Statistics and Probability
STA-301
Manual Formulas

1. Mode in case of frequency distribution of a continuous variable: $\hat{X}=l+\frac{f_m-f_1}{(f_m-f_1)+(f_m-f_2)}\,x\,h$

l = lower class boundary of the modal class,
f_m = frequency of the modal class,
f_1 = frequency of the class preceding the modal class,
f_2 = frequency of the class following modal class, and
h = length of class interval of the modal class

2. Arithmetic Mean in case of raw data: $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$

3. Arithmetic Mean in case of frequency distribution: $\overline{X} = \frac{\sum_{i=1}^{k} f_i X_i}{\sum_{i=1}^{k} f_i} = \frac{\sum_{i=1}^{k} f_i X_i}{n}$

4. Weighted Mean: $\overline{X}_w = \frac{\sum W_i X_i}{\sum W_i}$

5. Median: ***The central value.*** (In case of raw data)
6. Median in case of frequency distribution of a continuous variable: $\tilde{X} = l + \frac{h}{f}\left(\frac{n}{2} - c\right)$

Where
l =lower class boundary of the median class (i.e. that class for which the cumulative frequency is just in excess of n/2).
h=class interval size of the median class
f =frequency of the median class
n=Σf (the total number of observations)
c =cumulative frequency of the class preceding the median class

7. If Mean=Median=Mode it will be a symmetric curve.
8. Median - Mode 2 (Mean – Median) = Right skewed distribution.
9. Mean – Mode = 3 (Mean – Median) = Right skewed distribution.
10. Mode = 3 Median – 2 Mean.
11. First Quartile= $Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - c\right)$
12. Second quartile (i.e. median): $Q_2 = l + \frac{h}{f}\left(\frac{2n}{4} - c\right) = l + \frac{h}{f}\left(n/2 - c\right)$

Just modify the formula for 3rd quartile with 3n instead of 2n.

13. First decile: $D_1 = l + \frac{h}{f}\left(\frac{n}{10} - c\right)$
14. Second decile: $D_2 = l + \frac{h}{f}\left(\frac{2n}{10} - c\right)$

And so on.

15. First percentile: $P_1 = l + \frac{h}{f}\left(\frac{n}{100} - c\right)$
16. Second percentile: $P_2 = l + \frac{h}{f}\left(\frac{2n}{100} - c\right)$ And so on.

ABDUL SABOOR www.vchowk.com **bc080400280@vu.edu.pk**

17. Geometric Mean in case of raw data: $G = \sqrt[n]{X_1 X_2 \ldots X_n}$ $G = \text{anti}\log\left[\frac{\sum \log X}{n}\right]$

18. Geometric Mean in case of frequency distribution: $G = \sqrt[n]{X_1^{f_1} X_2^{f_2} \ldots X_k^{f_k}}$ $G = antil\,\text{og}\left[\frac{\sum f \log X}{n}\right]$

19. Harmonic Mean in case of raw data: $H.M. = \frac{n}{\sum\left(\frac{1}{X}\right)}$

20: Harmonic Mean in case of frequency distribution: $H.M. = \frac{n}{\sum f\left(\frac{1}{X}\right)}$

21. Where we should use Arithmetic, Harmonic, and Geometric mean?
There are three rules. Keep them in mind.
i. When values are given as x per y where x is constant and y is variable, the Harmonic Mean is the appropriate average to use.
ii. When values are given as x per y where y is constant and x is variable, the Arithmetic Mean is the appropriate average to use.
iii. When relative changes in some variable quantity are to be averaged, the geometric mean is the appropriate average to use.

22. RELATION BETWEEN ARITHMETIC, GEOMETRIC AND HARMONIC MEANS:

Arithmetic Mean ≥ Geometric Mean ≥Harmonic Mean.

