ALL FORMULAS

CA FOUNDATION STATISTICS

(100 formulas)

Use this PDF with Formula Revision Marathon

Marathon Link: https://youtu.be/cZJHcOf6OUs

› **Class Limits (CL)** : CL is the minimum and maximum value the class interval may contain

$$CL = UCL - LCL$$

UCL : Upper Class Limit

LCL : Lower Class Limit

1

STATS FORMULA MARATHON

CA Foundation Paper 3

Class Boundary (CB) : These are **actual class limits** of a class interval

› In case of Mutually Exclusive Classification

Class Boundary = Class Limit

$$UCB = UCL \text{ and } LCB = LCL$$

› In case of Mutually Inclusive Classification

$$UCB = UCL + 0.5$$
$$LCB = LCL - 0.5$$

2

Formula Marathon for Maths, LR and Stats by CAPRANAV

Mid Point of Class Interval : average of LCL and UCL or average of LCB and UCB

$$\frac{LCL+UCL}{2}$$

$$\frac{LCB+UCB}{2}$$

Mid point is also called as **Mid Value of a class or Class Mark**

3

Class Length / Width of Class / Size of Class

UCB - LCB

4

Formula Marathon for Maths, LR and Stats by CAPRANAV

Frequency Density

$$\frac{\text{Frequency of a class}}{\text{Class length of that class}}$$

5

STATS FORMULA MARATHON

CA Foundation Paper 3

Relative Frequency

$$\frac{\text{Frequency of class}}{\text{Total Frequency of distribution}}$$

Percentage Frequency

$$\frac{\text{Frequency of class}}{\text{Total Frequency of distribution}} \times 100$$

6

Arithmetic Mean of Discrete Distribution:

$$\overline{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

$$\overline{x} = \frac{\sum x}{n}$$

7

Arithmetic Mean in case of Frequency Distribution :

$$\overline{x} = \frac{\sum fx}{N}$$

Here,

x will be equal to mid point of class interval when frequency distribution is grouped.

N is number of observations which is also equal to total of all frequencies.

8

AM using Assumed Mean / Step Deviation Method

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

where $d = \dfrac{x - A}{C}$

A = assumed mean, C = class length

9

Property of AM

the algebraic sum of deviations of a set of observations from their AM is zero

$$\sum(x - \bar{x}) = 0$$

10

Property of AM

Combined AM

- If there are two groups containing $n_1 \, \& \, n_2$ observations with AMs as $\overline{x_1} \, \& \, \overline{x_2}$ respectively then the combined AM is given by

$$\overline{x}_c = \frac{n_1 \overline{x}_1 + n_2 \overline{x}_2}{n_1 + n_2}$$

11

Median in case of discrete distribution

- *If n = odd, then middle term*
- *If n = even, average of two middle terms*

Same formula will be used for *Simple Frequency distribution*

12

STATS FORMULA MARATHON

CA Foundation Paper 3

Median in case of Grouped Frequency Distribution

- **Step 1**: Prepare a less than type cumulative frequency distribution.

- **Step 2:** Calculate N/2 and check between which class boundaries it falls and call it as Median Class.

- **Step 4:** *Find l_1 = LCB of Median Class, N_u as Cum Freq. of Median Class, N_l as Cum. Freq. of Pre Median Class, C = Class length of Median Class*

- **Step 3**: Apply the below formula

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

13

STATS FORMULA MARATHON   *CA Foundation Paper 3*

Property of Median

*For a set of observations, the sum of **absolute** deviations is **minimum**, when the deviations are taken from the median.*

14

STATS FORMULA MARATHON

CA Foundation Paper 3

Formula Marathon for Maths, LR and Stats by CAPRANAV

Quartiles (in case of discrete observations)

Quartiles
$Q_1 = \left((n+1) \times \dfrac{1}{4} \right)^{th} term$
$Q_2 = \left((n+1) \times \dfrac{2}{4} \right)^{th} term$
$Q_3 = \left((n+1) \times \dfrac{3}{4} \right)^{th} term$

Formula Marathon for Maths, LR and Stats by CAPRANAV

π

Deciles (in case of discrete observations)

