Formula Sheet

Range (R)=$highest(data\ value) - lowest(data\ value)$

Class width=$R/(no.\ of\ Class)$

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

For pie graph: Degrees=$\frac{f}{n} \times 360^0$ Or $\% = \frac{f}{n} \times 100\%$

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=$\frac{lowest\ value+highest\ value}{2}$

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

Harmonic Mean HM=$\frac{n}{\sum_{i=1}^{n}(1/X_i)}$

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

Quadratic Mean QM=$\sqrt{\left(\frac{\sum_{i=1}^{n} X_i^2}{n}\right)}$

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

Standard Deviation $s = \sqrt{s^2}$

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

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

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

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=$\frac{data\ value - mean}{standard\ deviation}$

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

Finding k^{th} percentile data value, L=$(\frac{k}{100}) \times (n)$, where n=total no. of values and k=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)=$\sum XP(X)$

V(X)=$\sum X^2P(X) - [\sum XP(X)]^2$

In a binomial experiment, the probability of getting exactly X successes in n trials is

P(X)=$\frac{n!}{(n-X)!X!}.p^X q^{n-X}$, $X = 0, 1, ..., n,\ E(X) = np,\ V(X) = npq$

The Poisson distribution with parameter λ is

$P(X;\lambda) = \frac{e^{-\lambda}\lambda^X}{X!}$, where X = $0, 1, 2, ...$, E(X) = λ, V(X) = λ

For Hypergeometric distribution

P(X)=$\frac{\binom{a}{X}\binom{b}{n-X}}{\binom{a+b}{n}}$.

Normal distribution of a continuous random variable y

$f(y|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\Pi}}e^{-\frac{1}{2}(\frac{y-\mu}{\sigma})^2}$

E(y)=μ

V(y)=σ^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 μ is $\bar{X} \pm z_{\alpha/2}\ (\sigma/\sqrt{n})$

Small sample $100(1-\alpha)\%$ confidence interval for μ is $\bar{X} \pm t_{\alpha/2,\, n-1}\ (s/\sqrt{n})$

Large sample $100(1-\alpha)\%$ confidence interval for p is $\hat{p} \pm z_{\alpha/2}\ (\sqrt{\hat{p}\hat{q}/n})$

Sample size n=$(\frac{z_{\alpha/2}.\ \sigma}{E})^2$, E is the minimum error of estimation

When σ is known the test-statistic Z= $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

When σ is unknown the test-statistic t= $\frac{\bar{X}-\mu}{s/\sqrt{n}}$

Test-statistic with a specific population proportion Z= $\frac{\hat{p}-p}{\sqrt{pq/n}}$

Confidence interval for σ^2 is $\frac{(n-1)s^2}{\chi^2_R} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_L}$

Test-statistic for testing a claim about σ or σ^2 is $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$

When population variances are known the test-statistic Z= $\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})}}$

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=$\frac{(\hat{p_1}-\hat{p_2})-(p_1-p_2)}{\sqrt{\bar{p}\bar{q}(\frac{1}{n_1}+\frac{1}{n_2})}}$

where $\bar{p} = \frac{X_1+X_2}{n_1+n_2}$,

$\hat{p_1} = \frac{X_1}{n_1},\ \hat{p_2} = \frac{X_2}{n_2},\ \bar{q} = 1-\bar{p}$

Confidence interval for the difference between two proportions is $(\hat{p_1} - \hat{p_2}) \pm z_{\alpha/2}\sqrt{(\frac{\hat{p_1}\hat{q_1}}{n_1} + \frac{\hat{p_2}\hat{q_2}}{n_2})}$

Correlation Co-efficient r=$\frac{n(\sum xy)-(\sum x)(\sum y)}{\sqrt{[n(\sum x^2)-(\sum x)^2]\ [n(\sum y^2)-(\sum y)^2]}}$

t=$r\sqrt{\frac{n-2}{1-r^2}}$ 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=$\frac{n(\sum xy)-(\sum x)(\sum y)}{n(\sum x^2)-(\sum x)^2}$