23. Mid-Range: $\text{mid} - \text{range} = \frac{x_0 + x_m}{2}$

24. Mid-Quartile Range (Mid-Hinge): $\text{mid} - \text{quartile range} = \frac{Q_1 + Q_3}{2}$

25. Range: R = Xm – X0

26. Mid-range: $= \frac{X_m - X_0}{2}$

27. Co-efficient of dispersion: $= \frac{X_m - X_0}{X_m + X_0}$

28. Quartile Deviation (semi inter-quartile range): $Q.D. = \frac{Q_3 - Q_1}{2}$

29. Co-efficient of quartile deviation: $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

30. Mean deviation: $\text{M.D.} = \frac{\sum |d|}{n}$

31. Mean grouped data: $\text{M.D.} = \frac{\sum f_i \left|x_i - \bar{x}\right|}{n} = \frac{\sum f_i \left|d_i\right|}{n}$

32. Co-efficient of mean deviation: $= \frac{\text{M.D.}}{\text{Mean}}$

33. Variance: $= \frac{\sum (x - \bar{x})^2}{n}$

34. Standard deviation in case of raw data: $S = \sqrt{\left\{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2\right\}}$

35. Standard deviation in case of grouped data: $S = \sqrt{\left\{ \frac{\sum fx^2}{n} - \left(\frac{\sum fx}{n} \right)^2 \right\}}$

36. Co-efficient of standard deviation: $= \frac{\text{Standard Deviation}}{\text{Mean}}$

37. Co-efficient of variation: $\text{C.V.} = \frac{S}{\overline{X}} \times 100$

***38. Chebychev's theorem:** For any number k greater than 1, at least 1 – 1/k2 of the data-values fall within k standard deviations of the mean, i.e., within the interval ($\overline{X}$ – kS, $\overline{X}$ + kS).*

This means that:

a) At least 1-1/22 = 3/4 will fall within 2 standard deviations of the mean, i.e. within the interval ($\overline{X}$ – 2S, $\overline{X}$ + 2S).

b) At least 1-1/32=8/9 of the data-values will fall within 3 standard deviations of the mean, i.e. within the interval ($\overline{X}$ – 3S, $\overline{X}$ + 3S).

38. Empirical rule: (a) Approximately 68% of the measurements will fall within 1 standard deviation of the mean, i.e. within the interval ($\overline{X}$ – S, $\overline{X}$ + S)

(b) Approximately 95% of the measurements will fall within 2 standard deviations of the mean, i.e. within the interval ($\overline{X}$ – 2S, $\overline{X}$ + 2S).

(c) Approximately 100% (practically all) of the measurements will fall within 3 standard deviations of the mean, i.e. within the interval ($\overline{X}$ – 3S, $\overline{X}$ + 3S).

39. Right skewed distribution: median < mid-quartile range < midrange:

40. Left skewed distribution: midrange < mid-quartile range < median

41. Pearson co-efficient of skewness: $\frac{mean - mode}{standard\ deviation} = \frac{3\left(\text{mean} - \text{median}\right)}{\text{standard deviation}}$

42. Bowley's coefficient of skewness: $= \frac{\left(Q_1 + Q_3 - 2\tilde{X}\right)}{Q_3 - Q_1}$

43. MEASURE of kurtosis: $K = \frac{Q.D.}{P_{90} - P_{10}},$

***In case of a leptokurtic distribution, the percentile coefficient of kurtosis comes out to be LESS THAN 0.263, and in the case of a platykurtic distribution, the percentile coefficient of kurtosis comes out to be GREATER THAN 0.263**.*

44. Moments:

$$m_1 = \frac{\sum\left(x_i - \overline{x}\right)}{n}, \quad m_2 = \frac{\sum\left(x_i - \overline{x}\right)^2}{n}, \quad m_3 = \frac{\sum\left(x_i - \overline{x}\right)^3}{n}, \quad m_4 = \frac{\sum\left(x_i - \overline{x}\right)^4}{n}$$

$m_1 = 0$

45. Relationship between moments:

$$m_2 = m'_2 - (m'_1)^2;$$

$$m_3 = m'_3 - 3\ m'_2\ m'_1 + 2\ (m'_1)^3, \text{ and}$$

$$m_4 = m'_4 - 4\ m'_3\ m'_1 + 6\, m'_2\ (m'_1)^2 - 3\ (m$$

46. Moments ratios: $b_1 = \frac{(m_3)^2}{(m_2)^3}$ *and* $b_2 = \frac{m_4}{(m_2)^2}$

For symmetrical distributions, b1 is equal to zero

ABDUL SABOOR **bc080400280@vu.edu.pk**

For the normal distribution, b2 = 3
For a leptokurtic distribution, b2 > 3,
And for a platykurtic distribution, b2 < 3

