Formula Sheet

(Ungrouped) | **(Grouped)**

→ **Mean:**

Ungrouped: $\bar{x} = \dfrac{\Sigma x}{n}$

Grouped: $\bar{x} = \dfrac{\Sigma fx}{\Sigma f}$

→ **Median:**

n: odd

$\left(\frac{n+1}{2}\right)^{th}$ value

n: even

$\frac{1}{2}\left[\left(\frac{n}{2}\right)^{th} \text{val} + \left(\frac{n}{2}+1\right)^{th} \text{val}\right]$

Grouped: $l + \frac{h}{f}\left(\frac{n}{2} - c\right)$

→ **Mode:-**

Most frequent occuring value

Grouped: $l + \left[\dfrac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)}\right] h$

→ **Quartiles**

n: odd

$Q_1 = \left(\frac{n+1}{4}\right)^{th}$ val

$Q_3 = \left(\frac{3n+1}{4}\right)^{th}$ val

n: even

$Q_1 = \frac{1}{2}\left[\left(\frac{n}{4}\right)^{th} \text{val} + \left(\frac{n}{4}+1\right)^{th} \text{val}\right]$

Grouped:

$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - c\right)$

$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - c\right)$

Scanned with CamScanner

$$Q_3 = \frac{1}{2}\left[\left(\frac{3n}{4}\right)^{th} val + \left(\frac{3n}{4}+1\right)^{th} val\right]$$

⇒ **Decile:-**

n: odd

$$D_j = \left(\frac{jn+1}{10}\right)^{th} val$$

$$D_j = L + \frac{h}{f}\left(\frac{jn}{10} - c\right)$$

$j = 1, 2, 3 \ldots 9$

n: even

$$D_j = \frac{1}{2}\left[\left(\frac{jn}{10}\right)^{th} val + \left(\frac{jn}{10}+1\right)^{th} val\right]$$

⇒ **Percentile:-**

n: odd

$$P_i = \left(\frac{in+1}{100}\right)^{th} val$$

$$P_i = L + \frac{h}{f}\left(\frac{in}{100} - c\right)$$

$i = 1, 2, \ldots, 99$

n: even

$$P_i = \frac{1}{2}\left[\left(\frac{in}{100}\right)^{th} val + \left(\frac{in}{100}+1\right)^{th} val\right]$$

(Population)

(Sample)

Variance:-

$$\sigma^2 = \frac{\Sigma f(x-\mu)^2}{\Sigma f}$$

$$\sigma^2 = \frac{\Sigma x^2}{N} - \left(\frac{\Sigma x}{N}\right)^2$$

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

$$s^2 = \frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2$$

$$s^2 = \frac{\Sigma f(x-\bar{x})^2}{\Sigma f}$$

Standard Deviation:-

$$\sigma = \sqrt{\frac{\Sigma x^2}{N} - \left(\frac{\Sigma x}{N}\right)^2}$$

$$s = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

$$s = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}}$$

Scanned with CamScanner

IQR

$Q_3 - Q_1$

→ Semi-IQR

$(Q_3 - Q_1)/2$

→ relative frequency

$$\frac{\text{frequency of a particular class}}{\text{total no of observation}}$$

→ Skewness

centre : Symmetric

right : positively skewed

left : negatively skewed

∴ Mean of positively skewed data is greater than median

Opposite is the case in negatively skewed data.

(Probability)

- Arrangements for n distinct objects

$n!$

- Arrangement for n non-distinct elements

$$\frac{n!}{n_1!\, n_2! \dots n_k!}$$

- Circular Arrangement

$(n-1)!$

- Permutation (order is important)

$^nP_r = n! / (n-r)!$

- Combination (order is not important)

$^nC_r = n! / r!\,(n-r)!$

Scanned with CamScanner

→ Additive law

- Mutually exclusive:

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

- Non - mutually exclusive

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

(Conditional Probability)

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

$$P(A \cup B) = 1 - P(A' \cap B')$$

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

$$P(A \cap D') = P(A) - P(A \cap D)$$

⇒ Multiplicative law

$$P(A \cap B) = P(A) \cdot P(B)$$ independent

$$P(A \cap B) = P(A) \cdot P(B/A)$$ dependent

⇒ Bayes Theoram

$$P(D) = P(A)\,P(D/A) + P(B)\,P(D/B) + \cdots$$

Total probability

$$P(A/D) = \frac{P(A)\,P(D/A)}{P(D)}$$

⇒ Bayesian Spam Filter

$$P(S/E_1) = \frac{P(S)\,P(E_1/S)}{P(S)\,P(E_1/S) + P(S')\,P(E_1/S')}$$

Scanned with CamScanner

Continuous Probability Distribution

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

upper limit

$$P(x > 2) = \int_{2}^{} f(x)\, dx$$

$$P(x < 1) = \int^{1} f(x)\, dx$$

lower limit

CDF:-

$$F(x) = \int_{-\infty}^{x} f(x)\, dx$$

lower limit

$$P(a < x < b) = F(b) - F(a)$$

$$P(x < b) = F(b)$$

$$P(x \geq b) = 1 - F(b)$$

Joint Probability Distribution

Marginal (Discrete)

$$g(x) = \sum_{y} f(x,y)$$

$$h(y) = \sum_{x} f(x,y)$$

(continuous)

