given in this problem is less than cumulative frequency distribution. We need first convert it to a frequency distribution and then compute the coefficient of variation.

Table : Computation of coefficient of variation

Class Interval Age in years	No. of persons dying	Mid-value	$d_i = \frac{x_i - 25}{10}$	$f_i d_i$	$f_i d_i^2$
	(f_i)	(x_i)	10		
0-10	10	5	−2	−20	40
10-20	18−10= 8	15	−1	−8	8
20-30	30−18=12	25	0	0	0
30-40	45−30=15	35	1	15	15
40-50	60−45=15	45	2	30	60
50-60	80−60=20	55	3	60	180
Total	80	–	–	77	303

The AM is given by:

$$= \bar{x} = A + \frac{\sum f_i d_i}{N} \times C$$

$$= \left(25 + \frac{77}{80} \times 10\right) \text{ years}$$

= 34.63 years

The standard deviation is

$$s = \sqrt{\frac{\sum f_i d_i^2}{N} - \left(\frac{\sum f_i d_i}{N}\right)^2} \times C$$

$$= \sqrt{\frac{303}{80} - \left(\frac{77}{80}\right)^2} \times 10 \text{ years}$$

$$= \sqrt{3.79 - 0.93} \times 10 \text{ years}$$

= 16.91 years

Thus the coefficient of variation is given by

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

$$= \frac{16.91}{34.63} \times 100 = 48.83$$

Comparison between different measures of dispersion

We may now have a review of the different measures of dispersion on the basis of their relative merits and demerits.

1. Standard deviation, like AM, is the best measure of dispersion. It is rigidly defined, based on all the observations, not too difficult to compute, not much affected by sampling fluctuations and moreover it has some desirable mathematical properties. All these merits of standard deviation make SD as the most widely and commonly used measure of dispersion.
2. Range is the quickest to compute and as such, has its application in statistical quality control. However, range is based on only two observations and affected too much by the presence of extreme observation(s).
3. Mean deviation is rigidly defined, based on all the observations, and not much affected by sampling fluctuations. However, mean deviation is difficult to comprehend and its computation is also time consuming and laborious. Furthermore, unlike SD, mean deviation does not possess mathematical properties.
4. Quartile deviation is also rigidly defined, easy to compute and not much affected by sampling fluctuations. The presence of extreme observations has no impact on quartile deviation since quartile deviation is based on the central fifty-percent of the observations. However, quartile deviation is not based on all the observations and it has no desirable mathematical properties. Nevertheless, quartile deviation is the best measure of dispersion for open-end classifications.

CA FOUNDATION - PAPER 3 - BUSINESS MATHEMATICS, LOGICAL REASONING AND STATISTICS

At the Foundation level the concept of Probability is used in accounting and finance to understand the likelihood of occurrence or non-occurrence of a variable. It helps in developing financial forecasting in which you need to develop expertise at an advanced stage of chartered accountancy course. Here in this capsule an attempt is made for solving and understanding the concepts of probability.

Chapter 16 : Probability

The terms 'Probably' 'in all likelihood', 'chance', 'odds in favour', 'odds against' are too familiar nowadays and they have their origin in a branch of Mathematics, known as Probability. In recent time, probability has developed itself into a full-fledged subject and become an integral part of statistics.

Random Experiment: An experiment is defined to be random if the results of the experiment depend on chance only. For example if a coin is tossed, then we get two outcomes—Head (H) and Tail (T). It is impossible to say in advance whether a Head or a Tail would turn up when we toss the coin once. Thus, tossing a coin is an example of a random experiment. Similarly, rolling a dice (or any number of dice), drawing items from a box containing both defective and non—defective items, drawing cards from a pack of well shuffled fifty—two cards etc. are all random experiments.

Events: The results or outcomes of a random experiment are known as events. Sometimes events may be combination of outcomes. The events are of two types:
(i) Simple or Elementary,
(ii) Composite or Compound.
An event is known to be simple if it cannot be decomposed into further events. Tossing a coin once provides us two simple events namely Head and Tail. On the other hand, a composite event is one that can be decomposed into two or more events. Getting a head when a coin is tossed twice is an example of composite event as it can be split into the events HT and TH which are both elementary events.

Mutually Exclusive Events or Incompatible Events: A set of events A_1, A_2, A_3, ... is known to be mutually exclusive if not more than one of them can occur simultaneously. Thus, occurrence of one such event implies the non-occurrence of the other events of the set. Once a coin is tossed, we get two mutually exclusive events Head and Tail.

Exhaustive Events: The events A_1, A_2, A_3, ... are known to form an exhaustive set if one of these events must necessarily occur. As an example, the two events Head and Tail, when a coin is tossed once, are exhaustive as no other event except these two can occur.

Equally Likely Events or Mutually Symmetric Events or Equi-Probable Events: The events of a random experiment are known to be equally likely when all necessary evidence are taken into account, no event is expected to occur more frequently as compared to the other events of the set of events. The two events Head and Tail when a coin is tossed is an example of a pair of equally likely events because there is no reason to assume that Head (or Tail) would occur more frequently as compared to Tail (or Head).

CLASSICAL DEFINITION OF PROBABILITY OR A PRIORI DEFINITION
Let us consider a random experiment that result in n finite elementary events, which are assumed to be equally likely. We next assume that out of these n events, $n_A(\leq n)$ events are favourable to an event A. Then the probability of occurrence of the event A is defined as the ratio of the number of events favourable to A to the total number of events. Denoting this by P(A), we have

$$P(A) = \frac{n_A}{n} = \frac{\text{Number of equally likely events favourable to A}}{\text{Total Number of equally likely events}}$$

However, if instead of considering all elementary events, we focus our attention to only those composite events, which are mutually exclusive, exhaustive and equally likely and if $m(\leq n)$ denotes such events and is furthermore $m_A(\leq n_A)$ denotes the no. of mutually exclusive, exhaustive and equally likely events favourable to A, then we have

$$P(A) = \frac{m_A}{m} = \frac{\text{"Number of mutually exclusive,exhaustive and equally likely events favourable to A"}}{\text{"Total Number of mutually exclusive,exhaustive and equally likely events"}}$$

PROBABILITY AND EXPECTED VALUE BY MATHEMATICAL EXPECTATION
For this definition of probability, we are indebted to Bernoulli and Laplace. This definition is also termed as a priori definition because probability of the event A is defined based on prior knowledge.
This classical definition of probability has the following demerits or limitations:
(i) It is applicable only when the total no. of events is finite.
(ii) It can be used only when the events are equally likely or equi-probable. This assumption is made well before the experiment is performed.
(iii) This definition has only a limited field of application like coin tossing, dice throwing, drawing cards etc. where the possible events are known well in advance. In the field of uncertainty or where no prior knowledge is provided, this definition is inapplicable.
In connection with classical definition of probability, we may note the following points:
(a) The probability of an event lies between 0 and 1, both inclusive.
i.e. $0 \leq P(A) \leq 1$
When P(A) = 0, A is known to be an impossible event and when P(A) = 1, A is known to be a sure event.
(b) Non-occurrence of event A is denoted by A′ or A^C or and it is known as complimentary event of A. The event A along with its complimentary A′ forms a set of mutually exclusive and exhaustive events.
i.e. $P(A)+P(A')=1$, $P(A')=1-P(A) = 1 - \frac{m_A}{m} = \frac{m - m_A}{m}$

STATISTICS

Statistical definition of Probability: Owing to the limitations of the classical definition of probability, there are cases when we consider the statistical definition of probability based on the concept of relative frequency. This definition of probability was first developed by the British mathematicians in connection with the survival probability of a group of people.

Let us consider a random experiment repeated a very good number of times, say n, under an identical set of conditions. We next assume that an event A occurs F_A times. Then the limiting value of the ratio of F_A to n as n tends to infinity is defined as the probability of A.

i.e. $P(A) = \lim_{n \to \infty} \frac{F_A}{n}$

This statistical definition is applicable if the above limit exists and tends to a finite value.

Two events A and B are mutually exclusive if P (A ∩ B) = 0 or more precisely

P (A∪B) = P(A) + P(B)

Similarly, three events A, B and C are mutually exclusive if.

P (A∪B∪C) = P(A) + P(B) + P(C)

Two events A and B are exhaustive if.

P (A∪B) = 1

Similarly, three events A, B and C are exhaustive if.

P (A∪B∪C) = 1

Three events A, B and C are equally likely if

P(A) = P(B) = P(C)

Axiomatic or modern definition of probability: Then a real valued function p defined on s is known as a probability measure and p(a) is defined as the probability of A if P satisfies the following axioms:

(i) P(A) ≤ 0 for every A ⊆ S (subset)

(ii) P(S) = 1

(iii) For any sequence of mutually exclusive events A_1, A_2, A_3,..

$P(A_1 \cup A_2 \cup A_3 \cup ....) = P(A_1) + P(A_2) + P(A_3)$

Addition theorems or theorems on total probability: For any two mutually exclusive events A and B, the probability that either A or B occurs is given by the sum of individual probabilities of A and B.i.e. P (A∪B) or P(A + B) = P(A) + P(B) or P(A or B) whenever A and B are mutually exclusive

For any three events A, B and C, the probability that at least one of the events occurs is given by

P(A∪B∪C) = P(A) + P(B) + P(C) – P(A∩B) – P(A∩C) – P(B∩C)+ P(A∩B∩C)

(d) For any two events A and B, the probability that either A or B occurs is given by the sum of individual probabilities of A and B less the probability of simultaneous occurrence of the events A and B.

i. e. P(A∪B) = P(A) + P(B) – P(A∩B)

For any three events A, B and C, the probability that at least one of the events occurs is given by

P(A∪B∪C) = P(A) + P(B) + P(C) – P(A∩B) – P(A∩C) – P(B∩C)+ P(A∩B∩C)

(e) Two events A and B are mutually exclusive if

P (A∪B) = P(A) + P(B)

Similarly, three events A, B and C are mutually exclusive if (A∪B∪C) = P(A) + P(B) + P(C)

(f) P(A–B) = P (A∩B′) = P(A) – P(A∩B)

And P(B –A) = P(B ∩A′) = P(B) – P(A∩B)

Some important Results

1.If A and B are two independent events, then the probability of occurrence of both is given by P (A∩B) = P(A). P(B)

2.If A, B and C are three events, then. P(A∩B∩C) = P(A). P(B/A). P(C/A∩B)

3.If A and B are two mutually exclusive events of a random experiment, then.

A∩B = ϕ , P(AUB) = P(A) +P(B)

4. If A and B are associated with a random experiment, then.

P(AUB) = P(A) +P(B) -P(A∩B)

5. If A, B and C are three events connected with random experiment, then

P(AUBUC) = P(A) +P(B) +P(C) -P(A∩B) -P(B∩C)-P(C∩A) +P(A∩B∩C)

(g) Compound Probability or Joint Probability

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

(h) For any three events A, B and C, the probability that they occur jointly is given by

P(A∩B∩C) = P(A) P(B/A) P(C/(A∩B)) Provided P(A∩B) > 0

(i) $P(A'/B) = \frac{P(A' \cap B)}{P(B)} = \frac{P(B) - P(A \cap B)}{P(B)}$

(j) $P(A/B') = \frac{P(A \cap B')}{P(B')} = \frac{P(A) - P(A \cap B)}{1 - P(B)}$

(k) $P(A'/B') = \frac{P(A' \cap B')}{P(B')}$

$= \frac{P(A \cup B)'}{P(B')}$ [by De-Morgan's Law A′∩B′= (A∪B)′]

$= \frac{1 - P(A \cup B)}{1 - P(B)}$

CA FOUNDATION - PAPER 3 - BUSINESS MATHEMATICS, LOGICAL REASONING AND STATISTICS

This capsule is in continuation to the previous edition featured in March 2021. It presents the concepts of Random Variable, Expected value, Variance and Standard Deviation of a random variable. These concepts are extensively applied and widely used in areas such as Finance, Risk Management and Costing. Here an attempt is made to enable the students to understand these concepts of probability calculation with the help of examples and help them attempt diverse questions based on these concepts.

Chapter 16 : Probability - II

(7) **A random variable or stochastic variable** is a function defined on a sample space associated with a random experiment assuming any value from R and assigning a real number to each and every sample point of the random experiment.

(8) **Expected value or Mathematical Expectation or Expectation** of a random variable may be defined as the sum of products of the different values taken by the random variable and the corresponding probabilities.

When x is a discrete random variable with probability mass function f(x), then its expected value is given by

$$E(x) = \mu = \sum_x x f(x)$$

and its variance is

$$V(x) = \sigma^2 = E(x^2) - \mu^2$$

$$\text{Where } E(x^2) = \sum_x (x)^2 f(x)$$

For a continuous random variable x defined in [−∞, ∞], its expected value (i.e. mean) and variance are given by

$$E(x) = \int x f(x) dx$$

$$\text{and } \sigma^2 = E(x^2) - \mu^2$$

$$\text{where } E(x^2) = \int x^2 f(x) dx$$

Properties of Expected Values

1. Expectation of a constant is k
   i.e. E(k) = k for any constant k
2. Expectation of sum of two random variables is the sum of their expectations.
   i.e. E(x + y) = E(x) + E(y) for any two random variables x and y.
3. Expectation of the product of a constant and a random variable is the product of the constant and the expectation of the random variable.
   i.e. E(k x) = k.E(x) for any constant k
4. Expectation of the product of two random variables is the product of the expectation of the two random variables, provided the two variables are independent.
   i.e. E(xy) = E(x).E(y)
   Where x and y are independent.

IMPORTANT EXAMPLES:

1. A speaks truth in 60% and B in 75% of the cases. In what percentage of cases are they likely to contradict each other in stating the same fact?
   Solution:
   The Probability that A speaks the truth and B a lie =
   $$\frac{60}{100}\times\frac{(100-75)}{100} = \frac{60}{100}\times\frac{25}{100} = \frac{3}{20}$$
   The Probability that B speaks the truth and A a lie =
   $$\frac{75}{100}\times\frac{(100-60)}{100} = \frac{75}{100}\times\frac{40}{100} = \frac{3}{10}$$
   $$\therefore \text{ Total Probability} = \frac{3}{20}+\frac{3}{10}=\frac{9}{20}$$
   Hence, the percentage of cases in which they contradict each other = (9/20) × 100 or 45%

2. A Committee of 4 persons is to be appointed from 7 men and 3 women. The probability that the committee contains (i) exactly two women, and (ii) at least one woman is
   Solution:
   Total number of persons = 7+3 = 10. Since 4 out of them can be formed in 10 C_4 ways, where the exhaustive number of cases is 10 C_4 or 210 ways.
   (i) P (exactly 2 women in a committee) of four = 7 C_2X 3C_2/ 210 = 63/210 = 3/10.
   (ii) P (at least one women in committee)
   = 1−p (no women) =1-(7C_4/10C_4)=1-(35/210)=1-1/6=5/6

3. If A and B are two events, such that P(A) = 1/4, P(B) = 1/3 and P(A ∪ B) = 1/2; then P(B/A) is equal to
   Solution: $P(A \cup B)$ = P(A) + P (B) - P(A ∩ B)
   1/2 = 1/4 + 1/3 - P(A ∩ B)
   Or P(A ∩ B) = 1/4+1/3-1/2= 1/12
   $$\text{Hence, } P(B/A) = \frac{P(A\cap B)}{P(A)} = \frac{1/12}{1/4} = \frac{1}{3}$$

4. A person applies for a job in two firms, say X and Y. the probability of his being selected in firm X is 0.7 and being rejected in firm Y is 0.5. The probability of at least one of his applications being rejected is 0.6. What is the probability that he will be selected in one of the two firms?

Solution:
Event A; Person is selected in firm X, and
Event B : person is selected in Firm Y .
Then, $P(A) = 0.7$, $P(B^c) = 0.5$ and $P(A^c \cup B^c) = 0.6$
Therefore $P(B) = 1-0.5=0.5$
$P(A^c \cup B^c) = P[(A \cap B)^C] = 1-P(A \cap B)$
This implies that $P(A \cap B) = 1- P(A^c \cup B^c) = 1-0.6 = 0.4$
Hence, $P(A \cup B) = 0.7+0.5-0.4 = 0.8$

5. A person is known to hit a target in 5 out of 8 shots, whereas another person is known to hit in 3 out of 5 shots. Find the probability that the target is hit at all when they both try.