47. Standard error of estimate: $s_{y.x} = \sqrt{\dfrac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$

48. Pearson co-efficient of correlation: $r = \dfrac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}}$

49. Co-variance (used in the above formula): $Cov(X,Y) = \dfrac{\sum (X-\bar{X})(Y-\bar{Y})}{n}$

50. Short cut formula for Pearson co-efficient of correlation:

$$r = \frac{\sum XY - \left(\sum X\right)\left(\sum Y\right)/n}{\sqrt{\left[\sum X^2 - \left(\sum X\right)^2/n\right]\left[\sum Y^2 - \left(\sum Y\right)^2/n\right]}}$$

51. Sets laws: **Commutative laws:**

$A \cup B = B \cup A$ and
$A \cap B = B \cap A$

Associative laws:

$(A \cup B) \cup C = A \cup (B \cup C)$
and
$(A \cap B) \cap C = A \cap (B \cap C)$

Distributive laws

$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
and $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Idempotent laws

$A \cup A = A$ and
$A \cap A = A$

Identity laws

$A \cup S = S$,
$A \cap S = A$,
$A \cup \phi = A$, and
$A \cap \phi = \phi$.

Complementation laws

$A \cup \bar{A} = S$,
$A \cap \bar{A} = \phi$,

$(\bar{\bar{A}}) = A$,

$\bar{S} = \phi$, and
$\bar{\phi} = S$

De Morgan's laws: $\overline{(A \cup B)} = \bar{A} \cap \bar{B}$,
and $\overline{(A \cap B)} = \bar{A} \cup \bar{B}$

52. Permutations: nPr= $= \dfrac{n!}{(n-r)!}$. ${}^{n}P_r = r!\dbinom{n}{r}$

53. Combinations: ${}^{n}C_r$ or $\dbinom{n}{r}$, $\dbinom{n}{r} = \dfrac{n!}{r!(n-r)!}$

ABDUL SABOOR **bc080400280@vu.edu.pk**

54. nCr is also called binomial coefficient: $(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^{n-r}\ b^r.$

55. Mathematical Expectations: $E(X) = \sum_{i=1}^{n} x_i f(x_i)$

E(X) is also called the mean of X and is usually denoted by the letter μ.

56. If the values are equally likely: $E(X) = \frac{1}{n}\sum x_i,$

57. In case of discrete random variable E(X)= $= \sum_i H(x_i) f(x_i),$

58. Mean on probability distribution: μ= E(X)=∑xP(x)

59. Expected value of variance : Var(X) or σ2=E(X – μ)2 = Σ(xi - μ)2 f(x)

60. Short cut formula for variance: σ2 = E(X2) – [E(X)]2.

61. kth moment about the origin of the random variable X : ′ μk= E(Xk) = Σxik f(x)

62. kth moment about the mean of the random variable: μk = E(X – μ)k = Σ(xi – μ)k f(x)

63. Skew ness of a probability distribution: $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$

64. kurtosis of a probability distribution: $\beta_2 = \frac{\mu_4}{\mu_2^2}.$

65. **PROPERTIES OF MATHEMATICAL EXPECTATION:**

a. If c is a constant, then E(c) = c.

b. If X is a discrete random variable and if a and b are constants, then E(aX + b) = a E(X) + b.

66. Chebychev's Inequality: If X is a random variable having mean μ and variance σ2 > 0, and k is any positive constant, then the probability that a value of X falls within k standard deviations of the mean is at least. That is: $P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2},$

67. The function f(x) is called the probability density function if x is a random variable. <u>Properties:</u>

i) f(x) $\geq$ 0, for all x

ii) $\int_{-\infty}^{\infty} f(x)\ dx = 1$

iii) The probability that X takes on a value in the interval [c, d], c < d is given by:

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

68. To compute distribution function of a continuous probability function: **F(x) = P(X < x) =** $\int_{-\infty}^{x} f(x)\ dx$

ABDUL SABOOR **dr7wb@yahoo.com**

69. Mathematical expectation of *continuous* random variable: $E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$

70. Properties of a loint probability function:

   i) f(xi, yj)>0,for all(xi, yj),i.e. for i=1,2,…,m; j = 1, 2, …, n.

   ii) $\sum_i \sum_j f(x_i, y_j) = 1$

71. Marginal probability function of X: $g(x_i) = \sum_{j=1}^{n} f(x_i, y_j)$

72. Marginal probability function of Y: $h(y_j) = \sum_{i=1}^{m} f(x_i, y_j) = P(Y = y_j)$

73. The marginal p.d.f of random variable X: $g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$

74. The marginal p.d.f of random variable Y: $h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$

75. ***Two properties of mathematical expectations:***

   i. E(X + Y) = E(X) + E(Y). or E(X – Y) = E(X) – E(Y)
   ii. E(XY) = E(X) E(Y).