Deciles
$D_1 = \left((n+1) \times \dfrac{1}{10} \right)^{th} term$
$D_2 = \left((n+1) \times \dfrac{2}{10} \right)^{th} term$
$Q_9 = \left((n+1) \times \dfrac{9}{10} \right)^{th} term$

16

Percentiles (in case of discrete observations)

Percentiles
$P_1 = \left((n+1) \times \dfrac{1}{100} \right)^{th} term$
$P_2 = \left((n+1) \times \dfrac{2}{100} \right)^{th} term$
$P_{99} = \left((n+1) \times \dfrac{99}{100} \right)^{th} term$

17

Quartiles (Grouped FD)

– For first Quartile Q_1
  › Find Q_1 class
  › Use Formula :

$$Q_1 = l_1 + \left(\frac{\frac{N}{4} - N_l}{N_u - N_l} \right) \times C$$

– For third Quartile Q_3
  › Find Q_3 class
  › Use Formula :

$$Q_3 = l_1 + \left(\frac{\frac{3N}{4} - N_l}{N_u - N_l} \right) \times C$$

18

› Deciles (Grouped FD)

– For first Decile D_1
  › Find D_1 class
  › Use Formula:

$$D_1 = l_1 + \left(\frac{\frac{N}{10} - N_l}{N_u - N_l} \right) \times C$$

– For ninth Decile D_9
  › Find D_9 class
  › Use Formula:

$$D_9 = l_1 + \left(\frac{\frac{9N}{10} - N_l}{N_u - N_l} \right) \times C$$

19

Formula Marathon for Maths, LR and Stats by CAPRANAV

› Percentiles (Grouped FD)

– For first Percentile P_1
  › Find P_1 class
  › Use Formula:

$$P_1 = l_1 + \left(\frac{\dfrac{N}{100} - N_l}{N_u - N_l} \right) \times C$$

– For 99th Percentile P_{99}
  › Find P_{99} class
  › Use Formula:

$$P_{99} = l_1 + \left(\frac{\dfrac{99N}{10} - N_l}{N_u - N_l} \right) \times C$$

20

Mode in case of discrete observation

Observation repeating for maximum no. of times/ max frequency

› *When all observations have same frequency :* No Mode

› *When two observations have max frequency :* Bimodal Observation

› *In general, if we have two or more modes :* Multi Modal Observation

21

› # Mode in case of grouped frequency distribution

   – *First find modal class i.e. class with highest frequency*

   – *Now, Find*

      › f_0 = *frequency of the modal class,*

      › f_{-1} = *frequency of pre modal class,*

      › f_1 = *frequency of the post modal class*

   – Apply formula

$$Mo = l_1 + \left(\frac{f_0 - f_{-1}}{2f_0 - f_{-1} - f_1} \right) \times C$$

22

Relationship between Mean, Median and Mode

In case of symmetrical Distribution

Mean = Median = Mode

23

Relationship between Mean, Median and Mode

In case of moderately skewed distribution

$$\text{Mean} - \text{Mode} = 3 \, (\text{Median} - \text{Median})$$

24

STATS FORMULA MARATHON

CA Foundation Paper 3

Geometric Mean (in case of discrete positive observations)

$$G = \left(x_1 \times x_2 \times \ldots \times x_n \right)^{1/n}$$

25

Geometric Mean (in case of frequency distribution)

$$G = (x_1{}^{f_1} \times x_2{}^{f_2} \times x_3{}^{f_3} \times \cdots \times x_n{}^{f_n})^{1/N}$$

26

Harmonic Mean (in case of discrete observations)

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

27

STATS FORMULA MARATHON

CA Foundation Paper 3

Harmonic Mean (in case of frequency distribution)

$$H = \frac{N}{\Sigma(f/x)}$$

28

Combined HM

$$\bar{x}_c = \frac{n_1 + n_2}{\dfrac{n_1}{H_1} + \dfrac{n_2}{H_2}}$$

29

› Relationship between AM, GM and HM

When to use	Relationship
When all the observations are same	AM = GM = HM
When all the observations are distinct	AM > GM > HM
In General	AM ≥ GM ≥ HM