Solution: Event A = First person hits the target and
Event B = Another person hits the target.
$P(A) = 5/8$ and $P(B) = 3/5$
$P(A^C) = 1-5/8 = 3/8$ and $P(B^C) = 1-3/5 = 2/5$
Event X = target is hit when they both try i.e.,
When at least one of them hit the target.
$P(X^C)$ = P (the target is not hit at all)
$= P(A^C \cap B^C) = P(A^C) \times P(B^C) = 3/20$
Hence $P(X) = 1-P(X^C) = 1-3/20 = 17/20$

6. The probability that a man will be alive in 25 years is 3/5, and the probability that his wife will be alive in 25 years in 2/3. Find the probability that :
(i) Both will be alive (ii) at least one of the will be alive

Solution:
$P(M) = 3/5$ and $P(W) = 2/3$
$P(M^C) = 1 - 3/5$ and $P(W^C) = 1-2/3 = 1/3$.
The probability that both will be alive
$= P(M) \times P(W) = 3/5 \times 2/3 = 2/5$.
Probability that at least one of them will be alive is given by
$P(M \cup W) = P(M) + P(W) - P(M \cap W)$
$= 3/5+2/3-6/15 = 13/15$.

7. Given the data in Previous Problem find the probability that (i) only wife will be alive, (ii) only man will be alive.

Solution.
(i) Probability that only wife will be alive.
= Probability that wife will be alive but not man
$= P(W) \times P(M^C) = 2/3 \times 2/5 = 4/15$
(ii) Probability that only man will be alive
= Probability that man will be alive but not wife
$= P(M) \times P(W^C) = 3/5 \times 1/3 = 1/5$.

8. A random variable X has the following probability distribution:

Value of X	0	1	2	3
P[X = x]	1/3	1/2	0	1/6

Find $E(\{X - E(X)\}^2)$

Solution :
$E(X) = 0 \times 1/3+1 \times 1/2+2 \times 0+3 \times 1/6 = 1$
$E(X^2) = 0 \times 1/3+1 \times 1/2+4 \times 0+9 \times 1/6 = 2$
$E[X - E(X)]^2 = E(X^2) - [E(X)]^2 = 2-1 = 1$

9. Given the data in previous Problem, Find Var (Y), where Y = 2X – 1.

Solution:
$E(Y) = E(2X-1) = 2E(X) - 1 = 1$
$E(Y^2) = E(2x-1)^2 = 2E(X^2) - 4E(X)+1 = 1$
$Var.(Y) = E(Y^2) - [E(Y)]^2 = 1-1 = 0$

10 Daily demand for pen drive is having the following probability distribution. Determine the expected demand and variance of the demand:

Demand	1	2	3	4	5	6
Probability	0.10	0.15	0.20	0.25	0.18	0.12

Solution:
$E(X)=1 \times 0.10+2 \times 0.15+3 \times 0.20+4 \times 0.25+5 \times 0.18+6 \times 0.12 = 3.62$
Variance of the demand.
$E(X^2) = 1 \times 0.10+4 \times 0.15+9 \times 0.20+16 \times 0.25+25 \times 0.18+36 \times 0.12 = 15.32$
$Var.(X) = 15.32 - (3.62)^2 = 2.22$

11. An investment consultant predicts the odds against the price of a certain stock going up are 2:1 and odds in favor of the price remaining the same are 1:3. What is the price of stock will go down?
Solution:

P(the prices will go up) $= \frac{1}{2+1} = \frac{1}{3}$

P (the prices will remain the same) $= \frac{1}{1+3} = \frac{1}{4}$

Therefore P (the prices will go down)
= P (the price will neither go up nor remain same)
=1-P (the price will go up or will remain the same)

$$= 1 - \left(\frac{1}{3} + \frac{1}{4}\right) = 1 - \frac{7}{12} = \frac{5}{12}$$

12. A pair of dice is rolled. If the sum of the two dice is 9, find the probability that one of the dice shows 3
Solution:
Let A: Sum of on the two dice is 9. B: one of the dice showed 3.
Total outcomes when two dice are thrown = 6×6 = 36

$P(A) = P\{(6,3), (5,4), (4,5), (3,6)\} = \frac{4}{36}$

$P(A \cap B) = P\{(6,3), (3,6)\} = \frac{2}{36}$

Therefore, required probability = P(B/A) =

$$= \frac{P(A \cap B)}{P(A)} = \frac{2/36}{4/36} = \frac{2}{4} = \frac{1}{2}$$

13. The overall percentage of failures in a certain examination was 30. What is the probability that out of a group of 6 candidates at least four passed the examination?

Solution:

Let passing the examination be a success.

Take n= 6, P= P(a student passes) = $1-\frac{30}{100}=\frac{70}{100}=\frac{7}{10}$

q = P (a student fails) = $\frac{30}{100}=\frac{3}{10}$

Therefore P (at least 4 students pass) = P(4 or 5 or 6)

$$= 6_{C_4}\left(\frac{7}{10}\right)^4\left(\frac{3}{10}\right)^2 + 6_{C_5}\left(\frac{7}{10}\right)^5\left(\frac{3}{10}\right)^5 + 6_{C_6}\left(\frac{7}{10}\right)^6\left(\frac{3}{10}\right)^0$$

$$=15\left(\frac{7}{10}\right)^4\left(\frac{3}{10}\right)^2+6\left(\frac{7}{10}\right)^5\left(\frac{3}{10}\right)+\left(\frac{7}{10}\right)^6$$

$$=\left(\frac{7}{10}\right)^4\left[\frac{135}{100}+\frac{126}{100}+\frac{49}{100}\right]$$

$$=\frac{31}{10}\left(\frac{7}{10}\right)^4=0.74431.$$

14. What is the probability that a leap year selected at random would contain 53 Saturdays?

Solution:

A normal year has 52 Mondays, 52 Tuesdays, 52 Wednesdays, 52 Thursdays, 52 Fridays, 52 Saturdays and 52 Sundays

52 Saturdays, 52 + 1 day that could be anything depending upon the year under consideration.

- In addition to this, a leap year has an extra day which might be a Monday or Tuesday or Wednesday or Sunday.

Our sample space is S: {Monday-Tuesday, Tuesday-Wednesday, Wednesday-Thursday, Thursday-Friday, Friday-Saturday, Saturday-Sunday, Sunday-Monday} =Number of elements in S = n(S) = 7

set A (say) that comprises of the elements Friday-Saturday and Saturday-Sunday i.e. A : {Friday-Saturday, Saturday-Sunday}

Number of elements in set A = n(A) = 2,

By definition, probability of occurrence of A = n(A)/n(S) = 2/7

Therefore, probability that a leap year has 53 Saturdays is =2/7

15. If two unbiased coin is tossed three times, what is the probability of getting more than one head.

Solution:

One toss can give two (2) possible outcomes - head and tail. So, three tosses can give (2×2×2) = 8 possible outcomes.

2 heads and 1 tail out of 3 tosses can occur in $(3C_2)\times(1C_1) = 3$ ways.

So, the probability = (3/8).

16. If two unbiased are rolled, what is the probability of getting points neither 6 nor 9?

Solution:

Two dice can make 6 in 5 ways: {1,5}, {2,4}, {3,3}, {4,2} and {5,1}.

Two dice can make 9 in 4 ways: {3,6}, {4,5}, {5,4} and {6,3}.

There are 36 possible ways the two dice can fall. Therefore, the probability of 6 or 9 is (5+4)/36 = 1/4.

The probability of not (6 or 9) is therefore 1–1/4 = 3/4.

17. What is probability that 4 children selected at random would have different birthdays

Solution:

There are 365 out of 365 ways to select the birthday of first person. Therefore, the number of ways that we can choose a birthday for second person is 364 out of 365.

The probability that the second child has a different birthdate than the first is 364/365.

The probability that the third child has a different birthday than the first two is 363/365.

The probability that the fourth child has a different birthday than the first three is 362/365.

Since all three of these situations must occur, multiply the three probabilities.

364/365 × 363/365 × 362/365 = 98.364%

18. A box contains 5 white and 7 black balls. Two successive drawn of 3 balls are made (i) with replacement (ii) without replacement. The probability that the first draw would produce white balls and the second draw would produce black balls are respectively.

Solution:

Two successive drawn of 3 balls are made: TOTAL = 12 Balls: (5White balls +7 Black balls)

1st draw white ball and second draw black ball with replacement = $\frac{5_{C_3}}{12_{C_3}}\times\frac{7_{C_3}}{12_{C_3}}=\frac{10}{220}\times\frac{35}{220}=\frac{7}{968}$

1st draw white ball and second draw black ball without replacement = $\frac{5_{C_3}}{12_{C_3}}\times\frac{7_{C_3}}{9_{C_3}}=\frac{10}{220}\times\frac{35}{84}=\frac{5}{264}$

P (both happening) = $\frac{7}{968}$ and $\frac{5}{264}$

19. There are three boxes with the following composition:

Box I: 5 Red + 7 White + 6 Blue balls

Box II: 4 Red + 8 White + 6 Blue balls

Box III: 3 Red + 4 White + 2 Blue balls

If one ball is drawn at random, then what is the probability that they would be of same colour?

Solution:

Either balls would be Red or white or blue

$=P(R_1\cap R_2\cap R_3) + P(W_1\cap W_2\cap W_3) + P(B_1\cap B_2\cap B_3)$

$=P(R_1)\times P(R_2)\times P(R_3)+P(W_1)\times P(W_2)\times P(W_3)+P(B_1)\times P(B_2)\times P(B_3)$

$=\frac{5}{18}\times\frac{4}{18}\times\frac{3}{9}$(Red Balls) + $\frac{7}{18}\times\frac{8}{18}\times\frac{4}{9}$ (White Balls) + $\frac{6}{18}\times\frac{6}{18}\times\frac{2}{9}$ (Black Balls)

$=\frac{89}{729}$

20. A number is selected at random from the first 1000 natural numbers. What is the probability that the number so selected would be a multiple of 7 or 11?

Solution:

First 1000 natural numbers belong to the following set {1,2, 3,...,1000} with cardinality = 1000

Multiples of 7 less than 1000 = Quotient of (1000/7) = 142

Multiples of 11 less than 1000 = Quotient of (1000/11) = 90

As 7 & 11 are both primes so multiples of 7*11 = 77 will be included in both multiples of 7 and multiples of 11

Multiples of 77 less than 1000 = Quotient of (1000/77) = 12
Hence, all-natural numbers below 1000 which are either multiples of 7 or of 11 = 142 + 90–12 = 220
So, Prob (this event) = 220/1000 = 0.22

21. A bag contains 8 red and 5 white balls. Two successive draws of 3 balls are made without replacement. The probability that the first draw will produce 3 white balls and the second 3 red balls is

Solution:
There are total 13 balls out of which 8 are red and 5 are white.
Favourable case of first draw is to get 3 white balls out of 5 white balls.
Probability $P_1 = 5C_3/13C_3 = 5/143$
If this happens then remaining are - 2 white balls and 8 red balls.
Favourable case is to get 3 red balls out of 8 balls.
Probability $P_2 = 8C_3/10C_3 = 7/15$
Both the events are independent of each other, hence total probability is $P_1 * P_2 = (5C_3 \times 8C_3)/(13C_3 \times 10C_3) = 5/143 \times 7/15 = 7/429$

22. There are two boxes containing 5 white and 6 blue balls and 3 white and 7 blue balls respectively. If one of the the boxes is selected at random and a ball is drawn from it, then the probability that the ball is blue is

Solution:
First box: No. of white balls = 5, No. of blue balls = 6
Second box: No. of white balls = 3, No. of blue balls = 7
So, total no. of white balls = 8, Total no. of blue balls = 13
So, total no. of balls = 8+13=21
Now probability of getting blue ball: =13/21
Hence the probability of getting blue ball is =13/21

23. A problem in probability was given to three CA students A, B and C whose chances of solving it are 1/3, 1/5 and 1/2 respectively. What is the probability that the problem would be solved?

Solution:
Probability of A solving the problem = 1/3, Probability of A not solving the problem = 1-1/3 = 2/3
Probability of B solving the problem = 1/5, Probability of B not solving the problem = 1-1/5 = 4/5
Probability of C solving the problem = ½, Probability of C not solving the problem = 1-1/2 = ½
Probability of A, B and C not solving the problem = 2/3 x 4/5 x 1/2 = 4/15
Probability of A, B and C solving the problem = 1-4/15 = 11/15

24. There are three persons aged 60, 65 and 70 years old. The survival probabilities for these three persons for another 5 years are 0.7, 0.4 and 0.2 respectively. What is the probability that at least two of them would survive another five years?

Solution:
Probability (At least two alive) = P (two alive) +P (two alive)
=(0.7)(0.4)(1-0.2)+(0.7)(0.2)(1-0.4)+(0.4)(0.2)(1-0.7) +(0.7(0.4)(0.2)
$=0.28 \times 0.8 + 0.14 \times 0.6 + 0.08 \times 0.3 + 0.056 = 0.224 + 0.084 + 0.056 = 0.388$

25. Tom speaks truth in 30 percent cases and Dick speaks truth in 25 percent cases. What is the probability that they would contradict each other?

Solution:
P(Tom speaks truth)= $P(T_T)$ = 30/100= 3/10;
P(Dick speaks truth)= $P(D_T)$ = 25/100= ¼
$P(T_F)$ = 1-3/10= 7/10; $P(D_F)$ = 1-1/4= ¾
probability that they would contradict each other
$= P(T_T) \times P(D_F) + P(T_F) \times P(D_T)$
= (3/10×3/4) + (7/10×1/4) = 0.40

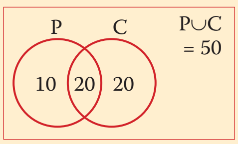
26. There are two urns. The first urn contains 3 red and 5 white balls whereas the second urn contains 4 red and 6 white balls. A ball is taken at random from the first urn and is transferred to the second urn. Now another ball is selected at random from the second urn. The probability that the second ball would be red is

Solution:
the first urn contains 3 red and 5 white balls => total = 8 Balls
second urn contains 4 red and 6 white balls. => Total = 10 Balls
There can be two cases : Ball taken from Urn is Red or White
Case 1 : Red is Taken from Urn:A
Probability of Red = (3/8)
then second urn contains 5 Red & 6 White => total = 11
Probability of Red from Urn:B = (3/8) (5/11) = 15/88
Probability of White from Urn:B = (3/8) (6/11) = 18/88
Case 1 : White is Taken from Urn:A
Probability of White = (5/8)
then second urn contains 4 Red & 7 White => total = 11
Probability of Red from Urn:B = (5/8) (4/11) = 20/88
Probability of White from Urn:B = (5/8) (7/11) = 35/88
probability of that the second ball would be Red = 15/88 + 20/88 = 35/88
probability of that the second ball would be White = 18/88 + 35/88 = 53/88

27. For a group of students, 30%, 40% and 50% failed in Physics, Chemistry and at least one of the two subjects, respectively. If an examinee is selected at random, what is the probability that he passed in Physics if it is known that he failed in Chemistry?

Solution:
Let the total number of students = 100
Number of students failed in physics = 30% of 100 =30
Number of students failed in chemistry = 40% of 100 =40
Number of students failed at least one of the two subjects = 50% of 100 =50
We need to calculate.
P (He passed in Physics but Failed in Chemistry)/P (Failed Chemistry)
$n(P \cup C) = n(p) + n(c) - n(P \cap C) 50 = 30+40- n(P \cap C)$
$n(P \cap C) = 20$
P (He passed in Physics but Failed in Chemistry)/P (failed Chemistry)

$$= P(C-P)/(P(P) = \frac{\frac{20}{100}}{\frac{40}{100}} = \frac{20}{40} = \frac{1}{2}$$



28. A packet of 10 electronic components is known to include 2 defectives. If a sample of 4 components is selected at random from the packet, what is the probability that the sample does not contain more than 1 defective?

Solution:
P (No more than one defective) = P (No defective) + P (One defective)

$= \frac{{}^{8}C_4}{{}^{10}C_4} + \frac{{}^{2}C_1 \times {}^{8}C_3}{{}^{10}C_4}$

$= \frac{70}{210} + \frac{112}{210}$

$= \frac{91}{105} = \frac{13}{15}$

29. 8 identical balls are placed at random in three bags. What is the probability that the first bag will contain 3 balls?

Solution:
The probability that a ball will be placed in the first bag is 1/3. The probability that exactly 3 of the 8 balls will end up in the first bag can be found by using the binomial distribution:

$${}^{8}C_3 \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^5 = \frac{{}^{8}C_3 \times 2^5}{3^8} = \frac{56 \times 32}{6561} = 0.2731$$

30. X and Y stand in a line with 6 other people. What is the probability that there are 3 persons between them?

Solution:
There are altogether 8 people $p_1, p_2, \ldots, p_8$ including X & Y and these 8 people can be arranged in 8!= 40320 ways

Now, there should be 3 people between X and Y and these 3 people can be selected out of 6 in C(6,3)=20 =20ways. (6 people because X & Y are excluded from 8).