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(xy)\, dx\, dy = 1$$

Y limits

x limits

(Marginal)

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\, dy \qquad h(y) = \int_{-\infty}^{\infty} f(x,y)\, dx$$

Y limits

x limit

Scanned with CamScanner

→ (Independent)

$f(x,y) = g(x)\, h(y)$ Independent

$f(0,1) = g(0)\, h(1)$ for all pairs of x & y

⇒ Mathematical Expectation / Mean

$\mu = E(x) = \sum x\, P(x)$ (Discrete)

$\mu = E(x) = \int_{-\infty}^{\infty} x\, P(x)$ (Continuous)

⇒ Variance

$Var(x) = E(x^2) - [E(x)]^2$ (Discrete)

$Var(x) = E(x^2) - [E(x)]^2$ (Continuous)

$\int_{-\infty}^{\infty} x^2\, p(x)$

⇒ Expectation of a function

$E(g(x)) = \sum g(x)\, P(x)$ (Discrete)

$E(g(x)) = \int_{-\infty}^{\infty} g(x)\, P(x)$ (Continuous)

⇒ Variance of a function

$Var(g(x)) = E(g(x)^2) - E(g(x))^2$ (Both)

Scanned with CamScanner

Binomial Distribution

- Trials are independent

$$P(X=x) = \binom{n}{x} p^x q^{n-x}$$

$\mu = np$

$\sigma^2 = npq$

$P(x=n) = \sum_0^n - \sum_0^{n-1}$

$P(x \le n) = \sum_0^n$

$P(n \le x \le 2) = \sum_0^2 - \sum_0^{n-1}$

$P(x \ge n) = 1 - \sum_0^{n-1}$

⇛ Multinomial Distribution

$$f(x_1, x_2 \dots x_k, n, p_1, p_2 \dots p_k) = \frac{n!}{x_1!\, x_2! \cdots x_k!} (P_1)^{x_1} (P_2)^{x_2} \dots (P_k)^{x_k}$$

$\mu = np_i \qquad i = 1, 2 \dots K$

$\sigma^2 = np_iq_i$

⇛ Hypergeometric Distribution

$$h(x, N, n, k) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

$\mu = \frac{nK}{N}$

$\sigma^2 = \frac{N-n}{N-1} \cdot n \cdot \frac{K}{N}\left(1 - \frac{K}{N}\right)$

(approximation)

$P = K/N$

$n/N < 0.05$ we hypergeometric

⇛ Poisson Distribution

$$P(x, \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Scanned with CamScanner

(approximation)

$$p \to 0$$
$$n \to \infty$$
$$p < 0.05$$

$$\lambda = np$$

⇒ Normal Distribution (Use table)

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

⇒ Point Estimation

Mean = $\frac{\Sigma x}{n}$

STD Error $\sigma / \sqrt{n}$

Std deviation = $\sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$

Var = $\frac{\sigma^2}{n}$

⇒ Hypothesis Testing

(Z-test)

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$z > z_\alpha$$
$$z < -z_\alpha$$

$z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$

(T-test)

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

$v = n - 1$

$$t > t_{\alpha v}$$
$$t < -t_{\alpha v}$$

$t < -t_{\alpha/2, v}$ or $t > t_{\alpha/2, v}$

(Confidence Interval)

Scanned with CamScanner

$$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

⇒ **Regression**

$$Y = \alpha + \beta x + \epsilon$$

$$\hat{Y} = a + bx$$

$$a = \bar{Y} - b\bar{x} \qquad b = \frac{n\sum xY - \sum x \sum Y}{n\sum x^2 - (\sum x)^2}$$

$$\bar{Y} = \frac{\sum Y}{n}$$

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

⇒ **Correlation**

range = $+1 \rightarrow -1$

$$r = \frac{n\sum xY - \sum x \sum Y}{\sqrt{[(n\sum x^2 - (\sum x)^2]\,.\,[(n\sum Y^2 - (\sum Y)^2]}}$$

⇒ **Coeffecient of determination**

$$r^2 = \frac{\text{Explained Variation}}{\text{total variation}}$$

OR

$$r^2 = 1 - \frac{\text{Unexplained}}{\text{total}}$$

Scanned with CamScanner

Total Variation $= \sum Y^2 - \frac{(\sum Y)^2}{n}$

Unexplained $= \sum Y^2 - a \sum Y - b \sum XY$

explained $=$ Total - unexplained.

Scanned with CamScanner