Formula Sheet

Range (R)= \text{highest(data value)} - \text{lowest(data value)}

Class width=\frac{R}{\text{no. of Class}}

Midpoint=\frac{\text{lower limit+upper limit}}{2}

For pie graph: Degrees=\frac{L \times 360^\circ}{\text{no. of Classes}}

Sample Mean \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}

Population Mean \mu = \frac{\sum_{i=1}^{N} X_i}{N}

Sample Mean (with frequency) \bar{X} = \frac{\sum_{i=1}^{n} f_i X_i}{n}

Midrange MR=\frac{\text{highest value}+\text{lowest value}}{2}

Weighted Mean \bar{X} = \frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i}

Harmonic Mean HM=\frac{\sum_{i=1}^{n} \frac{1}{w_i}}{\sum_{i=1}^{n} \frac{1}{w_i}}

Geometric Mean GM=\sqrt[n]{(X_1)(X_2)(X_3)\cdots(X_n)}

Quadratic Mean QM=\sqrt{\frac{\sum_{i=1}^{n} X_i^2}{n}}

Sample Variance s^2 = \frac{\sum_{i=1}^{n} X_i^2 - \left[\frac{\sum_{i=1}^{n} X_i}{n}\right]^2}{n-1}

Standard Deviation s = \sqrt{s^2}

Sample Variance (with frequency) s^2 = \frac{\sum_{i=1}^{n} f_i X_i^2 - \left[\frac{\sum_{i=1}^{n} f_i X_i}{n}\right]^2}{n-1}

Population Variance \sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}

Coefficient of Variation CV=\frac{\text{standard deviation}}{\text{mean}} \times 100\%

Chebyshev’s theorem: The proportion of values from a data set that will fall within \(k\) standard deviations of the mean will be at least \(1 - \frac{1}{k^2}\), where \(k\) is a number greater than 1

Z-score=\frac{\text{data value-mean}}{\text{standard deviation}}

percentile rank of X=\frac{\text{no. of data values below X}}{\text{total no. of data values}} \times 100\%

Finding \(k^{th}\) percentile data value, \(L=(\frac{k}{100})x(n)\), where \(n=\text{total no. of values and } k=\text{percentile}\). If \(L\) is a whole number then \(k^{th}\) percentile is the average of the \(L^{th}\) and \((L+1)^{th}\) data values. If \(L\) has a decimal value, round \(L\) up then \(k^{th}\) percentile is the rounded \(L^{th}\) data value.

Interquartile Range(IQR), IQR=\(Q_3 - Q_1\), where \(Q_1\) is the first quartile and \(Q_3\) is the third quartile

To check the outliers, data value which is smaller than \(Q_1 - 1.5 \times (IQR)\) or larger than \(Q_3 + 1.5 \times (IQR)\)

Permutation: The arrangements of \(n\) objects in a specific order using \(r\) objects at a time is called a permutation of \(n\) objects taking \(r\) objects at a time, i.e., \(nPr = \frac{n!}{(n-r)!}\)

Combination: The number of combinations of \(r\) objects selected from \(n\) objects is obtained by \(nCr = \frac{n!}{r!(n-r)!}\)

The conditional probability of an event \(B\) given an event \(A\) is \(P(B|A) = \frac{P(A \text{ and } B)}{P(A)}\)
Expected value of a discrete random variable is
\[ E(X) = \sum X P(X) \]
\[ V(X) = \sum X^2 P(X) - [\sum X P(X)]^2 \]
In a binomial experiment, the probability of getting exactly \( X \) success in \( n \) trials is
\[ P(X) = \binom{n}{X} \times p^X \times q^{n-X}, \quad X = 0, 1, \ldots, n, \quad E(X) = np, \quad V(X) = npq \]
The Poisson distribution with parameter \( \lambda \) is
\[ P(X; \lambda) = \frac{e^{-\lambda} \lambda^X}{X!}, \quad \text{where} \quad X = 0, 1, 2, \ldots, \quad E(X) = \lambda, \quad V(X) = \lambda \]
For Hypergeometric distribution
\[ P(X) = \binom{\frac{X}{n}}{\frac{\lambda}{\lambda}} \]
Normal distribution of a continuous random variable \( y \)
\[ f(y | \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(y - \mu)^2}{2\sigma^2}} \]
\[ E(y) = \mu \]
\[ V(y) = \sigma^2 \]
Standard Normal distribution of a continuous random variable \( y \)
\[ f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \]
Large sample 100(1 - \( \alpha \))% confidence interval for \( \mu \) is \( \hat{X} \pm z_{\alpha/2} \left( \frac{\sigma}{\sqrt{n}} \right) \)
Small sample 100(1 - \( \alpha \))% confidence interval for \( \mu \) is \( \hat{X} \pm t_{\alpha/2, n-1} \left( \frac{\sigma}{\sqrt{n}} \right) \)
Large sample 100(1 - \( \alpha \))% confidence interval for \( \sigma^2 \) is \( \hat{\sigma}^2 \pm z_{\alpha/2} \left( \frac{\hat{\sigma}^2}{n} \right) \)
Sample size \( n = \left( \frac{z_{\alpha/2} \sigma}{0.5 \mu} \right)^2 \), \( \mu \) is the minimum error of estimation
When \( \sigma \) is known the test-statistic \( Z = \frac{\hat{X} - \mu}{\sigma / \sqrt{n}} \)
When \( \sigma \) is unknown the test-statistic \( t = \frac{\hat{X} - \mu}{\hat{\sigma} / \sqrt{n}} \)
Test-statistic with a specific population proportion \( \hat{p} = \frac{\hat{X}}{n} \)
Confidence interval for \( \sigma^2 \) is \( \frac{(n-1)s^2}{\chi^2_{n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{n-2}} \)
Test-statistic for testing a claim about \( \sigma \) or \( \sigma^2 \) is \( \chi^2 = \frac{(n-1)s^2}{\sigma^2} \)
When population variances are known the test-statistic \( Z = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2} \)
Confidence interval for the difference between two population means is \( (\bar{X}_1 - \bar{X}_2) \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \)
Z-test for the difference between two population proportions, \( z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}_1 \hat{q}_1/n_1 + \hat{p}_2 \hat{q}_2/n_2}} \)
where \( \hat{p} = \frac{X_1 + X_2}{n_1 + n_2} \), \( \hat{q}_1 = \frac{X_1}{n_1} \), \( \hat{q}_2 = \frac{X_2}{n_2} \), \( \tilde{q} = 1 - \hat{p} \)
Confidence interval for the difference between two proportions is \( (\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1/n_1 + \hat{p}_2 \hat{q}_2/n_2}{n_1 + n_2}} \)
Correlation Co-efficient \( r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2] [n \sum y^2 - (\sum y)^2]}} \)
and \( t = \frac{n - 2}{\sqrt{\frac{\hat{r}^2}{n - 2}}} \) for testing \( H_0: \rho = 0 \) and \( H_1: \rho \neq 0. \)
For Regression \( a = \frac{\sum y \sum x^2 - (\sum x)(\sum xy)}{n \sum x^2 - (\sum x)^2} \) \( b = \frac{\sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2} \)