Now, take (X,*,*,*,Y)(X,*,*,*,Y)as one set of people and together with the remaining 3 people we can think of a total of 4 people which can be arranged in 4!=24 ways.

Again, the 3 people between X & Y can be arranged in 3!=6 ways.
Also, the position of X and Y can also be arranged in 2!=2ways.
So, total arrangements with 3 people between X & Y is 20×24×6× 2=5760

Hence, the required probability is =5760/40320=1/7.

31. Given that P (A) = 1/2, P (B) = 1/3, P (AB) = 1/4, what is P (A'/B')

Solution:

$$P\left(\frac{A'}{B'}\right) = \frac{P(A' \cap B')}{P(B')} = \frac{P(AUB)'}{1-P(B)} = \frac{1-P(AUB)}{1-P(B)} = \frac{1-\left(\frac{1}{2}+\frac{1}{3}-\frac{1}{4}\right)}{1-\frac{1}{3}} = \frac{1-\left(\frac{6+4-3}{12}\right)}{\frac{3-1}{3}}$$

$$= \frac{1-\frac{7}{12}}{\frac{2}{3}} = \frac{\frac{5}{12}}{\frac{2}{3}} = \frac{5}{12} \times \frac{3}{2} = \frac{5}{8}$$

32. Four digits 1, 2, 4 and 6 are selected at random to form a four-digit number. What is the probability that the number so formed, would be divisible by 4?

Solution:
From four digits 1,2,4 and 6 last two digits (12,16,24,64) can be selected in (4 ways)

Total possible numbers are divisible by 4 are 4612, 6412, 2416, 4216, 1624, 6124, 1264, 2164 =8

Here are four ways of filling the last two digits. The remaining two places(100's, 1000's digits) can be filled in two ways. Thus there are total 4×2=8 ways

Total possible 4 digit numbers =4! =24

$\text{Probability} = \frac{8}{4!} = \frac{8}{24} = \frac{1}{3}$

33. A bag containing 6 white and 4 red balls. Rs 10 is received if he draws white ball and Rs. 20 for red ball. Find the expected amount when the person draws 2 balls.

Solution:
The probability of both being white ball would be $= \frac{{}^{6}C_2}{{}^{10}C_2} = \frac{15}{45}$

The probability of both being red ball would be $= \frac{{}^{4}C_2}{{}^{10}C_2} = \frac{6}{45}$

The probability of one being red ball and another being white ball would be $= \frac{{}^{6}C_1 \times {}^{4}c_1}{{}^{10}C_2} = \frac{24}{45}$

Hence the expected amount when person draws two balls will be

$$= \frac{15}{45} \times (10+10) + \frac{6}{45}(20+20) + \frac{24}{45}(10+20)$$

= 6.7+5.33+16 = Rs.28

34. If two random variables x and y are related as Y =-3x +4 and Standard Deviation of Y is

Solution:
Given Y =-3x +4
Y = a + bx

$\sigma_y = \sigma(ax+b)$, $\sigma(b) = 0$

$\sigma_y = |a| . \sigma_x$

SD of y=$\sigma_y = 3 \times 2 = 6$

35. If 2x+3y+ 4 = 0 and v(x) = 6 then V(y) is

Solution:
Given that 2x+3y+ 4 = 0 and v(x) = 6 then SD of x = √6:V(y) = ?

$$SD_y = \left|-\frac{2}{3}\right| . SD_x = \frac{2}{3}\sqrt{6}$$

$$y^2 = \frac{4}{9} . 6 = \frac{8}{3}$$

$$V(y) = \frac{8}{3}$$

36. A pocket of 10 electronic components is known to include 3 defectives. If 4 components are selected from the packet at random, what is the expected value of the number of defective?

Solution:
10 electronic components If 4 components are selected from the packet at random = $\frac{4}{10}$

Expected value of the number of defective = $3 \times \frac{4}{10} = \frac{12}{10} = 1.2$

37. The Probability there is atleast one error in a account statement prepared by 3 persons A , B and C are 0.2 , 0.3 and 0.1 respectively . If A, B and C prepare 60, 70 and 90 such statements then expected number of correct statements.

Solution:
E(x)=A.(1-P(A))+B.(1-P(B))+C.(1-P(C))=(60×(1-0.2) + (70×(1-0.3) + (90×(1-0.1)
= (60×0.8) +(70×0.70) +(90×0.9) =178

CA FOUNDATION - PAPER 3 - BUSINESS MATHEMATICS, LOGICAL REASONING AND STATISTICS

This capsule on Foundation - Paper 3 - Business Mathematics, Logical Reasoning and Statistics , will enable the students to understand and apply the techniques of developing discrete and continuous probability distributions.

Chapter 17 : Theoretical Distributions

In this chapter we will discuss the probability theory by considering a concept and analogous to the idea of frequency distribution. In frequency distribution where to the total frequency to different class intervals, the total probability (i.e. one) is distributed to different mass points is known as theoretical probability distributions.

- Discrete Random variable.
- Continues Random variable.

Importance of theoretical probability distribution.☛:

(a) An observed frequency distribution, in many a case, may be regarded as a sample i.e. a representative part of a large, unknown, boundless universe or population and we may be interested to know the form of such a distribution.
For Example: By fitting a theoretical probability distribution.
- Length of life of the lamps produced by manufacturer up to a reasonable degree of accuracy.
- The effect of a particular type of missiles, it may be possible for our scientist to suggest the number of such missiles necessary to destroy an army position.
- By knowing the distribution of smokers, a social activist may warn the people of a locality about the nuisance of active and passive smoking and so on.

(b) Theoretical probability distribution may be profitably employed to make short term projections for the future.

(c) Statistical analysis is possible only on the basis of theoretical probability distribution.

A probability distribution also possesses all the characteristics of an observed distribution. We define population mean, population median, population mode, population standard deviation etc. exactly same way we have done earlier. These characteristics are known as population parameters.

A probability distribution may be either a discrete probability distribution or a Continuous probability distribution depending on the random variable under study.

Two important discrete probability distributions

(a) Binomial Distribution
(b) Poisson distribution.

Important continuous probability distribution

Normal Distribution

Binomial Distribution

One of the most important and frequently used discrete probability distribution is Binomial Distribution. It is derived from a particular type of random experiment known as Bernoulli process named after the famous mathematician Bernoulli. Noting that a 'trial' is an attempt to produce a particular outcome which is neither certain nor impossible, the characteristics of Bernoulli trials are stated below:

(i) Each trial is associated with two mutually exclusive and exhaustive outcomes, the occurrence of one of which is known as a 'success' and as such its non occurrence as a 'failure'. As an example, when a coin is tossed, usually occurrence of a head is known as a success and its non–occurrence i.e. occurrence of a tail is known as a failure.

(ii) The trials are independent.

We may note the following important points in connection with binomial distribution:

(a) As $n > 0, p, q \geq 0$, it follows that $f(x) \geq 0$ for every x
Also $\sum f(x) = f(0) + f(1) + f(2) + \ldots.. + f(n) = 1$

(b) Binomial distribution is known as biparametric distribution as it is characterised by two parameters n and p. This means that if the values of n and p are known, then the distribution is known completely.

(c) The mean of the binomial distribution is given by $\mu = np$

(d) Depending on the values of the two parameters, binomial distribution may be unimodal or bi- modal., the mode of binomial distribution, is given by μ_0=the largest integer contained in (n+1)p
if (n+1)p is a non-integer (n+1)p and (n+1)p - 1 if (n+1)p is an integer

(e) The variance of the binomial distribution is given by
$\sigma^2 = npq$
Since p and q are numerically less than or equal to 1, npq < np variance of a binomial variable is always less than its mean.
Also variance of X attains its maximum value at p = q = 0.5 and this maximum value is n/4.

(f) Additive property of binomial distribution.
If X and Y are two independent variables such that
$X \sim \beta(n_1, P)$
and $Y \sim \beta(n_2, P)$
Then $(X+Y) \sim (n_1 + n_2, P)$

Applications of Binomial Distribution

Binomial distribution is applicable when the trials are independent and each trial has just two outcomes success and failure. It is applied in coin tossing experiments, sampling inspection plan, genetic experiments and so on.

Poisson Distribution

Poisson distribution is a theoretical discrete probability distribution which can describe many processes. Simon Denis Poisson of France introduced this distribution way back in the year 1837.

Poisson Model

Let us think of a random experiment under the following conditions:

I. The probability of finding success in a very small time interval (t, t + dt) is kt, where k (>0) is a constant.

II. The probability of having more than one success in this time interval is very low.

III. The probability of having success in this time interval is independent of t as well as earlier successes.

The above model is known as Poisson Model. The probability of getting x successes in a relatively long time interval T containing m small time intervals t i.e. T = mt. is given by

$$\frac{e^{-kt}.(kt)^x}{x!} \quad \text{for } x = 0, 1, 2, \ldots\ldots\ldots\ldots$$

Taking kT = m, the above form is reduced to

$$\frac{e^{-m}.m^x}{x!} \quad \text{for } x = 0, 1, 2, \ldots\ldots\ldots\ldots$$

Definition of Poisson Distribution

A random variable X is defined to follow Poisson distribution with parameter λ, to be denoted by X ~ P (m) if the probability mass function of x is given by

$$f(x) = P(X = x) = \frac{e^{-m}.m^x}{x!} \text{ for } x = 0, 1, 2, \ldots$$

$$= 0 \quad \text{otherwise}$$

Here e is a transcendental quantity with an approximate value as 2.71828.

Important points in connection with Poisson distribution:

(i) **Since $e^{-m} = 1/e^m > 0$, whatever may be the value of m, m > 0, it follows that f (x)≥ 0 for every x.**
Also it can be established that $\sum_x f(x) = 1$ i.e. f(0) + f(1) + f(2) +....... = 1

(ii) **Poisson distribution is known as a uniparametric distribution as it is characterised by only one parameter m.**

(iii) **The mean of Poisson distribution is given by m i,e μ = m**

(iv) **The variance of Poisson distribution is given by $\sigma^2 = m$**

(v) **Like binomial distribution, Poisson distribution could be also unimodal or bimodal depending upon the value of the parameter m.**
We have μ_0 = The largest integer contained in m if m is a non-integer
= m and m–1 if m is an integer

(vi) **Poisson approximation to Binomial distribution**
If n, the number of independent trials of a binomial distribution, tends to infinity and p, the probability of a success, tends to zero, so that m = np remains finite, then a binomial distribution with parameters n and p can be approximated by a Poisson distribution with parameter m (= np).
In other words when n is rather large and p is rather small so that m = np is moderate then β(n, p) ≅ P (m)

(vii) **Additive property of Poisson distribution**
If X and y are two independent variables following Poisson distribution with parameters m_1 and m_2 respectively, then Z = X + Y also follows Poisson distribution with parameter ($m_1 + m_2$).
i.e. if X ~ P (m_1)
and Y ~ P (m_2)
and X and Y are independent, then
Z = X + Y ~ P ($m_1 + m_2$)

Application of Poisson distribution

Poisson distribution is applied when the total number of events is pretty large but the probability of occurrence is very small. Thus we can apply Poisson distribution, rather profitably, for the following cases:

a) The distribution of the no. of printing mistakes per page of a large book.

b) The distribution of the no. of road accidents on a busy road per minute.

c) The distribution of the no. of radio-active elements per minute in a fusion process.

d)

Normal or Gaussian distribution

The two distributions discussed so far, namely binomial and Poisson, are applicable when the random variable is discrete. In case of a continuous random variable like height or weight, it is impossible to distribute the total probability among different mass points because between any two unequal values, there remains an infinite number of values. Thus a continuous random variable is defined in term of its probability density function f (x), provided, of course, such a function really exists, f (x) satisfies the following condition:

$$f(x) \geq 0 \text{ for } x(-\infty,\infty) \text{ and } = \int_{-\infty}^{+\infty} f(x) = 1$$

The most important and universally accepted continuous probability distribution is known as normal distribution. Though many mathematicians like De-Moivre, Laplace etc. contributed towards the development of normal distribution, Karl Gauss was instrumental for deriving normal distribution and as such normal distribution is also referred to as Gaussian Distribution.

A continuous random variable x is defined to follow normal distribution with parameters μ and σ^2, to be denoted by

$$X \sim N(\mu, \sigma^2)$$

If the probability density function of the random variable x is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}.e^{-(\bar{x}-\mu)^2/2\sigma^2} \quad \text{for } -\infty<x<\infty$$

where μ and σ are constants, and > 0

CA FOUNDATION - PAPER 3: BUSINESS MATHEMATICS, LOGICAL REASONING AND STATISTICS

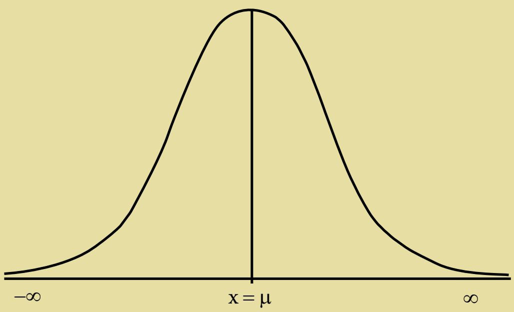
This capsule is in continuation to the previous edition featured in November 2021. Further here presented properties of Normal Distribution and their applications. Here an attempt is made to the enable the students to understand the concepts Binomial, Poisson and Normal distribution with the help of examples.

Chapter 17 : Theoretical Distributions

Some important points relating to normal distribution are listed below:

(a) The name Normal Distribution has its origin some two hundred years back as the then mathematician were in search for a normal model that can describe the probability distribution of most of the continuous random variables.

(b) If we plot the probability function y = f (x), then the curve, known as probability curve, takes the following shape:



Mean = Median = Mode

Showing Normal Probability Curve. A quick look at figure reveals that the normal curve is bell shaped and has one peak, which implies that the normal distribution has one unique mode. The line drawn through x = μ has divided the normal curve into two parts which are equal in all respect. Such a curve is known as symmetrical curve and the corresponding distribution is known as symmetrical distribution. Thus, we find that the normal distribution is symmetrical about x = μ. It may also be noted that the binomial distribution is also symmetrical about p = 0.5. We next note that the two tails of the normal curve extend indefinitely on both sides of the curve and both the left and right tails never touch the horizontal axis. The total area of the normal curve or for that any probability curve is taken to be unity i.e. one. Since the vertical line drawn through x = μ divides the curve into two equal halves, it automatically follows that,
The area between – ∞ to μ= the area between μ to ∞ = 0.5
When the mean is zero, we have
the area between – ∞ to 0 = the area between 0 to ∞ = 0.5

(c) If we take μ = 0 and σ = 1 , we have

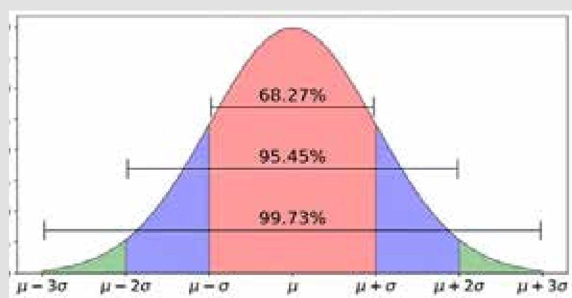
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-2^2/2} \qquad \text{for } -\infty < z < \infty$$

The random variable z is known as standard normal variate (or variable) or standard normal deviate. The probability that a standard normal variate X would take a value less than or equal to a particular value say X = x is given by
φ(x) = p (X ≤ x)
(x) is known as the cumulative distribution function.
We also have (0) = P (X ≤ 0) = Area of the standard normal curve between –∞ and 0 = 0.5

(d) **The normal distribution is known as biparametric distribution as it is characterised by two parameters μ and σ^2. Once the two parameters are known, the normal distribution is completely specified.**

Properties of Normal Distribution

1. Since π = 22/7 , $e^{-\theta} = 1 / e^{\theta} > 0$, whatever θ may be,
it follows that f (x) 0 for every x.
It can be shown that $\int_{-\infty}^{\infty} f(x)dx = 1$
2. The mean of the normal distribution is given by μ. Further, since the distribution is symmetrical about x = μ, it follows that the mean, median and mode of a normal distribution coincide, all being equal to μ.
3. The standard deviation of the normal distribution is given by
Mean deviation of normal distribution is $\sigma \frac{\sqrt{2}}{\pi}$
The first and third quartiles are Q_1 = μ– 0.675 σ and Q_3 = μ+ 0.675 σ
so that, quartile deviation = 0.675 σ
4. The normal distribution is symmetrical about x =μ. As such, its skewness is zero i.e. the normal curve is neither inclined move towards the right (negatively skewed) nor towards the left (positively skewed).
5. The normal curve y = f (x) has two points of inflexion to be given by x = μ – σ and
x = μ + σ i.e. at these two points, the normal curve changes its curvature from concave to convex and from convex to concave.
6. If x ~ N (μ, σ^2) then z = x – μ/σ ~ N (0, 1), z is known as standardised normal variate or normal deviate.
We also have P (z ≤ k) = ϕ (k) The values of ϕ(k) for different k are given in a table known as "Biometrika."
7. Area under the normal curve is shown in the following figure:



(z=3) (z=-2) (z=-1) (z=0) (z=1) (z=2) (z=3)

Area Under Normal Curve

From the figure , we have

$P(\mu - \sigma < x < \mu + \sigma) = 0.6828$

=> $P(-1 < z < 1) = 0.6828$

$P(\mu - 2\sigma < x < \mu + 2\sigma) = 0.9546$

=> $P(-2 < z < 2) = 0.9546$

and $P(\mu - 3\sigma < x < \mu + 3\sigma) = 0.9973$

=> $P(-3 < z < 3) = 0.9973$.

