Statistics Formula – Sheet 1

Mean & Standard Deviation

$$\bar{x} = \frac{\sum f \cdot x}{\sum f}, \qquad \bar{x} = \frac{\sum w \cdot x}{\sum w}$$

$$s_x = \sqrt{\frac{\sum f.(x-\bar{x})^2}{n-1}} = \sqrt{\frac{n\sum f.x^2 - (\sum f.x)^2}{n(n-1)}}$$

$$\sigma = \sqrt{\frac{\sum f.(x-\mu)^2}{N}} = \sqrt{\frac{N\sum f.x^2 - (\sum f.x)^2}{N^2}}$$

Counting Rules

$$n! = n \cdot (n-1) \cdot (n-2) \ldots 3 \cdot 2.1$$

$$\begin{pmatrix} & n & \\ n_1 \quad n_2 & \ldots & n_k \end{pmatrix} = \frac{n!}{n_1!\, n_2! \ldots .. n_k!}$$

$$\binom{n}{r} = C_r^n = \frac{n!}{(n-r)!\, r!} \quad (Combinations)$$

$$P_r^n = \frac{n!}{(n-r)!} \quad (Permutations)$$

Probability

$$0 \le P(x) \le 1 \quad \& \quad \sum P(x) = 1$$

$$P(A \text{ and } B) = P(A)P(B|A)$$

If And B are independent.

$$P(A \text{ and } B) = P(A) \cdot P(B)$$

$$P(A^c) = 1 - P(A)$$

Probability Distribution

$$\mu = \sum x \cdot p(x)$$

$$\sigma = \sqrt{\sum x^2 \cdot P(x) - \mu^2}$$

Bayes' Formula

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}$$

Bayes' Formula – Two – Event Case

$$P(A|B) = \frac{P(A \,\&\, B)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

Chebyshev's Theorem

$$P(|x-\mu| < k\sigma) \ge 1 - \frac{1}{k^2} \quad \text{where } k > 0$$

or

$$P(|x-\mu| \ge k\sigma) \le \frac{1}{k^2}$$

Coefficient of Variation

$$CV = \frac{\sigma}{\mu} \cdot 100\%$$

Expected Value

$$\boldsymbol{E(x) = \sum x \cdot p(x)} \quad (Probability\ Distributions)$$

$$\boldsymbol{E(x) = \int x \cdot f(x) dx} \quad (Contineous\ \ Distributions)$$

Variance = Cov(X, X) = V(x)

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

Covariance

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

Central Limit Theorem

$$x \sim N(\mu, \sigma) \quad or \quad n \ge 30 \quad \Rightarrow \quad \bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Linear Correlation

$$r = \frac{\sum Z_x \cdot Z_y}{n-1} = \frac{SS_{xy}}{\sqrt{SS_{xx} \cdot SS_{yy}}} = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2}\sqrt{\sum (y-\bar{y})^2}}$$

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

Regression

$$\hat{y} = b_1 x + b_0$$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = r\frac{S_y}{S_x}$$

$$b_0 = \bar{y} - b_1 x$$

Test Statistics (One Population)

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} \quad (Mean - \sigma - unknown)$$

$$Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \quad (Mean - \sigma - known)$$

$$Z = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} \quad (Proportion)$$

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \quad (\sigma^2 \text{ or } \sigma)$$

Skewness

$$S_k = \frac{Mean - Mode}{SD} \quad (Karl\ Pearson)$$

If mode doesn't exist,

$$S_k = \frac{3(Mean - Median)}{SD}$$