Range (in case of Discrete Observations)

$$L - S$$

L = Largest Observation

S = Smallest Observation

31

Range (in case of Grouped FD)

$$L - S$$

L = UCB of last class interval

S = LCB of first class interval

32

Coefficient of Range

$$\frac{L-S}{L+S} \times 100$$

33

Mean Deviation (in case of discrete observations)

$$MD_A = \frac{1}{n}\Sigma|x - A|$$

Here A is mean or median or mode as given in question

34

Mean Deviation (in case of grouped frequency distributions)

$$MD_A = \frac{1}{N}\Sigma f|x - A|$$

Here A is mean or median or mode as given in question

35

Coefficient of Mean Deviation

$$\frac{\text{Mean Deviation about A}}{A} \times 100$$

36

CA Foundation Paper 3

Standard Deviation (in case of discrete observations)

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

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

37

Standard Deviation (in case of grouped frequency observations)

$$\sigma_x = SD_x = \sqrt{\frac{\sum f(x - \bar{x})^2}{N}}$$

$$\sigma_x = SD_x = \sqrt{\frac{\sum fx^2}{N} - (\bar{x})^2}$$

38

Coefficient of Variation

$$\frac{SD_x}{\overline{x}} \times 100$$

39

STATS FORMULA MARATHON

CA Foundation Paper 3

For any two numbers, SD = half of range

$$SD = \frac{|a-b|}{2}$$

40

› Standard Deviation of first n natural numbers

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

41

Formula Marathon for Maths, LR and Stats by CAPRANAV

Combined SD

$$SD_c = \sqrt{\frac{n_1 s_1^{\,2} + n_2 s_2^{\,2} + n_1 d_1^{\,2} + n_2 d_2^{\,2}}{n_1 + n_2}}$$

$$d_1 = \bar{x}_c - \bar{x}_1$$

$$d_2 = \bar{x}_c - \bar{x}_2$$

42

Property of SD

If all the observations are constant, SD is $\boxed{\text{Zero}}$

43

Property of SD

› *No effect of change of origin but affected by change of scale in the magnitude (ignore sign)*

44

Quartile Deviation

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

45

Coefficient of Quartile Deviation

$$\frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$$

46

STATS FORMULA MARATHON

CA Foundation Paper 3

Relationship between SD, MD, QD

For a moderately skewed distribution

$$4SD = 5MD = 6QD$$

$$SD : MD : QD = 15 : 12 : 10$$

47

Basic formula of Probability

$$P(A) = \frac{(no.\,of\,favorable\,events\,to\,A)}{Total\,no.\,of\,events}$$

48

CA Foundation Paper 3

Odds in Favor of Event A

$$\frac{\text{no. of favorable events}}{\text{no. of unfavorable events}}$$

49

STATS FORMULA MARATHON

CA Foundation Paper 3

› Odds against an Event

$$\frac{\text{no. of unfavorable events}}{\text{no. of favorable events}}$$

50

› No. of total outcome of a random experiment

If an experiment results in p outcomes and if it is repeated q times then

Total number of outcomes is

$$p^q$$

51

Relative Frequency Probability

$$\text{Relative Frequency} = \frac{\text{no. of times the event occurred during experimental trials}}{\text{total no. of trials}} = \frac{f_A}{n}$$

Probability by this method is defined as

$$P(A) = \lim_{n \to \infty} \frac{f_A}{n}$$

52

› Set Based Probability

$$P(A) = \frac{\text{no.of sample points in A}}{\text{no.of sample points in S}} = \frac{n(A)}{n(S)}$$

53

Addition Theorem 1

In case of two mutually exclusive events A and B

$$P(A \cup B) = P(A+B) = P(A \text{ or } B) = P(A) + P(B)$$

54

Addition Theorem 2

In case of two or more mutually exclusive events

$$P\left(A_1 \cup A_2 \cup A_3 \cup \cdots\right) = P(A_1) + P(A_2) + P(A_3) + \ldots$$