We note that 99.73 per cent of the values of a normal variable lies between $(\mu - 3\sigma)$ and $(\mu + 3\sigma)$. Thus the probability that a value of x lies outside that limit is as low as 0.0027.

8. If x and y are independent normal variables with means and standard deviations as μ_1 and μ_2 and σ_1, and σ_2 respectively, then $z = x + y$ also follows normal distribution with mean $(\mu_1 + \mu_2)$ and SD = $\sqrt{\sigma_1^2 + \sigma_2^2}$ respectively.

   i.e. If $x \sim N(\mu_1, \sigma_1^2)$ and $y \sim N(\mu_2, \sigma_2^2)$ and x and y are independent,

   then $z = x + y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Applications of Normal Distribution

Most of the continuous variables like height, weight, wage, profit etc. follow normal distribution. If the variable under study does not follow normal distribution, a simple transformation of the variable, in many a case, would lead to the normal distribution of the changed variable.

When n, the number of trials of a binomial distribution, is large and p, the probability of a success, is moderate i.e. neither too large nor too small then the binomial distribution, also, tends to normal distribution. Poisson distribution, also for large value of m approaches normal distribution. Such transformations become necessary as it is easier to compute probabilities under the assumption of a normal distribution.

Some Important Problems

1. What is the probability that out of 10 missiles fired, atleast 2 will hit the target

Solution: Probability of atleast 2 will hit the target is given as,

$P(X \geq 2) = 1 - P(X < 2)$

The probability of a missile hitting a target is 1 /8

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

$= 1 - [{}^{10}C_0 \times (1/8)^0 \times (7/8)^{10} + {}^{10}C_1 \times (1/8)^1 \times (7/8)^9]$

$= 1 - [7^{10}/8^{10} + (10 \times 7^9)/8^{10}]$

$= 1 - [7^{10} + 10 \times 7^9] / 8^{10}$

$= 1 - (17 \times 7^9) / 8^{10}$

$P(X \geq 2) \approx 0.3611 = 36.11\%$

Therefore, the probability that out of 10 missiles fired, atleast 2 will hit the target is 0.3611.

2. Given X is a binomial variable such that 2 P(X = 2) = P(X = 3) and mean of X is known to be 10/3. What would be the probability that X assumes at most the value 2

Solution:

mean = 10/3 , Mean = np = 10/3

$P(X) = {}^nC_x p^x (1-p)^{n-x}$

=> $P(2) = {}^nC_2 p^2 (1-p)^{n-2}$

$P(3) = {}^nC_3 p^3 (1-p)^{n-3}$

$P(3) = 2\,P(2)$

=> ${}^nC_3 p^3 (1-p)^{n-3} = 2\,{}^nC_2 p^2 (1-p)^{n-2}$

=> $p/3!(n-3)! = (2/2!(n-2)!)(1 - p)$

=> $p/6 = (1-p)/(n - 2)$

=> $np - 2p = 6 - 6p$

=> $4p = 6 - np$

=> $4p = 6 - 10/3$

=> $4p = 8/3$

=> $p = 2/3$

$np = 10/3$ => $n = 5$

probability that X assumes at most the value 2 = P(0) + P(1) + P(2)

$= {}^5C_0(2/3)^0(1/3)^5 + {}^5C_1(2/3)^1(1/3)^4 + {}^5C_2(2/3)^2(1/3)^3$

$= 1/243 + 10/243 + 40/243 = 51/243 = 17/81$

3. Assuming that one-third of the population is tea drinkers and each of 1000 enumerators takes a sample of 8 individuals to find out whether they are tea drinkers or not, how many enumerators are expected to report that five or more people are tea drinkers?

Solution: Assume that being a tea drinker is like taking a flip of coin.(i.e. either the person drinks tea or not) with the probability of heads being 1/3 and tails being 2/3

now $P(X\geq5) = P(X=5)+P(X=6)+P(X=7)+P(X=8) = (56*8+28*4+8*2+1)/6561 = 577/6561 = 0.087943$ so for 1000 trials the number of trials which report greater than or equal to 5 is $1000*P(X>=5) = 87.94 = 88$

4. If a random variable x follows binomial distribution with mean as 5 and satisfying the condition 10. P (x= 0) = P(x= 1) , what is the value of $P(X\geq1/X>1)$?

Solution:

Here Mean np = 5

$10.P(x = 0) = P(x = 1)$

$10.{}^nC_0.p^0.(1-p)^n = {}^nC_1.p.(1-p)^{n-1}$

$10.(1-p)^n = np(1-p)^{n-1}$

$10(1-p) = 5$

$1 - p = 1/2$

$p = 1/2, n = 10$

$$p\left(x \geq 1 \middle/ x \geq 0\right) = \frac{p(x \geq 1)}{p(x \geq 0)}$$

$= 1 - p(x = 0)$

$$= 1 - \frac{1}{2^{10}}$$

$= 0.99$

5. Out of 128 families with four children each, how many are expected to have atleast have one boy and one girl?

Solution: 4 children in a family can be in

2 * 2 * 2 * 2 = 16 ways

at least one boy and one girl = Total cases - all boys - all girs

All boys = 1 case, All girls = 1 case

=> at least one boy and one girl = 16 - 1 - 1 = 14

Probability of at least one boy and one girl = 14/16

out of 128 families expected to have = 128 * 14/16 = 8 * 14 = 112

112 Families expected to have at least one boy and one girl

6. In 10 independent rollings of a biased die, the probability that an even number will appear 5 times is twice the probability that an even number will appear 4 times. What is the probability that an even number will appear twice when the die is rolled 8 times?

Solution: Probability of even number p = p

Then probability of odd number (or not even number) = q = 1 - p

Probability Appearing 5 times

${}^{10}C_5 * p^5 * q^{10-5}$

$= {}^{10}C_5 * p^5 * q^5$

Probability Appearing 5 times

$^{10}C_4 * p^4 * q^{10-4}$

$= {}^{10}C_4 * p^4 * q^6$

$^{10}C_5 * p^5 * q^5 \;\; = 2 \; * {}^{10}C_4 * p^4 * q^6$

=>p * 10!/5!5! = 2q * 10!/6!4!

=> p * 6 = 2q * 5

=> 3p = 5q => 3p = 5(1 - p)

=> 8p = 5

=> p = 5/8 & q = 1 - 5/8 = 3/8

probability that an even number will appear twice when the die is rolled 8 times

$= {}^8C_2 * p^2q^{8-2}$

$= 28 \; * (5/8)^2 \, (3/8)^6$

$= 28 * 25 * 3^6 / 8^8$

$= 700 * 3^6 / 8^8$

= 5,10,300/1,67,77,216 = 0.0304

7. What is the probability of making 3 correct guesses in 5 true-False answer type questions?

Solution: Here P = 0.5 and q = 0.5 (Since answer can be either True or False) n= 5, r = 3

$P(x=r) = {}^nC_r \, p^x.q^{n-x}$

$P(x=3) = {}^5C_3 \, (0.5)^{5}.(0.5)^{5-3}$ =0.3125

8. Suppose that weather records show that on an average 5 out of 31 days in October are rainy days. Assuming a binomial distribution which each day of October as an independent trail, then the probability that the next October will have at most three rainy days is:

Solution: p = Probability of a rainy day in October :p = 5/31, q = probability of a non-rainy day in October q = 1- p = 1-5/31 = 26/31

n = 31 (number of days in October)

$$P(x) = 31Cr.\left(\frac{5}{31}\right)^r.\left(\frac{26}{31}\right)^{31-r}$$

Required Probability = P (0) + P (1) +P(2) +P(3)

$$= 31C_0.\left(\frac{5}{31}\right)^0.\left(\frac{26}{31}\right)^{31} + 31C_1.\left(\frac{5}{31}\right)^1.\left(\frac{26}{31}\right)^{31-1} + 31C_2.\left(\frac{5}{31}\right)^2.\left(\frac{26}{31}\right)^{31-2} + 31C_3.\left(\frac{5}{31}\right)^3.\left(\frac{26}{31}\right)^{31-3}$$

= 0.2403

9. If 5 days are selected at random, then the probability of getting two Sundays is:

Solution: Let P = Probability of getting a Sunday in a week (P) = 1/7

Therefore P = 1/7 and q= 1-p = 1-1/7 = 6/7

$$\text{Required probability} = 15C2\,.\left(\frac{1}{7}\right)^2.\left(\frac{6}{7}\right)^{15-2} = 0.288 = 0.29$$

10. An experiment of succeeds twice as often as it falls. What is the probability that in next five trials there will be at least three successes?

Solution: According to the given statement p = 2q

We know that p = 2/3 q= 1/3

Required probability P(X≥3) = P(3) +P(4) +P(5)

$$= 5C3\,.\left(\frac{2}{3}\right)^2.\left(\frac{6}{7}\right)^{15-2} + 5C4\,.\left(\frac{2}{3}\right)^2.\left(\frac{6}{7}\right)^{15-2} + 5C5\,.\left(\frac{2}{3}\right)^3.\left(\frac{1}{3}\right)^2$$

11. What is the probability of getting 3 head if 6 unbiased coins are tossed simultaneously?

Solution: if x denotes the number of heads, then x follows binomial distribution with parameters n = 6 and p = probability of success = ½

q= probability of failure = 1-1/2 = ½ , being given the coins are unbiased

The probability mass function of x is given by

$$f(x) = {}^6C_3\,.\left(\frac{1}{2}\right)^3.\left(\frac{1}{2}\right)^3 = 20\times\left(\frac{1}{2}\right)^3 = 0.3125$$

12. In Binomial Distribution n = 9 and p = 1/3 , What is the value of variance .

Solution: In Binomial Distribution variance = npq , here n= 9 , p = 1/3 and q= 2/3

Therefore variance = 9*1/3*2/3 = 2

13. For a Binomial Distribution E(x) = 2 , V(X)= 4/3 . Find the value of n

Solution: Here E(X) = np = 2

V(x)= npq = 4/3 then substituting the value of np

2×q = 4/3

2q = 4/3

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

np = 2, n×1/3 = 2 , n= 6

14. The mode of Binomial Distribution for which the mean is 4 and variance 3 is equal to?

Solution: In Binomial Distribution Mean = np = 4 and Variance = npq = 3

Then 4q = 3 q = ¾ , p = 1-q = 1-3/4 = ¼

n×1/4 = 4 therefore n = 16

(n+1) p = (16+1) ×1/4 = 4.25 which is no longer integer.

So mode = 4

15. In a Binomial Distribution with 5 independent trials, probability of 2 and 3 successes are 0.4362 and 0.2181 respectively. Parameter 'p 'of the Binomial Distribution is

Solution: Given n = 5 , P(x=2) = 0.4362

P(x=3) =0.2181

$P(x=3) = 5C3\,.(p)^3.(q)^{5-3} = 10\,(p)^3.(q)^2$

$0.2181 = 10.P^{3}.q^{2}$

And $P(X=2) = 5C_2\,.(p)^3.(q)^{5-2} = 10\,(p)^2.(q)^3$

$0.4362 = 10\,(p)^2.(q)^3$

By dividing

$$0.2181/0.4362 = \frac{10.p^3.q^2}{10.p^2.q^3}$$

½= q/p

q= 2p; 2p+p = 1

3p = 1, then p = 1/3

16. What is the first quartile of X having the following probability density function?

$$f(x) = \frac{1}{\sqrt{72\pi}}\,e^{\frac{-(x-10)^2}{72}} \qquad \text{for} - \propto < x < \propto$$

Solution: First Quartile Deviation $(Q_1) = \mu - 0.675\sigma$ here

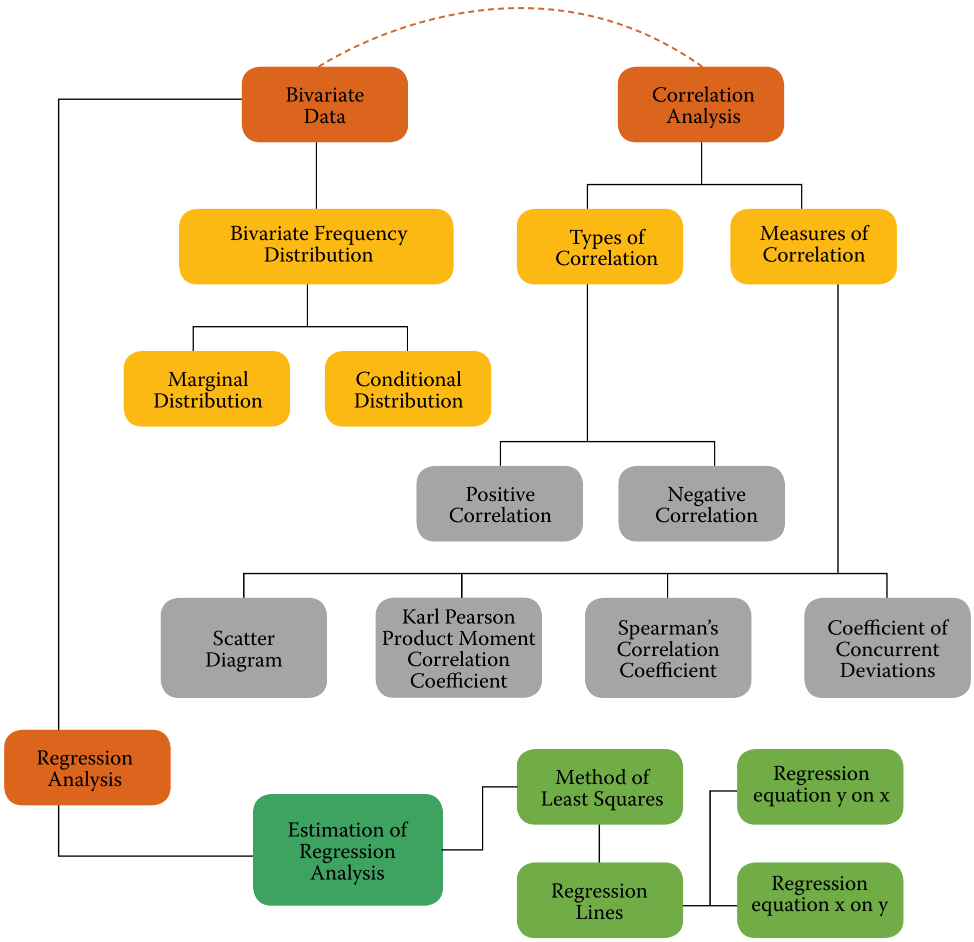
By comparing probability density function of Normal Distribution, Mean = 10 and σ = 4

Q_1 = 10-0.675(6) =10-4.05 = 5.95

CA FOUNDATION - PAPER 3 - BUSINESS MATHEMATICS, LOGICAL REASONING AND STATISTICS

At the foundation level with regards to Paper 3: Business Mathematics, Logical Reasoning and Statistics, Chapter 17: Correlation and Regression is very important for students not only to acquire professional knowledge but also for examination point of view. Here in this capsule an attempt is made for solving and understanding the concepts of Correlation and Regression.