76. Expected value of X in a joint continuous probability distribution: E(X) = ∑ xj g(xj)
77. Expected value of Y in a joint continuous probability distribution: E(Y) = ∑ yi h (yi)
78. To compute the E(X+Y): $E(X+Y) = \sum_i \sum_j (x_i + y_j)\, f(x_i, y_j)$
79. To compute the E(XY): $E(XY) = \sum_i \sum_j (x_i y_j)\, f(x_i, y_j)$

80. Expected value of X in a j.p.d (continuous): $E(X) = \int_{-\infty}^{\infty} x\, g(x)\, dx$

81. Expected value of Y in a j.p.d (continuous): $E(Y) = \int_{-\infty}^{\infty} y\ h(y)\, dy$

82. Co-variance of two random variables X & Y: Cov (X, Y) = E(XY) – E(X) E(Y).

Cov (X, Y) = E {[X – E(X)] [Y – E(Y)]}

If x and Y are independent then: E(XY) = E(X) E(Y), and Cov (X, Y) = E(XY) – E(X) E(Y) =0

ABDUL SABOOR **admin@vchowk.com**

83. Correlation co-efficient of two random variables:

$$\rho_{XY} = \frac{E\left[X - E(X)\right]\left[Y - E(X)\right]}{\sigma_X \sigma_Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}}$$

84. PROPERTIES OF A BINOMIAL EXPERIMENT:

1. ***Every trial results in a success or a failure.***
2. ***The successive trials are independent.***
3. ***The probability of success, p, remains constant from trial to trial.***
4. ***The number of trials, n, is fixed in advanced.***

85. The binomial distribution: $P(X = x) = \binom{n}{x} p^x q^{n-x}$

Where
n = the total no. of trials
p = probability of success in each trial
q = probability of failure in
each trial (i.e. q = 1 - p)
x = no. of successes in n trials.
x = 0, 1, 2, … n

86. For binomial distribution: E(X) = np and Var(X) = npq

86. Standard deviation of binomial distribution: $S.D.(X) = \sqrt{npq}$

87. co-efficient of variation of binomial distribution: $C.V. = \frac{\sigma}{\mu} \times 100$

88. Skew ness of binomial distribution:
If p=q it will be a symmetric distribution.
p<q it will be right skewed.
p>q it will be left skewed.

89. ***PROPERTIES OF HYPERGEOMETRIC EXPERIMENT:***

i) The outcomes of each trial may be classified into one of two categories, success and failure.
ii) The probability of success changes on each trial.
iii) The successive trials are not independent.
iv) The Experiment is repeated a fixed number of times.

90. Hyper geometric probability distribution: $P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$,

where
N = number of units in the population,
n = number of units in the sample, and
k = number of successes in the population.

91. Mean of the hyper geometric p.d: $\mu = n\frac{k}{N}$

92. Variance of hyper geometric p.d: $\sigma^2 = n\frac{k}{N}\frac{N-k}{N}\frac{N-n}{N-1}$,

ABDUL SABOOR **dr7wb@yahoo.com**

93. Limiting the binomial p.d: $\lim_{\substack{n\to\infty \\ p\to 0}} b(x;n,p)=\frac{e^{-\mu}\mu^x}{x!}, \quad x=0,1,2,\ldots,\infty$
Where e=2.71828 and μ= np

The Poisson distribution has only one parameter μ > 0.

94. Poisson process: $P(X=x)=\frac{e^{-\lambda t}(\lambda t)^x}{x!},$
Where

λ = average number of occurrences of the outcome of interest per unit of time,
t = number of time-units under consideration, and
x = number of occurrences of the outcome of interest in t units of time.