55

STATS FORMULA MARATHON

CA Foundation Paper 3

Addition Theorem 3

For any two events

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

56

π

Addition Theorem 4

In case of any three events

$$P\ (A \cup B \cup C)$$

$$= P(A)\ +\ P(B)\ +\ P(C)\ -\ P(A \cap B)\ -\ P(B \cap C)$$

$$-\ P(A \cap C)\ +\ P(A \cap B \cap C)$$

57

STATS FORMULA MARATHON

CA Foundation Paper 3

Conditional Probability of Event B

knowing that Event A is already occurred

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

$$provided\ P(A) \neq 0$$

58

CA Foundation Paper 3

Formula Marathon for Maths, LR and Stats by CAPRANAV

π

Conditional Probability of Event A

knowing that Event B is already occurred

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

$$provided\ P(B) \neq 0$$

59

STATS FORMULA MARATHON

CA Foundation Paper 3

Compound Theorem

In case of two dependent events

$$P(A \cap B) = P(B) \times P(A/B) \text{ or}$$

$$P(A \cap B) = P(A) \times P(B/A)$$

60

Compound Theorem

In case of two independent events

$$P(A \cap B) = P(A) \times P(B)$$

Same thing applies for two or more independent events

61

CA Foundation Paper 3

Expected value of a Probability Distribution

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

$$E(x) = \mu$$

62

CA Foundation Paper 3

Formula Marathon for Maths, LR and Stats by CAPRANAV

Variance of Probability Distribution

$$V(x) = \sigma^2$$

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

Probability Mass Function in case of Binomial Distribution

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

64

CA Foundation Paper 3

Mean and Variance in case of Binomial Distribution

$$\mu = np$$

$$\sigma^2 = npq$$

65

CA Foundation Paper 3

Formula Marathon for Maths, LR and Stats by CAPRANAV

π

Mode in case of Binomial Distribution

Calculate $(n + 1)p$

if the resulting value is integer then Bi-modal	$\mu_0 = (n + 1)p$ and $[(n + 1)p - 1]$
If the resulting value is non-integer then Uni-modal	$\mu_0 = $ largest integer contained in $(n + 1)p$

66

Probability Mass Function in case of Poisson Distribution

$$f(x) = P(X = x) = \frac{(e^{-m}.m^x)}{x!}$$

67

Formula Marathon for Maths, LR and Stats by CAPRANAV

π

Mean, Variance and SD in case of Poisson Distribution

$$\mu = m$$

$$\sigma^2 = m$$

$$\sigma = \sqrt{m}$$

68

Mode in case of Poisson Distribution

if the value of m is integer then Bi-modal	$\mu_0 = m$ *and* $[m-1]$
If the value of m is non-integer then Uni-modal	$\mu_0 = $ *largest integer contained in m*

69

STATS FORMULA MARATHON   *CA Foundation Paper 3*

Probability Density Function in case of Normal Distribution

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

70

Mean Deviation in case of Normal Distribution

$$MD = 0.8\,\sigma$$

71

CA Foundation Paper 3

π

Quartiles in case of Normal Distribution

$$Q_1 = \mu - 0.675\sigma$$

$$Q_3 = \mu + 0.675\sigma$$

72

CA Foundation Paper 3

Quartile Deviation in case of Normal Distribution

$$QD = 0.675\sigma$$

73

STATS FORMULA MARATHON

CA Foundation Paper 3

Formula Marathon for Maths, LR and Stats by CAPRANAV

π

Points of Inflexion

$$\mu - \sigma \quad \& \quad \mu + \sigma$$

74

Ratio between

$$QD : MD : SD = 10:12:15$$

75

Conditions of Standard Normal Distribution

Mean μ	0
Standard Deviation σ	1

76

STATS FORMULA MARATHON

CA Foundation Paper 3

Z Value

$$Z = \frac{x - \mu}{\sigma}$$

77

Area under Normal Curve

From	To	Area/Probability
μ	+σ	34.135%
+σ	+2σ	13.59%
+2σ	+3σ	2.14%
3σ	∞	0.135%

78

For a bivariate frequency distribution

No. of marginal distributions = 2

No. of conditional distributions = m + n

79

Karl Pearson's Product Moment Correlation Coefficient

$$r_{xy} = \frac{Cov(x, y)}{\sigma_x . \sigma_y}$$

80

Covariance

$$Cov(x, y) = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{n} \ or \ \frac{\Sigma xy}{n} - \bar{x}.\bar{y}$$