CHAPTER 17 OVERVIEW: CORRELATION AND REGRESSION



Univariate Distribution: Statistical measure relating to Univariate distribution i.e. distribution of one variable like height, weight, mark, profit, wage and so on. However, there are situations that demand study of more than one variable simultaneously. A businessman may be keen to know what amount of investment would yield a desired level of profit or a student may want to know whether performing better in the selection test would enhance his or her chance of doing well in the final examination. With a view to answering this series of questions, we need to study more than one variable at the same time.

Bivariate Data: When data are collected on two variables simultaneously, they are known as bivariate data and the corresponding frequency distribution, derived from it, is known as Bivariate Frequency Distribution. If x and y denote marks in Maths and Stats for a group of 30 students, then the corresponding bivariate data would be (x_i, y_i) for i= 1, 2, ..., 30 where (x_1, y_1) denotes the marks in Mathematics and Statistics for the student with serial number or Roll Number 1, (x_2, y_2), that for the student with Roll Number 2 and so on and lastly (x_{30}, y_{30}) denotes the pair of marks for the student bearing Roll Number 30.

Correlation Analysis and Regression Analysis are the two analyses that are made from a multivariate distribution i.e. a distribution of more than one variable. In particular, when there are two variables, say x and y, we study bivariate distribution. We restrict our discussion to bivariate distribution only.

Correlation analysis helps us to find an association or the lack of it between the two variables x and y. Thus, if x and y stand for profit and investment of a firm or the marks in Statistics and Mathematics for a group of students, then we may be interested to know whether x and y are associated or independent of each other. The extent or amount of correlation between x and y is provided by different measures of Correlation namely Product Moment Correlation Coefficient or Rank Correlation Coefficient or Coefficient of Concurrent Deviations. In Correlation analysis, we must be careful about a cause-and-effect relation between the variables under consideration because there may be situations where x and y are related due to the influence of a third variable although no causal relationship exists between the two variables.

Regression analysis, on the other hand, is concerned with predicting the value of the dependent variable corresponding to a known value of the independent variable on the assumption of a mathematical relationship between the two variables and also an average relationship between them.

As in the case of a Univariate Distribution, we need to construct the frequency distribution for bivariate data. Such a distribution takes into account the classification in respect of both the variables simultaneously. Usually, we make horizontal classification in respect of x and vertical classification in respect of the other variable y. Such a distribution is known as Bivariate Frequency Distribution or Joint Frequency Distribution or Two way classification of the two variables x and y. Frequency Distribution, we can obtain two types of univariate distributions which are known as:

(a) Marginal distribution: Marginal distributions always divide the column or row totals by the table total

(b) Conditional distribution: To calculate a conditional distribution, you must first establish a condition. For instance, we could ask what the distribution of gender is among students who watched the last football game. So, the condition here would be that the student watched the game. In particular, if there are m classifications for x and n classifications for y, then there would be altogether (m + n) conditional distribution.

Correlation Analysis: While studying two variables at the same time, if it is found that the change in one variable is reciprocated by a corresponding change in the other variable either directly or inversely, then the two variables are known to be associated or correlated. Otherwise, the two variables are known to be dissociated or uncorrelated or independent. There are two types of correlation.

Positive correlation

If two variables move in the same direction i.e. an increase (or decrease) on the part of one variable introduces an increase (or decrease) on the part of the other variable, then the two variables are known to be positively correlated.

For example, height and weight yield and rainfall, profit and investment etc. are positively correlated.

Negative correlation

If the two variables move in the opposite directions i.e. an increase (or a decrease) on the part of one variable results a decrease (or an increase) on the part of the other variable, then the two variables are known to have a negative correlation.

The price and demand of an item, the profits of Insurance Company and the number of claims it has to meet etc. are examples of variables having a negative correlation.

The two variables are known to be uncorrelated if the movement on the part of one variable does not produce any movement of the other variable in a particular direction. As for example, Shoe-size and intelligence are uncorrelated

Measures of correlation:

- Scatter diagram
- Karl Pearson's Product moment correlation coefficient
- Spearman's rank correlation coefficient
- Coefficient of concurrent deviations

(a) SCATTER DIAGRAM: This is a simple diagrammatic method to establish correlation between a pair of variables. Unlike product moment correlation coefficient, which can measure correlation only when the variables are having a linear relationship, scatter diagram can be applied for any type of correlation – linear as well as non-linear i.e. curvilinear. Scatter diagram can distinguish between different types of correlation although it fails to measure the extent of relationship between the variables.

Each data point, which in this case a pair of values (x_i, y_i) is represented by a point in the rectangular axes of cordinates. The totality of all the plotted points forms the scatter diagram. The pattern of the plotted points reveals the nature of correlation. In case of a positive correlation, the plotted points lie from lower left corner to upper right corner, in case of a negative correlation the plotted points concentrate from upper left to lower right and in case of zero correlation, the plotted points would be equally distributed without depicting any particular pattern. The following figures show different types of correlation and the one-to-one correspondence between scatter diagram and product moment correlation coefficient.

Figure 1: Showing Positive Correlation

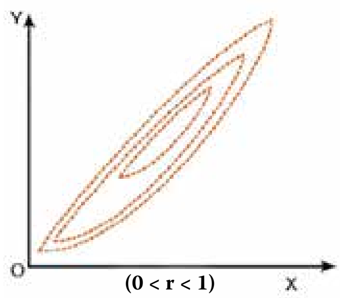


Figure 2: Showing Perfect Correlation

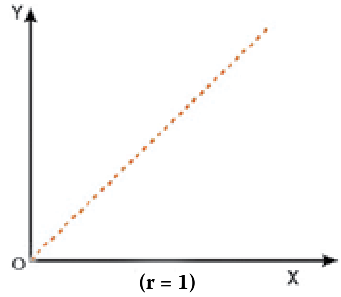


Figure 3: Showing Negative Correlation

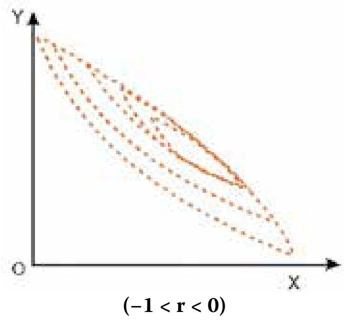


Figure 4: Showing Perfect Negative Correlation

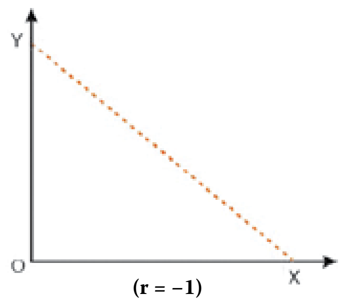


Figure 5: Showing No Correlation

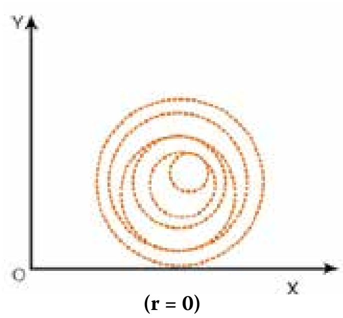
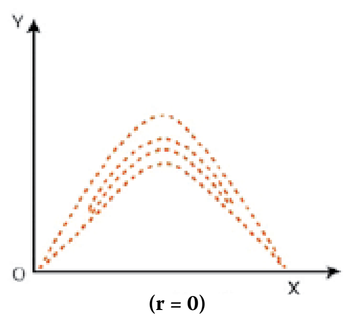


Figure 6: Showing Curvilinear Correlation



(b) KARL PEARSON'S PRODUCT MOMENT CORRELATION COEFFICIENT

This is by far the best method for finding correlation between two variables provided the relationship between the two variables is linear. Pearson's correlation coefficient may be defined as the ratio of covariance between the two variables to the product of the standard deviations of the two variables. If the two variables are denoted by x and y and if the corresponding bivariate data are (x_i, y_i) for i = 1, 2, 3, ….., n, then the coefficient of correlation between x and y, due to Karl Pearson, is given by

$$r = r_{xy} = \frac{Cov(x,y)}{S_x \times S_y}$$

where , $$cov(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n} = \frac{\sum x_i y_i}{n} - \bar{x}\,\bar{y}$$

$$S_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2} \text{ and } S_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n}} = \sqrt{\frac{\sum y_i^2}{n} - \bar{y}^2}$$

A single formula for computing correlation coefficient is given by

$$r = \frac{n\sum x_i y_i - \sum x_i \times \sum y_i}{\sqrt{n\sum x_i^2 - \left(\sum x_i\right)^2}\sqrt{n\sum y_i^2 - \left(\sum y_i\right)^2}}$$

In case of a bivariate frequency distribution, we have

$$\text{Cov}(x,y) = \frac{\sum_{i,j} x_i y_j f_{ij}}{N} - \bar{x} \times \bar{y}$$

$$S_x = \sqrt{\frac{\sum_i f_{io} x_i^2}{N} - \bar{x}^2} \text{ and } S_y = \sqrt{\frac{\sum_j f_{oj} y_j^2}{N} - \bar{y}^2}$$

where x_i = Mid-value of the i[th] class interval of x. ,y_j =Mid-value of the j[th] class interval of y
f_{io} = Marginal frequency of x f_{oj} =Marginal frequency of y
f_{ij} = frequency of the (i, j)[th] cell , N = Total frequency

PROPERTIES OF CORRELATION COEFFICIENT

(i) **The Coefficient of Correlation is a unit-free measure:** This means that if x denotes height of a group of students expressed in cm and y denotes their weight expressed in kg, then the correlation coefficient between height and weight would be free from any unit.

(ii) **The coefficient of correlation remains invariant under a change of origin and/or scale of the variables under consideration depending on the sign of scale factors.**

This property states that if the original pair of variables x and y is changed to a new pair of variables u and v by effecting a change of origin and scale for both x and y i.e.

$$u = \frac{x-a}{b} \text{ and } v = \frac{y-c}{d}$$

where a and c are the origins of x and y and b and d are the respective scales and then we have

$$r_{xy} = \frac{bd}{|b||d|} r_{uv}.$$

r_{xy} and r_{uv} being the coefficient of correlation between x and y and u and v respectively, the two correlation coefficients remain equal and they would have opposite signs only when b and d, the two scales, differ in sign.

(iii) **The coefficient of correlation always lies between –1 and 1, including both the limiting values i.e.** $-1 \le r \le 1$

(c) SPEARMAN'S RANK CORRELATION COEFFICIENT: When we need finding correlation between two qualitative characteristics, say, beauty and intelligence, we take recourse to using rank correlation coefficient. Rank correlation can also be applied to find the level of agreement (or disagreement) between two judges so far as assessing a qualitative characteristic is concerned. As compared to product moment correlation coefficient, rank correlation coefficient is easier to compute, it can also be advocated to get a first hand impression about the correlation between a pair of variables.

Spearman's rank correlation coefficient is given by

$$r_R = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

where r_R denotes rank correlation coefficient and it lies between –1 and 1 inclusive of these two values.

$d_i = x_i - y_i$ represents the difference in ranks for the i-th individual and n denotes the number of individuals.

In case u individuals receive the same rank, we describe it as a tied rank of length u. In case of a tied rank, formula is changed to

$$r_R = 1 - \frac{6\left[\sum_i d_i + \sum_j \frac{(tj^3 - t_j)}{12}\right]}{n(n^2-1)}$$

In this formula, t_j represents the j[th] tie length and the summation extends over the lengths of all the ties for both the series.

(d) COEFFICIENT OF CONCURRENT DEVIATIONS

A very simple and casual method of finding correlation when we are not serious about the magnitude of the two variables is the application of concurrent deviations. This method involves in attaching a positive sign for a x-value (except the first) if this value is more than the previous value and assigning a negative value if this value is less than the previous value. This is done for the y-series as well. The deviation in the x-value and the corresponding y-value is known to be concurrent if both the deviations have the same sign.

Denoting the number of concurrent deviation by c and total number of deviations as m (which must be one less than the number of pairs of x and y values), the coefficient of concurrent deviation is given by

$$r_C = \pm\sqrt{\pm\frac{(2c-m)}{m}}.$$

If (2c–m) >0, then we take the positive sign both inside and outside the radical sign and if (2c–m) <0, we are to consider the negative sign both inside and outside the radical sign.

Like Pearson's correlation coefficient and Spearman's rank correlation coefficient, the coefficient of concurrent deviations also lies between –1 and 1, both inclusive.

Spurious Correlation: There are some cases when we may find a correlation between two variables although the two variables are not causally related. This is due to the existence of a third variable which is related to both the variables under consideration. Such a correlation is known as spurious correlation or non-sense correlation. As an example, there could be a positive correlation between production of rice and that of iron in India for the last twenty years due to the effect of a third variable time on both these variables. It is necessary to eliminate the influence of the third variable before computing correlation between the two original variables.

Correlation Coefficient: Correlation Coefficient measuring a linear relationship between the two variables indicates the amount of variation of one variable accounted for by the other variable. A better measure for this purpose is provided by the square of the correlation coefficient, known as 'coefficient of determination'. This can be interpreted as the ratio between the explained variance to total variance i.e.

$$r^2 = \frac{\text{Explained variance}}{\text{Total variance}}$$

Thus, a value of 0.6 for r indicates that $(0.6)^2 \times 100\%$ or 36 per cent of the variation has been accounted for by the factor under consideration and the remaining 64 per cent variation is due to other factors.

Coefficient of non-determination: The 'coefficient of non-determination' is given by $(1-r^2)$ and can be interpreted as the ratio of unexplained variance to the total variance.

Coefficient of non-determination = $(1-r^2)$

CA FOUNDATION - PAPER 3 - BUSINESS MATHEMATICS, LOGICAL REASONING AND STATISTICS

This capsule is in continuation to the previous edition featured in June 2022. Further here presented properties of Regression and their applications. Here an attempt is made to the enable the students to understand the Correlation and Regression with the help of examples.

CHAPTER 17: CORRELATION AND REGRESSION - PART 2

Regression Lines: (i) The two lines of regression coincide i.e. become identical when r = –1 or 1 or in other words, there is a perfect negative or positive correlation between the two variables. (ii) If r = 0, Regression lines are perpendicular to each other.

Regression Analysis: In regression analysis, we are concerned with the estimation of one variable for a given value of another variable (or for a given set of values of a number of variables) on the basis of an average mathematical relationship between the two variables (or a number of variables). Regression analysis plays a very important role in the field of every human activity. A businessman may be keen to know what would be his estimated profit for a given level of investment on the basis of the past records. Similarly, an outgoing student may like to know her chance of getting a first class in the final University Examination on the basis of her performance in the college selection test.

When there are two variables x and y and if y is influenced by x i.e., if y depends on x, then we get a simple linear regression or simple regression. y is known as dependent variable or regression or explained variable and x is known as independent variable or predictor or explanator. In the previous examples since profit depends on investment or performance in the University Examination is dependent on the performance in the college selection test, profit or performance in the University Examination is the dependent variable and investment or performance in the selection test is the independent variable.

In case of a simple regression model if y depends on x, then the regression line of y on x in given by

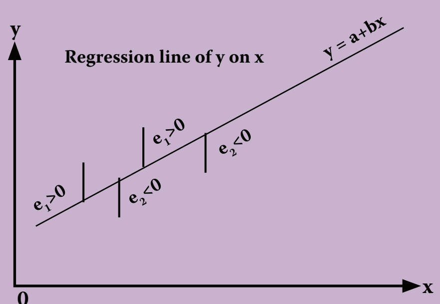
$$y = a + bx$$

Here, a and b are two constants and they are also known as regression parameters. Furthermore, b is also known as the regression coefficient of y on x and is also denoted by b_{yx}. We may define the regression line of y on x as the line of best fit obtained by the method of least squares and used for estimating the value of the dependent variable y for a known value of the independent variable x.

The method of least squares involves in minimising

$$\Sigma e_i^2 = \Sigma (y_i^2 - y^\wedge_i)^2 = \Sigma(y_i - a - bx_i)^2$$

where y_i denotes the actual or observed value and $y^\wedge_i = a + b_{xi}$, the estimated value of y_i for a given value of x_i, e_i is the difference between the observed value and the estimated value and ei is technically known as error or residue. This summation intends over n pairs of observations of (x_i, y_i). The line of regression of y or x and the errors of estimation are shown in the following figure.