95. Uniform distribution: $\int_{-\infty}^{\infty} f(x)dx=\int_a^b \frac{1}{b-a}dx=\frac{1}{b-a}[x]_a^b=\frac{b-a}{b-a}=1$

96. U.D can also be measured by the simple rectangle formula= (base) x (height) $=(b-a)\times\left(\frac{1}{b-a}\right)=1$

97. Mean and variance of U.D: $\mu=\frac{a+b}{2}$ **and <u>variance</u> is** $\sigma^2=\frac{(b-a)^2}{12}$

98. Normal distribution: $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}, \quad -\infty<x<\infty$

$$\begin{pmatrix} \text{where} \\ \pi = 3.1416 \simeq 22/7, \\ e \simeq 2.71828 \end{pmatrix}$$

99. Moment rations of normal distribution: $\beta_1=\frac{\mu^2{}_3}{\mu_2{}^3}=\frac{0^2}{(\sigma^2)^3}=0, \quad \beta_2=\frac{\mu^4}{\mu_2{}^2}=\frac{3\sigma^4}{(\sigma^2)^2}=3$

100. μ ± σ= 68.26% of the total area.

μ ± 2σ= 95.44% of the total area.

μ ± 3σ= 99.73% of the total area.

101. The standardization formula: $Z=\frac{X-\mu}{\sigma}$

102. Mean of the population distribution: $\mu=E(X)=\sum xf(x)=2$

103. Variance of the population distribution:
$$\sigma^2=Var(X)=E(X)^2-[E(X)]^2$$
$$=\sum x^2 f(x)-\left[\sum x\,f(x)\right]^2$$
$$=6-2^2=6-4=2$$

104. Mean of the sampling distribution of X: $\mu_{\bar{x}}=E(\bar{X})=\sum \bar{x}f(\bar{x})$

Atiq Muhammad **bc080200822@vu.edu.pk**

105. Variance of the sampling distribution of X:

$$\sigma^2_{\bar{x}} = Var(\bar{X}) = E(\bar{X})^2 - [E(\bar{X})]^2 = \sum \bar{x}^2 f(\bar{x}) - [\sum \bar{x} f(\bar{x})]^2$$

106. standard deviation: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$

107. $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n} \frac{N-n}{N-1}},$

108. Standard error of p^: $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$

109. Modified formula for sample variance: $s^2 = \frac{\sum (x-\bar{x})^2}{n-1}$

110. Exponential distribution: $f(x) = \theta e^{-\theta x}, \; x > 0, \; \theta > 0$,

111. the standardized version of $\bar{X}$: $Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

112. Confidence interval for mew: $\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$

113. **Confidence Interval for the difference between the means of two Populations (i.e. μ1 – μ2):**

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \sqrt{\frac{s_1^{\;2}}{n_1} + \frac{s_2^{\;2}}{n_2}}$$

114. **Confidence Interval for a Population Proportion (p):** $\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

115. **Confidence Interval for p1-p2:** $(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

116. test statistics: $z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

117. Test statistics for two populations: $Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

118. $\chi^2 = \frac{nS^2}{\sigma_0^2}$

119. The F distribution: $f(x) = \frac{\Gamma[(\nu_1+\nu_2)/2] \, (\nu_1/\nu_2)^{\nu_1/2} \; x^{(\nu_1/2)-1}}{\Gamma(\nu_1/2)\,\Gamma(\nu_2/2)\,[1+\nu_1 x/\nu_2]^{(\nu_1+\nu_2)/2}}, \quad 0 < x < \infty$

Atiq Muhammad **bc080200822@vu.edu.pk**

120. $$\sigma^2 = \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}$$

120. Conditional probability: $$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

121. Bay's theorem: $$P(A_i / B) = \frac{P(A_i)P(B/A_i)}{\sum_{i=1}^{k} P(A_i)P(B/A_i)},$$

for i = 1, 2, …, k.

$$P(A_1 \mid B) = \frac{P(A_1).P(B \mid A_1)}{P(A_1).P(B \mid A_1) + P(A_2).P(B \mid A_2)}$$

122. Standard deviation: $$S = \sqrt{\frac{\sum f(X-\overline{X})^2}{\sum f}} \qquad \sqrt{\frac{\sum X^2 f}{\sum f} - \left(\frac{\sum Xf}{\sum f}\right)^2}$$

123. in case of probability distribution: = S.D.(X) $$= \sqrt{\frac{\sum X^2 P(X)}{\sum P(X)} - \left[\frac{\sum XP(X)}{\sum P(X)}\right]^2}$$

$$= \sqrt{\sum X^2 P(X) - [\sum XP(X)]^2}$$

$$\left(\because \quad \sum P(X) = 1\right)$$

124. Co-efficient of variation: $$C.V. = \frac{\sigma}{\mu} \times 100$$

ABDUL SABOOR BS BA MARKETING VU **bc080400280@vu.edu.pk**