81

Spearman's Rank Correlation Coefficient

$$r_R = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}$$

here d means difference in ranks

82

Formula Marathon for Maths, LR and Stats by CAPRANAV

π

Adjustment value in case of Tie

$$\Sigma \frac{(t^3 - t)}{12}$$

83

CA Foundation Paper 3

Coefficient of Concurrent Deviations

$$r_c = \pm\sqrt{\pm\frac{2c - m}{m}}$$

STATS FORMULA MARATHON

CA Foundation Paper 3

Regression Coefficients

$$b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x} \ \ or \ \ b_{yx} = \frac{Cov(x,y)}{\sigma_x^{\,2}}$$

$$b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y} \ \ or \ \ b_{xy} = \frac{Cov(x,y)}{\sigma_y^{\,2}}$$

85

CA Foundation Paper 3

π

Regression Property

$$r = \pm \sqrt{b_{xy} \times b_{yx}}$$

b_{xy}, b_{yx} and r all will have same sign

86

Regression Property

› Change of Origin and Scale

- *Origin: No Impact*
- *Scale: If original pair is x, y and modified pair is u, v*

then,

$$b_{vu} = b_{yx} \times \frac{change\ of\ scale\ of\ y}{change\ of\ scale\ of\ x}$$

$$b_{uv} = b_{xy} \times \frac{change\ of\ scale\ of\ x}{change\ of\ scale\ of\ y}$$

87

Regression Property

› Intersection of two regression lines:

— *Two regression lines (if not identical) will intersect at the point [means]*

$$(\overline{x}, \overline{y})$$

88

Coefficient of determination

Explained Variance

Accounted Variance

$$r^2$$

CA Foundation Paper 3

89

Formula Marathon for Maths, LR and Stats by CAPRANAV

π

Coefficient of non-determination

Unexplained Variance

Unaccounted Variance

$$1 - r^2$$

90

Probable Error

$$0.6745 \times \frac{1 - r^2}{\sqrt{N}}$$

– *where, r = correlation coefficient, N = no. of pairs of observation*

91

Limits of Population Correlation Coefficient

$$r \pm PE$$

92

Index Relatives

Price Relative	Price Relative $= \dfrac{P_n}{P_0}$
Quantity Relative	Quantity Relative $= \dfrac{Q_n}{Q_0}$
Value Relative	Value Relative $= \dfrac{V_n}{V_0} = \dfrac{P_n Q_n}{P_0 Q_0}$
Link relative	$$\frac{P_1}{P_0}, \frac{P_2}{P_1}, \frac{P_3}{P_2}, \cdots, \frac{P_n}{P_{n-1}}$$ *Same can be created for quantities also*
Chain relatives	*When the above relatives are in respect to a fixed base period these are also called the chain relatives* $$\frac{P_1}{P_0}, \frac{P_2}{P_0}, \frac{P_3}{P_0}, \cdots, \frac{P_n}{P_0}$$

93

Simple Aggregative Method of Index

Price Index is expressed as total of commodity prices in a given year as a percentage of total of commodity prices in the base year

› Formula

$$\frac{\Sigma P_n}{\Sigma P_0} \times 100$$

94

Simple Average of Relatives – Method of Index

› *Under this method, we invert the actual price for each variable into percentage of the base period. These percentages are called relatives.*

› Formula

$$\frac{\Sigma \frac{P_n}{P_0}}{N}$$

95

Weighted Index Methods

› *Laspeyres' Index*

$$\frac{\Sigma P_n Q_0}{\Sigma P_0 Q_0} \times 100$$

96

STATS FORMULA MARATHON

CA Foundation Paper 3

Weighted Index Methods

› *Passche's Index*

$$\frac{\Sigma P_n Q_n}{\Sigma P_0 Q_n} \times 100$$

97

STATS FORMULA MARATHON

CA Foundation Paper 3

Weighted Index Methods

› *Marshall-Edgeworth Index*

$$\frac{\Sigma P_n(Q_0 + Q_n)}{\Sigma P_0(Q_0 + Q_n)} \times 100$$

98

Weighted Index Methods

› *Fisher's Index*

(GM of Laspeyres' and Paasche's)

$$\sqrt{\frac{\Sigma P_n Q_0}{\Sigma P_0 Q_0} \times \frac{\Sigma P_n Q_n}{\Sigma P_0 Q_n}} \times 100$$

99

Formula Marathon for Maths, LR and Stats by CAPRANAV

Deflated Value

$$\text{Deflated Value} = \frac{\text{Current Value}}{\text{Price Index of the current year}}$$

100