SHOWING REGRESSION LINE OF y on x AND ERRORS OF ESTIMATION

Minimisation of the equation yields the following equations known as 'Normal Equations'

$$\Sigma y_i = na + b\Sigma x_i$$
$$\Sigma x_i y_i = a\Sigma x_i + b\Sigma x_i^2$$

Solving there, two equations for b and a, we have the "least squares" estimates of b and a as

$$b = \frac{cov(x_i y_i)}{S_x^2} = \frac{r.S_x.S_y}{S_x^2}$$
$$= r.\frac{S_y}{S_x}$$

After estimating b, estimate of a is given by

$$a = \bar{y} - b\bar{x}$$

Substituting the estimates of b and a in equation, we get

$$\frac{(y-\bar{y})}{S_y} = \frac{r(x-\bar{x})}{S_x}$$

There may be cases when the variable x depends on y and we may take the regression line of x on y as

$$x = a^\wedge + b^\wedge y$$

Unlike the minimisation of vertical distances in the scatter diagram as shown in figure for obtaining the estimates of a and b, in this case we minimise the horizontal distances and get the following normal equation in a^ and b^, the two regression parameters:

$$\Sigma x_i = na^\wedge + b^\wedge \Sigma y_i$$
$$\Sigma x_i y_i = a^\wedge \Sigma y_i + b^\wedge \Sigma y_i^2$$

or solving these equations, we get

$$b^\wedge = b_{xy} = \frac{cov(x, y)}{S_y^2} = \frac{r.S_x}{S_y}$$
$$a^\wedge = \bar{x} - b^\wedge \bar{y}$$

A single formula for estimating b is given by

$$b^\wedge = b_{yx} = \frac{n\sum x_i y_i - \sum x_i . \sum y_i}{n\sum x^2 - \left(\sum x\right)^2}.$$

$$\text{Similarly, } b^\wedge = b_{xy} = \frac{n\sum x_i y_i - \sum x_i . \sum y_i}{n\sum y_i^2 - \left(\sum y_i\right)^2}$$

The standardised form of the regression equation of x on y is given by

$$\frac{x-\bar{x}}{S_x} = r\frac{\left(y-\bar{y}\right)}{S_y}$$

PROPERTIES of Regression lines: We consider the following important properties of regression lines:

(i) **The regression coefficients remain unchanged due to a shift of origin but change due to a shift of scale.**

This property states that if the original pair of variables is (x, y) and if they are changed to the pair (u, v) where

$$u = \frac{x-a}{p} \text{ and } v = \frac{y-c}{q}$$

$$b_{xy} = \frac{p}{q} \times b_{vu} \text{ and } b_{yx} = \frac{q}{p} \times b_{uv}$$

(ii) **The two lines of regression intersect at the point, where x and y are the variables under consideration.**

According to this property, the point of intersection of the regression line of y on x and the regression line of x on y is the solution of the simultaneous equations in x and y.

(iii) **The coefficient of correlation between two variables x and y in the simple geometric mean of the two regression coefficients. The sign of the correlation coefficient would be the common sign of the two regression coefficients.**

This property says that if the two regression coefficients are denoted by b_{yx} (=b) and b_{xy} (=b') then the coefficient of correlation is given by

$$r = \pm\sqrt{b_{yx} \times b_{xy}}$$

If both the regression coefficients are negative, r would be negative and if both are positive, r would assume a positive value.

1. If for two variables x and y, the covariance, variance of x and variance of y are 40, 16 and 256 respectively, what is the value of the correlation coefficient?

Solution:

Cov(x,y)=30, V(x)=25, V(y)=144

As we know formula of Corelation coefficient is:

Let r be Corelation coefficient of x,y

$$r = \frac{\text{Cov }(x,y)}{\sqrt{V(x) \text{ x } V(y)}} = \frac{30}{\sqrt{25 \text{ x } 144}} = \frac{30}{5 \text{ x } 12} = 0.5$$

$\Rightarrow$r=0.5

2. If the covariance between two variables is 20 and the variance of one of the variables is 16, what would be the variance of the other variable?

Solution: Given, Cov (x, y) =20 and variance of one of the variables is 16.

so the standard deviation SD is 4.

we know the formula,

r= cov (x, y) /(SD of x * SD of Y)

r= 20 /4 * SD of the other variable

r= 5/ SD of the other variable

we also know that coefficient of correlation, r, lies between –1 and +1 including them.

so, SD of the other variable has to be atleast 5 or more.

so the variance will be 5^2 = 25 atleast or more. $S^2 \geq 25$

3. If r = 0.6, then the coefficient of non-determination is

Solution: Given r = 0.6

The coefficient of non-determination = 1 – r^2= 1 – 0.36 = 0.64

4. If u+5x=6 and 3y+7v=20 and correlation coefficient between x and y is 0.58, then what is correlation coefficient between u and v ?

Solution: Correlation coefficient between x and y is 0.58

u + 5x = 6

=> u = 6 – 5x

–5 is the factor (constant does not have any impact)

3y + 7v = 20

=> 7v = –3y + 20

=> v = (–3/7) y + 20/7

(–3/7) is the factor (constant does not have any impact)

correlation coefficient between u and v = 0.58 x (–5)(–3/7) / $\sqrt{(-5)^2}\sqrt{(-3/7)^2}$

= 0.58

correlation coefficient between u and v = 0.58

5. If the relation between x and u is 3x + 4u + 7 = 0 and the correlation coefficient between x and y is –0.6, then what is the correlation coefficient between u and y?

Solution: Given x and u is 3x + 4u + 7 = 0

$$\therefore u = \frac{-3x-7}{4}$$

We can write this as

$$u = \left(-\frac{3}{4}\right)x - \left(\frac{7}{4}\right)$$

Therefore, perfect negative correlation between x and y and that is –0.6

$$= \frac{-(0.6) \times \left(-\frac{3}{4}\right)}{\left(\frac{3}{4}\right)} = 0.6$$

So, the correlation between u and y is 0.6

6. If the sum of squares of difference of ranks, given by two judges A and B of 8 students is 21, what is the value of rank correlation coefficient?

Solution: Here, n = 8.

and 'Σ (d^2)' = 21.

Now, n x {(n^2) – 1} = 8 x 63 = 504. So, rank correlation coefficient

= 1 – (6 x 21 / 504)

= 1 – 0.25 = 0.75

7. If the rank correlation coefficient between marks in management and mathematics for a group of student is 0.6 and the sum of squares of the differences in ranks is 66, what is the number of students in the group?

Solution: Rank correlation coefficient = 1 – 6 Σ $(di)^2$ / ($n(n^2$ – 1))

n = number of students of group

Σ $(di)^2$ = sum of squares of the differences in ranks = 66

rank correlation coefficient = 0.6

=> 0.6 = 1 – 6 Σ $(di)^2$ / ($n(n^2$ – 1))

=> –0.4 = –6 x 66 / ($n(n^2$ – 1))

=> ($n(n^2$ – 1)) = 990

=> ($n(n^2$ – 1)) = 10 x 99

=> ($n(n^2$ – 1)) = 10 x (100 – 1)

=> ($n(n^2$ – 1)) = 10 x (10^2 – 1)

=> n = 10

Therefore, the number of students of group = 10.

8. While computing rank correlation coefficient between profit and investment for the last 6 years of a company the difference in rank for a year was taken 3 instead of 4. What is the rectified rank correlation coefficient if it is known that the original value of rank correlation coefficient was 0.4?

Solution: rank correlation coefficient $= 1 - 6\ \Sigma\ (di)^2 / (n(n^2 - 1))$

n = number of years = 6

$0.4 = 1 - 6\ \Sigma\ (di)^2 / (6(6^2 - 1)$

$\Rightarrow 0.6 = \Sigma\ (di)^2/35$

$\Rightarrow \Sigma\ (di)^2 = 21$

rank for a year was taken 3 instead of 4.

=> Actual $\Sigma\ (di)^2 = 21 - (3)^2 + 4^2 = 28$

Actual rank correlation coefficient $= 1 - 6 \times 28 / (n(n^2 - 1))$

$= 1 - 28/35 = 7/35 = 1/5 = 0.2$

9. For 10 pairs of observations, No. of concurrent deviations was found to be 4. What is the value of the coefficient of concurrent deviation?

Solution: Step-by-step explanation:

The Formula for Coefficient of Concurrent deviation is:

$$R = \pm \sqrt{\left|\frac{2c-m}{m}\right|}$$

here, m = n – 1 (and n is the total number of observation) = 10 – 1 = 9

c = Number of pair of concurrent deviation = 4

Substituting all values in formula,

$$R = \pm \sqrt{\left|\frac{2x4-9}{9}\right|}$$

Also, $2\times4-9 < 0$ so take negative value

$\Rightarrow R = \frac{-1}{3}$

10. For 10 pairs of observations, No. of concurrent deviations was found to be 4. What is the value of the coefficient of concurrent deviation?

Solution: Given coefficient of concurrent deviation is given by $R = \pm\sqrt{2}\ c - m\ /\ m$

Now m = p – 1 and c = 6 and $R = 1/\sqrt{3}$

So $1/\sqrt{3} = \sqrt{2} \times 6 - (p - 1) / p - 1$

$1/\sqrt{3} = \sqrt{12} - p + 1 / p - 1$

Squaring both sides, we get

$1/3 = 13 - p / p - 1$

Or $p - 1 = 39 - 3p$

$4p = 40$ Or $p = 40 / 4$ Or p = 10 pairs

11. Following are the two normal equations obtained for deriving the regression line of y and x:

5a + 10b = 40, 10a + 25b = 95. The regression line of y on x is given by

Solution:

The normal equations obtained for deriving the regression line of y and x: 5a + 10b = 40 and 10a + 25b = 95.

In order to find the regression line of y on x, we need to solve the 5a + 10b = 40 and 10a + 25b = 95 to find out the values of a and b. So, we have,

Given,

5a + 10b = 40 ....(1)

10a + 25b = 95 ....(2)

2 × (1) gives,

10a + 20b = 80 ....(3)

(2) – (3) gives,

10a + 25b = 95

10a + 20b = 80

5b = 15, **b = 3** from (1)

5a + 10b = 40

5a + 10 (3) = 40

5a + 30 = 40

5a = 40 – 30

5a = 10

a = 2

As, "a" represents the y-intercept and "b" represents the slope, so we have,

The regression line of y on x is given by **y = 2x + 3**

12. Given the regression equations as 2x+3y = 6 and 5x+7y =12, then which one is regression equation x on y?

Solution: For regression equations, both coefficients both b_{xy} and b_{yx} need to be of the same sign.

If 2x+3y = 6 is y on x and 5x+7y–12 = 0 is x on y

then y=6/3–2/3x then $x = \frac{12-7y}{5}$

$b_{yx}1 = -2/3$ and $b_{xy}1 = -7/5$

If 2x+3y = 6 is x on y and 5x+7y = 12 is y on x

Then $x = \frac{6-3y}{2}$ and $y = \frac{12-5x}{7}$

$b_{xy}2 = -3/2$ and $b_{yx}2 = -5/7$

$b_{xy}1 \times b_{yx}1 = 14/15 < 1$

$b_{yx}2 \times b_{xy}2 = 15/14 > 1$

We know $b_{xy} \times b_{yx} = r^2$ which is always <1

So first assumption is correct.

Regression equation of y on x is 2x+3y = 6

Regression equation of x on y is 5x+7y–12 = 0

13. If y = 3x + 4 is the regression line of y on x and the arithmetic mean of x is –1, what is the arithmetic mean of y?

Solution: Given regression line y on x is y = 3x + 4

Given $\bar{y} = 3\bar{x} + 4 = 3(-1) + 4 = 1$

Therefore, arithmetic mean of y =1

14. Compute the correlation coefficient between x and y from the following data n = 10, $\Sigma xy = 220$, $\Sigma x^2 = 200$, $\Sigma y^2 = 262$ $\Sigma x = 40$ and $\Sigma y = 50$

Solution: From the given data, we have by applying

$$r = \frac{n\sum xy - \sum x \times \sum y}{\sqrt{n\sum x^2 - \left(\sum x\right)^2}\sqrt{n\sum y^2 - \left(\sum y\right)^2}}$$

$$= \frac{10 \times 220 - 40 \times 50}{\sqrt{10 \times 200 - (40)^2} \times \sqrt{10 \times 262 - (50)^2}}$$

$$= \frac{2200 - 2000}{\sqrt{2000 - 1600} \times \sqrt{2620 - 2500}}$$

$$= \frac{200}{20 \times 10.9545} = 0.91$$

Thus, there is a good amount of positive correlation between the two variables x and y.

Alternately

As given, $\bar{x} = \frac{\sum x}{n} = \frac{40}{10} = 4$, $\bar{y} = \frac{\sum y}{n} = \frac{50}{10} = 5$

$$\text{Cov}(x, y) = \frac{\sum xy}{n} - \bar{x}.\bar{y} = \frac{220}{10} - 4 \times 5 = 22 - 20 = 2$$

$$S_x = \sqrt{\frac{\sum x^2}{n} - (\bar{x})^2} = \sqrt{\frac{200}{10} - 4^2} = 2$$

$$S_y = \sqrt{\frac{\sum y_i^2}{n} - \bar{y}^2} = \sqrt{\frac{262}{10} - 5^2} = \sqrt{26.20 - 25} = 1.0954$$

Thus, applying formula, we get

$$r = \frac{\text{cov}(x,y)}{S_x.S_y} = \frac{2}{2 \times 1.0954} = 0.91$$

As before, we draw the same conclusion.

15. For a group of 10 students, the sum of squares of differences in ranks for Mathematics and Statistics marks was found to be 50. What is the value of rank correlation coefficient?

Solution: As given n = 10 and $\sum d_i^2 = 50$. Hence the rank correlation coefficient between marks in Mathematics and Statistics is given by

$$r_R = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} = 1 - \frac{6 \times 50}{10(10^2 - 1)} = 1 - 0.30 = 0.70$$

16. For a number of towns, the coefficient of rank correlation between the people living below the poverty line and increase of population is 0.50. If the sum of squares of the differences in ranks awarded to these factors is 82.50, find the number of towns.

Solution: As given $r_R = 0.50$, $\sum d_i^2 = 82.50$.

$$\text{Thus } r_R = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

$$0.50 = 1 - \frac{6 \times 82.50}{n(n^2 - 1)}$$

$= n(n^2 - 1) = 990$

$= n(n^2 - 1) = 10(10^2 - 1)$

Therefore n = 10 as n must be a positive integer.

17. While computing rank correlation coefficient between profits and investment for 10 years of a firm, the difference in rank for a year was taken as 7 instead of 5 by mistake and the value of rank correlation coefficient was computed as 0.80. What would be the correct value of rank correlation coefficient after rectifying the mistake?

Solution: We are given that n = 10,

r_R = 0.80 and the wrong $d_i = 7$ should be replaced by 5.

$$r_R = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

$$0.80 = 1 - \frac{6\sum d_i^2}{10(10^2 - 1)}$$

$$\sum d_i^2 = 33$$

Corrected $\sum d_i^2 = 33 - 7^2 + 5^2 = 9$

Hence rectified value of rank correlation coefficient $= 1 - \frac{6 \times 9}{10(10^2 - 1)}$ = 0.95

18. Find product moment correlation coefficient from the following information:

x :	2	3	5	5	6	8
y :	9	8	8	6	5	3

Solution: In order to find the covariance and the two-standard deviation, we prepare the following table:

Table: Computation of Correlation Coefficient

x_i	y_i	x_iy_i	x_i^2	y_i^2
(1)	(2)	(3)= (1) x (2)	(4)= $(1)^2$	(5)= $(2)^2$
2	9	18	4	81
3	8	24	9	64
5	8	40	25	64
5	6	30	25	36
6	5	30	36	25
8	3	24	64	9
29	39	166	163	279

We have

$$\bar{x} = \frac{29}{6} = 4.8333 \quad \bar{y} = \frac{39}{6} = 6.50$$

$$\text{cov}(x, y) = \frac{\sum x_iy_i}{n} - \bar{x}\,\bar{y}$$

= 166/6 – 4.8333 × 6.50 = –3.7498

$$= \sqrt{\frac{\sum x_i^2}{n} - (\bar{x})^2}$$

$$= \sqrt{\frac{163}{6}\ (4.8333)^2}$$

$$= \sqrt{27.1667 - 23.3608} = 1.95$$

$$S_y = \sqrt{\frac{\sum y_i^2}{n} - (\bar{y})^2}$$

$$= \sqrt{\frac{279}{6} - (6.50)^2}$$

$$= \sqrt{46.50 - 42.25} = 2.0616$$

Thus the correlation coefficient between x and y is given by

$$r = \frac{\text{cov}(x,y)}{S_x.S_y} = \frac{-3.7498}{1.9509 \times 2.0616} = -0.93$$

We find a high degree of negative correlation between x and y.

19. If y is independent variable and x is independent variable and SD of x and y are 5 and 8 respectively and coefficient of correlation between x and y is 0.8. Find the regression coefficient y on x

Solution: SD of x (σ_x) = 5, SD of y (σ_y) = 8,

Coefficient of correlation (r) = 0.8

Regression coefficient y on x = $b_{yx} = r \times \frac{\sigma_y}{\sigma_x} = \frac{0.8 \times 8}{5} = \frac{6.4}{5} = 1.28$

CA FOUNDATION - PAPER 3 - BUSINESS MATHEMATICS, LOGICAL REASONING AND STATISTICS

This capsule is in continuation to the previous edition featured in November 2022. Here, an attempt is made to enable the students to understand the Correlation and Regression with the help of examples.

CHAPTER 17: CORRELATION AND REGRESSION - PART 3

1. Given that the correlation coefficient between x and y is 0.8, write down the correlation coefficient between u and v where

(i) $2u + 3x + 4 = 0$ and $4v + 16x + 11 = 0$
(ii) $2u - 3x + 4 = 0$ and $4v + 16x + 11 = 0$
(iii) $2u - 3x + 4 = 0$ and $4v - 16x + 11 = 0$
(iv) $2u + 3x + 4 = 0$ and $4v - 16x + 11 = 0$

Solution: Using formula, we find that

$$r_{xy} = \frac{bd}{|b||d|} r_{uv}$$

i.e., $r_{xy} = r_{uv}$ if b and d are of same sign and $r_{uv} = -r_{xy}$ when b and d are of opposite signs, b and d being the scales of x and y respectively. In (i), $u = (-2) + (-3/2)\,x$ and $v = (-11/4) + (-4)y$.

Since $b = -3/2$ and $d = -4$ are of same sign, the correlation coefficient between u and v would be the same as that between x and y, i.e., $r_{xy} = 0.8 = r_{uv}$

In (ii), $u = (-2) + (3/2)x$ and $v = (-11/4) + (-4)y$. Hence $b = 3/2$ and $d = -4$ are of opposite signs and we have $r_{uv} = -r_{xy} = -0.8$

Proceeding in a similar manner, we have $r_{uv} = 0.8$ and -0.8 in (iii) and (iv).

2. For the variables x and y, the regression equations are given as $7x - 3y - 18 = 0$ and $4x - y - 11 = 0$

(i) Find the arithmetic means of x and y.
(ii) Identify the regression equation of y on x.
(iii) Compute the correlation coefficient between x and y.
(iv) Given the variance of x is 9, find the SD of y.

Solution:

(i) Since the two lines of regression intersect at the point, $(\bar{x}, \bar{y})$ replacing x and y by $\bar{x}$ and $\bar{y}$ respectively in the given regression equations, we get

$$7\bar{x} - 3\bar{y} - 18 = 0$$

and $4\bar{x} - \bar{y} - 11 = 0$

Solving these two equations, we get $\bar{x} = 3$ and $\bar{y} = 1$

Thus, the arithmetic means of x and y are given by 3 and 1 respectively.

(ii) Let us assume that $7x - 3y - 18 = 0$ represents the regression line of y on x and $4x - y - 11 = 0$ represents the regression line of x on y.

Now $7x - 3y - 18 = 0$

$$\Rightarrow \quad y = (-6) + \frac{(7)}{3}x$$

$$\Rightarrow \quad b_{yx} = \frac{7}{3}$$

Again $4x - y - 11 = 0$

$$\Rightarrow \quad x = \frac{(11)}{4} + \frac{(11)}{4}y \qquad \therefore b_{xy} = \frac{1}{4}$$

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

$$= \frac{7}{3} \times \frac{1}{4}$$

$$= \frac{7}{12} < 1$$

Since $|r| \le 1 \Rightarrow r^2 \le 1$, our assumptions are correct. Thus, $7x - 3y - 18 = 0$ truly represents the regression line of y on x.

(iii) Since $r^2 = \frac{7}{12}$

$$\therefore r = \sqrt{\frac{7}{12}}$$ (We take the sign of r as positive since both the regression coefficients are positive)

$= 0.7638$

(iv) $b_{yx} = r \times \frac{S_y}{S_x}$

$$\Rightarrow \frac{7}{3} = 0.7638 \times \frac{S_y}{3} \quad (\because S_x^2 = 9 \text{ as given})$$

$$\Rightarrow S_y = \frac{7}{0.7638} = 9.1647$$

3. The following data relate to the test scores obtained by eight salesmen in an aptitude test and their daily sales in thousands of rupees:

Salesman :	1	2	3	4	5	6	7	8
Scores :	60	55	62	56	62	64	70	54
Sales :	31	28	26	24	30	35	28	24

Solution:

Let the scores and sales be denoted by x and y respectively. We take a, origin of x as the average of the two extreme values, i.e., 54 and 70. Hence, a = 62 similarly, the origin of y is taken as $b = \frac{24 + 35}{2} \cong 30$

Table
Computation of Correlation Coefficient Between Test Scores and Sales

Scores	Sales in	u_i	v_i	$u_i v_i$	u_i^2	v_i^2
	Rs. 1000	$= x_i - 62$	$= y_i - 30$			
(x_i)	(y_i)					
(1)	(2)	(3)	(4)	(5)=(3) x(4)	$(6)=(3)^2$	$(7)=(4)^2$
60	31	–2	1	–2	4	1
55	28	–7	–2	14	49	4
62	26	0	–4	0	0	16
56	24	–6	–6	36	36	36
62	30	0	0	0	0	0
64	35	2	5	10	4	25
70	28	8	–2	–16	64	4
54	24	–8	–6	48	64	36
Total	—	–13	–14	90	221	122

Since correlation coefficient remains unchanged due to change of origin, we have

$$r = r_{xy} = r_{uv} = \frac{n\sum u_i v_i - \sum u_i \times \sum v_i}{\sqrt{n\sum u_i^2 - \left(\sum u_i\right)^2} \times \sqrt{n\sum v_i^2 - \left(\sum v_i\right)^2}}$$

$$= \frac{8 \times 90 - (-13) \times (-14)}{\sqrt{8 \times 221 - (-13)^2} \times \sqrt{8 \times 122 - (-14)^2}} = \frac{538}{\sqrt{1768 - 169} \times \sqrt{976 - 196}}$$

$= 0.48$

In some cases, there may be some confusion about selecting the pair of variables for which correlation is wanted. This is explained in the following problem.

4. Examine whether there is any correlation between age and blindness on the basis of the following data:

Age in years :	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of Persons (in thousands):	90	120	140	100	80	60	40	20
No. of blind Persons:	10	15	18	20	15	12	10	06

Solution:

Let us denote the mid-value of age in years as x and the number of blind persons per lakh as y. Then as before, we compute correlation coefficient between x and y.

Table : Computation of correlation between age and blindness

Age in	Mid-value	No. of	No. of	No. of	xy	x^2	y^2
years	x	Persons	blind	blind per	(2)×(5)	$(2)^2$	$(5)^2$
(1)	(2)	('000)	B	lakh	(6)	(7)	(8)
		P	(4)	y=B/P × 1 lakh			
		(3)		(5)			
0-10	5	90	10	11	55	25	121
10-20	15	120	15	12	180	225	144
20-30	25	140	18	13	325	625	169
30-40	35	100	20	20	700	1225	400
40-50	45	80	15	19	855	2025	361
50-60	55	60	12	20	1100	3025	400
60-70	65	40	10	25	1625	4225	625
70-80	75	20	6	30	2250	5625	900
Total	320	—	—	150	7090	17000	3120

The correlation coefficient between age and blindness is given by

$$r = \frac{n\sum xy - \sum x.\sum y}{\sqrt{n\sum x^2 - \left(\sum x\right)^2} \times \sqrt{n\sum y^2 - \left(\sum y\right)^2}}$$

$$= \frac{8\times 7090 - 320\times 150}{\sqrt{8\times 17000 - (320)^2} \times \sqrt{8\times 3120 - (150)^2}}$$

$$= \frac{8720}{183.3030 \times 49.5984} = 0.96$$

which exhibits a very high degree of positive correlation between age and blindness.

5. Coefficient of correlation between x and y for 20 items is 0.4. The AM's and SD's of x and y are known to be 12 and 15 and 3 and 4 respectively. Later on, it was found that the pair (20, 15) was wrongly taken as (15, 20). Find the correct value of the correlation coefficient.

Solution:

We are given that n = 20 and the original r = 0.4, $\bar{x} = 12$, $\bar{y} = 15$, $S_x = 3$ and $S_y = 4$

$$r = \frac{cov(x,y)}{S_x.S_y} = 0.4 = \frac{cov(x,y)}{3\times 4}$$

= Cov (x, y) = 4.8

$$= \frac{\sum xy}{n} - \bar{x}\bar{y} = 4.8$$

$$= \frac{\sum xy}{20} - 12\times 15 = 4.8$$

= $\sum xy = 3696$

Hence, corrected $\sum xy$= 3696 – 20 × 15 + 15 × 20 = 3696

Also, $S_x^2 = 9$

= $(\sum x^2/20) - 12^2 = 9$

$\sum x^2 = 3060$

Similarly, $S_y^2 = 16$

$$S_y^2 = \frac{\sum y^2}{20} - 15^2 = 16$$

$\sum y^2 = 4820$

Thus, corrected $\sum x = n\bar{x}$ – wrong value + correct value.

= 20 × 12 – 15 + 20

= 245

Similarly, corrected $\sum y$ = 20 × 15 – 20 + 15 = 295

Corrected $\sum x^2 = 3060 - 15^2 + 20^2 = 3235$

Corrected $\sum y^2 = 4820 - 20^2 + 15^2 = 4645$

Thus, corrected value of the correlation coefficient by applying formula

$$= \frac{20\times 3696 - 245\times 295}{\sqrt{20\times 3235 - (245)^2} \times \sqrt{20\times 4645 - (295)^2}}$$

$$= \frac{73920 - 72275}{68.3740 \times 76.6480}$$

= 0.31

6. If the regression line of y on x is given by y = x + 2 and Karl Pearson's coefficient of correlation is 0.5,

then $\frac{\sigma_y^2}{\sigma_x^2} =$

Solution: The regression line of y on x is given by y = x + 2

x – y + 2 = 0

$$b_{yx} = -\frac{\text{coefficent of x}}{\text{coefficent of y}} = \frac{-1}{-1} = 1$$

$b_{yx} = 1$

$$b_{yx} = r\frac{\sigma_y}{\sigma_x} [Coefficent\ of\ correlation\ (r) = 0.5]$$

$$1 = 0.5\frac{\sigma_y}{\sigma_x}$$

$$\frac{\sigma_y}{\sigma_x} = \frac{1}{0.5} = 2$$

$$\left(\frac{\sigma_y}{\sigma_x}\right)^2 = 2^2 \rightarrow \frac{\sigma_y^2}{\sigma_x^2} = 4$$

7. Two variables x and y are related according to 4x + 3y = 7, then x and y are:

Solution: Given regression equation

4x + 3y = 7 and 4x + 3y = 7

3y = 7 – 4x and 4x = 7 – 3y

$$y = \frac{7}{3} - \frac{4x}{3} \text{ and } x = \frac{7}{4} - \frac{3}{4}y$$

y = a + bx and x = a + by

We get

$b_{yx} = -4/3$ and $b_{xy} = -3/4$

$$r = \pm\sqrt{b_{yx}\times b_{xy}} = \pm\sqrt{\left(\frac{-4}{3}\right)\left(-\frac{3}{4}\right)} = -\sqrt{1}$$ (both b_{xy},b_{yx} &r are same sign)

r = –1(Perfectly Negative correlation)

CA FOUNDATION - PAPER 3 - BUSINESS MATHEMATICS AND LOGICAL REASONING & STATISTICS

Often, we encounter news of price rise, GDP growth, production growth, etc. It is important for students of Chartered Accountancy to learn techniques of measuring growth/rise or decline of various economic and business data and how to report them objectively. After reading this capsule, students will be able to understand Purpose of constructing index number and its important applications in understanding rise or decline of production, prices, etc. different methods of computing index number.

CHAPTER 18: INDEX NUMBERS

INTRODUCTION

Index numbers are convenient devices for measuring relative changes of differences from time to time or from place to place. Just as the arithmetic mean is used to represent a set of values, an index number is used to represent a set of values over two or more different periods or localities.

The basic device used in all methods of index number construction is to average the relative change in either quantities or prices since relatives are comparable and can be added even though the data from which they were derived cannot themselves be added. For example, if wheat production has gone up to 110% of the previous year's production and cotton production has gone up to 105%, it is possible to average the two percentages as they have gone up by 107.5%. This assumes that both have equal weight; but if wheat production is twice as important as cotton, percentage should be weighted 2 and 1. The average relatives obtained through this process are called the index numbers.

- **Relatives:** One of the simplest examples of an index number is a price relative, which is the ratio of the price of single commodity in a given period to its price in another period called the base period or the reference period. It can be indicated as follows:

Price relative = $\frac{P_n}{P_o} \times 100$

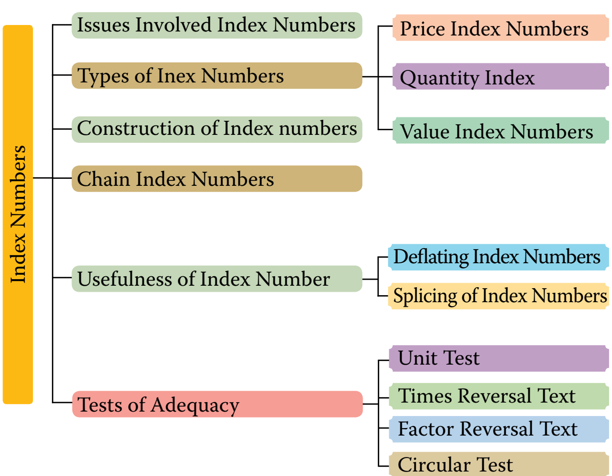
There can be other relatives such as of quantities, volume of consumption, exports, etc. The relatives in that case will be:

Quantity relative = $\frac{Q_n}{Q_o} \times 100$

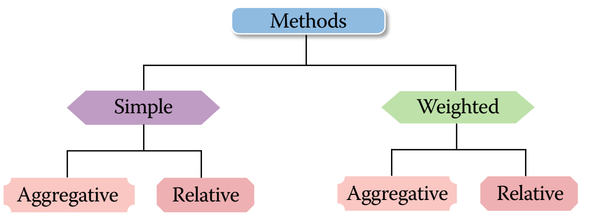
Similarly, there are value relatives:

Value relative = $\frac{V_n}{V_o} = \frac{P_nQ_n}{P_oQ_o} = \left(\frac{P_n}{P_o} \times \frac{Q_n}{Q_o}\right) \times 100$

Index number Overview



Methods of Index numbers:



Price Index numbers

(a) Simple aggregative price index = $\frac{\sum P_n}{\sum P_0} \times 100$

(b) Laspeyres' Index: In this Index base year quantities are used as weights:

Laspeyres Index = $\frac{\sum P_nQ_o}{\sum P_oQ_o} \times 100$

(c) Paasche's Index: In this Index current year quantities are used as weights:

Passche's Index = $\frac{\sum P_nQ_n}{\sum P_oQ_n} \times 100$

(d) The Marshall-Edgeworth index uses this method by taking the average of the base year and the current year

Marshall-Edgeworth Index = $\frac{\sum P_n(Q_o + Q_n)}{\sum P_o(Q_o + Q_n)} \times 100$

(e) Fisher's ideal Price Index: This index is the geometric mean of Laspeyres' and Paasche's.

Fisher's Index = $\sqrt{\frac{\sum P_nQ_o}{\sum P_oQ_o} \times \frac{\sum P_nQ_n}{\sum P_oQ_n}} \times 100$

(g) Weighted Average of Relative Method: $\frac{\sum\left(\frac{P_n}{P_o}\right) \times (P_oQ_o)}{\sum P_oQ_o} \times 100 = \frac{\sum P_nQ_o}{\sum P_oQ_o} \times 100$

(h) Chain Index = $\frac{\text{Link relative of current year X Chain Index of the previous year}}{100}$

Quantity Index Numbers

- Simple aggregate of quantities: $\frac{\sum Q_n}{\sum Q_0} \times 100$
- The simple average of quantity relatives: $\frac{\frac{\sum Q_n}{\sum Q_0}}{N} \times 100$
- Weighted aggregate quantity indices:
  - (i) With base year price as weight : $\frac{\sum Q_nP_o}{\sum Q_oP_o} \times 100$ (Laspeyre's index)
  - (ii) With current year price as weight : $\frac{\sum Q_nP_n}{\sum Q_oP_n} \times 100$ (Paasche's index)
  - (iii) Geometric mean of (i) and (ii) : $\sqrt{\frac{\sum Q_nP_o}{\sum Q_oP_o} \times \frac{\sum Q_nP_n}{\sum Q_oP_n}} \times 100$ (Fisher's Ideal)
- Base-year weighted average of prices as relatives in Marshall-Edgeworth quantity index number.

  Weighted Relative method formula $\frac{\sum\left(\frac{Q_n}{Q_o}P_oQ_o\right)}{\sum P_oQ_o} \times 100$
- Value Indices $\frac{V_n}{V_o} = \frac{\sum P_nQ_n}{\sum P_oQ_o} \times 100$

THE CHAIN INDEX NUMBERS

So far we concentrated on a fixed base but it does not suit when conditions change quite fast. In such a case the changing base for example, 2018 for 2019, and 2019 for 2020, and so on, may be more suitable. If, however, it is desired to associate these relatives to a common base the results may be chained. Thus, under this method the relatives of each year are first related to the preceding year called the link relatives and then they are chained together by successive multiplication to form a chain index.

The formula is:

$$\text{Chain Index} = \frac{\text{Link relative of current year X Chain Index of the previous year}}{100}$$

Illustrations

Commodities	2018	2019	2020
Cheese (per 100 gms)			
Egg (per piece)	120	150	160
Potato (per kg)	30	36	40
	50	60	60
Aggregrate	200	246	260
Index	100	123	130

Simple Aggregative Index for 2019 over 2018 $= \frac{\sum P_n}{\sum P_o} = \frac{246}{200} \times 100 = 123$

and for 2020 over 2018 $= \frac{\sum P_n}{\sum P_o} \times 100 = \frac{260}{200} \times 100 = 130$

The above method is easy to understand but it has a serious defect. It shows that the first commodity exerts greater influence than the other two because the price of the first commodity is higher than that of the other two. Further, if units are changed then the Index numbers will also change. Students should independently calculate the Index number taking the price of eggs per dozen i.e., ` 36, ` 43.20, ` 39.60 for the three years respectively. This is the major flaw in using absolute quantities and not the relatives. Such price quotations become the concealed weights which have no logical significance.

Limitations of Index Numbers

So far we have studied various types of index numbers. However, they have certain limitations. They are:

1. As the indices are constructed mostly from deliberate samples, chances of errors creeping in cannot be always avoided.
2. Since index numbers are based on some selected items, they simply depict the broad trend and not the real picture.
3. Since many methods are employed for constructing index numbers, the result gives different values and this at times create confusion.

Usefulness of Index Numbers

In spite of its limitations, index numbers are useful in the following areas:

1. Framing suitable policies in economics and business. They provide guidelines to make decisions in measuring intelligence quotients, research etc.
2. They reveal trends and tendencies in making important conclusions in cyclical forces, irregular forces, etc.
3. They are important in forecasting future economic activity. They are used in time series analysis to study long-term trend, seasonal variations and cyclical developments.
4. Index numbers are very useful in deflating, i.e., they are used to adjust the original data for price changes and thus transform nominal wages into real wages.
5. Cost of living index numbers measure changes in the cost of living over a given period.

DEFLATING TIME SERIES USING INDEX NUMBERS

Sometimes a price index is used to measure the real values in economic time series data expressed in monetary units. For example, GNP initially is calculated in current price so that the effect of price changes over a period of time gets reflected in the data collected. Thereafter, to determine how much the physical goods and services have grown over time, the effect of changes in price over different values of GNP is excluded. The real economic growth in terms of constant prices of the base year therefore is determined by deflating GNP values using price index.

Year	*Wholesale Price Index*	*GNP at Current Prices*	*Real GNP*
2019	113.1	7499	6630
2020	116.3	7935	6823
2021	121.2	8657	7143
2022	127.7	9323	7301

The formula for conversion can be stated as

$$\text{Deflated Value} = \frac{\text{Current Value}}{\text{Price Index of the current year}} \times 100$$

$$\text{Or Current Value} \times \frac{\text{Base Price } (P_0)}{\text{Current Price } (P_n)}$$

Shifting and Splicing of Index Numbers

These refer to two technical points: (i) how the base period of the index may be shifted, (ii) how two index covering different bases may be combined into single series by splicing.

Shifted Price Index

Year	Original Price Index	Shifted Price Index to base 2021
2011	100	71.4
2012	104	74.3
2013	106	75.7
2014	107	76.4
2015	110	78.6
2016	112	80.0
2017	115	82.1
2018	117	83.6
2019	125	89.3
2020	131	93.6
2021	140	100.0
2022	147	105.0

The formula used is, $\text{Shifted Price Index} = \frac{\text{Original Price Index}}{\text{Price Index of the year on which it has to be shifted}} \times 100$

Splicing two sets of price index numbers covering different periods of time is usually required when there is a major change in quantity weights. It may also be necessary on account of a new method of calculation or the inclusion of new commodity in the index.

Splicing Two Index Number Series

Year	Old Price Index [2010 = 100]	Revised Price Index [2015=100]	Spliced Price Index [2015 = 100]
2010	100.0		87.6
2011	102.3		89.6
2012	105.3		92.2
2013	107.6		94.2
2014	111.9		98.0
2015	114.2	100.0	100.0
2016		102.5	102.5
2017		106.4	106.4
2018		108.3	108.3
2019		111.7	111.7
2020		117.8	117.8

You will notice that the old series 2010 has to be converted shifting to the base. 2015 i.e, 114.2 to have a continuous series, even when the two parts have different weights

TEST OF ADEQUACY

There are four tests:

1. Unit Test: This test requires that the formula should be independent of the unit in which or for which prices and quantities are quoted. Except for the simple (unweighted) aggregative index all other formulae satisfy this test.

(ii) Time Reversal Test: It is a test to determine whether a given method will work both ways in time, forward and backward. The test provides that the formula for calculating the index number should be such that two ratios, the current on the base and the base on the current should multiply into unity. In other words, the two indices should be reciprocals of each other. Symbolically,

$$P_{01} \times P_{10} = 1$$

where P_{01} is the index for time 1 on 0 and P_{10} is the index for time 0 on 1.

You will notice that Laspeyres' method and Paasche's method do not satisfy this test, but Fisher's Ideal Formula does.

While selecting an appropriate index formula, the Time Reversal Test and the Factor Reversal test are considered necessary in testing the consistency.

Laspeyres: $P_{01} = \frac{\sum P_1Q_0}{\sum P_0Q_0}$ $\quad P_{10} = \frac{\sum P_0Q_1}{\sum P_1Q_1}$

$$P_{01} \times P_{10} = \frac{\sum P_1Q_0}{\sum P_0Q_0} \times \frac{\sum P_0Q_1}{\sum P_1Q_1} \neq 1$$

Paasche's: $P_{01} = \frac{\sum P_1Q_1}{\sum P_0Q_1}$ $\quad P_{10} = \frac{\sum P_0Q_0}{\sum P_1Q_0}$

$$P_{01} \times P_{10} = \frac{\sum P_1Q_1}{\sum P_0Q_1} \times \frac{\sum P_0Q_0}{\sum P_1Q_0} \neq 1$$

Fisher's: $P_{01} = \sqrt{\frac{\sum P_1Q_0}{\sum P_0Q_0} \times \frac{\sum P_1Q_1}{\sum P_0Q_1}}$ $\quad P_{10} = \sqrt{\frac{\sum P_0Q_1}{\sum P_1Q_1} \times \frac{\sum P_0Q_0}{\sum P_1Q_0}}$

$$P_{01} \times P_{10} = \sqrt{\frac{\sum P_1Q_0}{\sum P_0Q_0} \times \frac{\sum P_1Q_1}{\sum P_0Q_1} \times \frac{\sum P_0Q_1}{\sum P_1Q_1} \times \frac{\sum P_0Q_0}{\sum P_1Q_0}} = 1$$

(iii) Factor Reversal Test: This holds when the product of price index and the quantity index should be equal to the corresponding value index, i.e.,

Symbolically: $P_{01} \times Q_{01} = V_{01}$

Fishers' $P_{01} = \sqrt{\frac{\sum P_1Q_0}{\sum P_0Q_0} \times \frac{\sum P_1Q_1}{\sum P_0Q_1}}$ $\quad Q_{01} = \sqrt{\frac{\sum Q_1P_0}{\sum Q_0P_0} \times \frac{\sum Q_1P_1}{\sum Q_0P_1}}$

$$P_{01} \times Q_{01} = \sqrt{\frac{\sum P_1Q_0}{\sum P_0Q_0} \times \frac{\sum P_1Q_1}{\sum P_0Q_1} \times \frac{\sum Q_1P_0}{\sum Q_0P_0} \times \frac{\sum Q_1P_1}{\sum Q_0P_1}} = \sqrt{\frac{\sum P_1Q_1}{\sum P_0Q_0} \times \frac{\sum P_1Q_1}{\sum P_0Q_0}} = \frac{\sum P_1Q_1}{\sum P_0Q_0} = V_{01}$$

Thus Fisher's Index satisfies Factor Reversal test. Because Fisher's Index number satisfies both the tests in (ii) and (iii), it is called an Ideal Index Number.

(iv) Circular Test: It is concerned with the measurement of price changes over a period of years, when it is desirable to shift the base. For example, if the 1970 index with base 1965 is 200 and 1965 index with base 1960 is 150, the index 1970 on base 1960 will be 300. This property therefore enables us to adjust the index values from period to period without referring each time to the original base. The test of this shiftability of base is called the circular test.

- This test is not met by Laspeyres, or Paasche's or the Fisher's ideal index. **The simple geometric mean of price relatives and the weighted aggregative with fixed weights meet this test.**

Example 1: Compute Fisher's Ideal Index from the following data:

Commodities	Base Year		Current Year	
	Price	*Quantity*	*Price*	*Quantity*
A	4	3	6	2
B	5	4	6	4
C	7	2	9	2
D	2	3	1	5

Show how it satisfies the time and factor reversal tests.

Solution:

Commodities	P_0	Q_0	P_1	Q_1	P_0Q_0	P_1Q_0	P_0Q_1	P_1Q_1
A	4	3	6	2	12	18	8	12
B	5	4	6	4	20	24	20	24
C	7	2	9	2	14	18	14	18
D	2	3	1	5	6	3	10	5
					52	63	52	59

Fisher's Ideal Index: $P_{01} = \sqrt{\frac{\sum P_1Q_0}{\sum P_0Q_0} \times \frac{\sum P_1Q_1}{\sum P_0Q_1}} \times 100 = \sqrt{\frac{63}{52} \times \frac{59}{52}} \times 100$

$$= \sqrt{1.375} \times 100 = 1.172 \times 100 = 117$$

Time Reversal Test:

$$P_{01} \times P_{10} = \sqrt{\frac{63}{52} \times \frac{59}{52} \times \frac{52}{59} \times \frac{52}{63}} = \sqrt{1} = 1$$

∴ Time Reversal Test is satisfied.

Factor Reversal Test:

$$P_{01} \times Q_{01} = \sqrt{\frac{63}{52} \times \frac{59}{52} \times \frac{52}{59} \times \frac{52}{63}} = \sqrt{\frac{59}{52} \times \frac{59}{52}} = \frac{59}{52}$$

Since, $\frac{\sum P_1Q_1}{\sum P_0Q_0}$ is also equal to $\frac{59}{52}$, the Factor Reversal Test is satisfied.

Example 2: If the ratio between Laspeyre's index number and Paasche's Index number is 28 : 27. Then the missing figure in the following table P is :

Commodity	Base Year		Current Year	
	Price	Quantity	Price	Quantity
X	L	10	2	5
Y	L	5	P	2

Solution: So Laspeyre's index number = Σ PxQo / Σ PoQo
= 2 x 10 + P x 5 / L x 10 + L x 5
= 20 + 5P / 15 L
= 5 (4 + P) / 15 L
= 4 + P / 3L

Now for the Paasche's index number we have, Σ Px Qx / Σ Po Qx
= 2 x 5 + P x 2 / L x 5 + L x 2
= 2P + 10 / 7L
= 2 (P + 5) / 7L

Given Ratio = L : P = 28:27

So 4 + p / 3L / 2 (P + 5) / 7L = 28/27
or 7 (4 + P) / 6(P + 5) = 28/27
or 9(4 + P) = 8(P + 5)
36 + 9P = 8P + 40
or P = 40 – 36
or P = 4

Example 3: The consumer price index for 2006 on the basis for 2006 on the basis of 2005 from the following data is:

Commodities in 2005	Quantities consumed in 2005	Prices in 2005	Prices in 2006
A	6	5.75	6.00
B	6	5.00	8.00
C	1	6.00	9.00
D	6	8.00	10.00
E	4	2.00	1.50
F	1	20.00	15.00

Solution:

Commodities in 2005	Quantities consumed in 2005(q_0)	Price in 2005 (P_0)	Prices in 2006 (P_1)	P_1q_0	P_0q_0
A	6	5.75	6.00	36	34.50
B	6	5.00	8.00	48	30.00
C	1	6.00	9.00	9	6.00
D	6	8.00	10.00	60	48.00
E	4	2.00	1.50	6	8.00
F	1	20.00	15.00	15	20.00
				ΣP_1q_0=174	ΣP_0q_0=146.5

Consumer Price Index = $\frac{\sum P_1q_0}{\sum P_0q_0}\times 100=\frac{174}{146.5}\times 100=118.77$

Example 4: Net monthly salary of an employee was ₹30,000 in 2000. The consumer price index 2015 is 250 with 2000 as base year, if he has to rightly compensated, then dearness allowance to be paid to the employee is:

Solution: The consumer price index number in 2015 is 250 with 2000 as base year.

if in 2000 = 100 then in 2015 = 250

if in 2000 = 1 then in 2015 = 250/100 = 2.5

if in 2000 = 30000 then 250/100* 30000 = 75,000

additional dearness allowance to be paid to the employee is = 75000 – 30000 = ₹45000

additional dearness allowance to be paid to the employee = ₹45,000

Example 5: Consumer price index number goes up from 110 to 200 and the Salary of a worker is also raised from ₹330 to ₹500. Therefore, in real terms, to maintain his previous standard of living, he should get an additional amount of:

Solution: Cost of Living Index in base year = 110,

Cost of Index in current year = 200

Salary of worker in base year = 330

Salary of worker in current year = 500

$$Real\ wages\ (for\ base\ year)=\frac{\text{Money wages in base year}}{\text{Cost of living index in base year}}\times 100$$

$$=\frac{330}{110}\times 100$$
$$=300$$

$$Real\ wages\ (for\ current\ year)=\frac{\text{Money wages in current year}}{\text{Cost of living index number in current year}}\times 100$$

$$=\frac{300}{200}\times 100=250$$

Thus, we can say that even though money wage of the worker had increased from ₹330 to ₹500, his real wage has fallen from ₹300 to ₹250. This implies a loss of ₹50 in real terms.

Example 6: If the price of a commodity in a place have decreased by 30% over the base period prices, then the index number of that place is:

Solution: Base price of any commodity = 10 decreased price = 30% of 100 = 30

Index number of that place now = 100-30 =70

Example 7: If with an increase of 10% in prices, the rise in wages is 20% then the real wages has increased by

$$\text{Solution : Real wages}=\frac{\text{Real wage of current year}}{\text{Real wage of base year}}\times 100$$

$$=\frac{120}{100}\times 100$$
=120-100=20%

Example 8: In the year 2010, the monthly salary of clerk was ₹24,000. The consumer price index was 140 in the year 2010, which rises to 224 in 2016. If he has to be rightly compensated, what additional monthly salary to be paid to him?

Solution

Year	CPI	Salary
2010	140	24000
2016	224	X

= 140/224 = 24000/x

$$x=\frac{24000\times 224}{140}=38{,}400$$

DA=38,400-24,000

=14,400

Important Points:

1. Circular Test is an extension of time reversal test.
2. Time Reversal Test and Factor Reversal Test is satisfied by: Fisher's Ideal Index.
3. The ratio of price of single commodity in a given period to its price in the preceding year price is called the relative price.
4. Fisher's Ideal Formula does not satisfy Circular test.
5. The best average for constructing an index number is: Geometric Mean.
6. The time reversal test is satisfied by Fisher's index number.
7. The factor reversal test is satisfied by : Fisher's index.
6. The circular test is satisfied by Simple GM price relative.
7. Fisher's index number is based on Geometric mean of Laspeyre's and Paasche's index numbers.
8. Paasche index is based on: Current year quantities.
9. Fisher's Ideal Formula does not satisfy Circular test.
10. Purchasing Power of Money is the Reciprocal of price index number and inverse relationship between Purchasing power of money and price index number.