

Basic Statistics Formulas

Population Measures

$$\text{Mean } \mu = \frac{1}{n} \sum x_i \quad (1)$$

$$\text{Variance } \sigma^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 \quad (2)$$

$$\text{Standard Deviation } \sigma = \sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2} \quad (3)$$

Sampling

$$\text{Sample mean } \bar{x} = \frac{1}{n} \sum x_i \quad (4)$$

$$\text{Sample variance } s_x^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 \quad (5)$$

$$\text{Std. Deviation } s_x = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} \quad (6)$$

$$\text{z-score } z = \frac{x - \mu}{\sigma} \quad (7)$$

Correlation $r =$

$$\frac{1}{n-1} \sum_{i=1}^n \left(\frac{(x_i - \bar{x})}{s_x} \right) \left(\frac{(y_i - \bar{y})}{s_y} \right) \quad (8)$$

Linear Regression

$$\text{Line } \hat{y} = a + bx \quad (9)$$

$$b = r \frac{s_y}{s_x}, a = \bar{y} - b\bar{x} \quad (10)$$

$$s = \sqrt{\frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y})^2} \quad (11)$$

$$SE_b = \frac{s}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (12)$$

$$\text{To test } H_0 : b = 0, \text{ use } t = \frac{b}{SE_b} \quad (13)$$

$$CI = b \pm t^* SE_b \quad (14)$$

Probability

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) \quad (15)$$

$$P(\text{not } A) = 1 - P(A) \quad (16)$$

$$P(A \text{ and } B) = P(A)P(B) \text{ (independent)} \quad (17)$$

$$P(B|A) = P(A \text{ and } B)/P(A) \quad (18)$$

$$0! = 1; n! = 1 \times 2 \times 3 \cdots \times (n-1) \times n \quad (19)$$

$$\binom{n}{k} = \frac{n!}{n!(n-k)!} \quad (20)$$

Binomial Distribution :

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (21)$$

$$\mu = np, \sigma = \sqrt{np(1-p)} \quad (22)$$

One-Sample z-statistic

$$\text{To test } H_0 : \mu = \mu_0 \text{ use } z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \quad (23)$$

$$\text{Confidence Interval for } \mu = \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} \quad (24)$$

$$\text{Margin of Error } ME = z^* \frac{\sigma}{\sqrt{n}} \quad (25)$$

$$\text{Minimum sample size } n \geq \left[\frac{z^* \sigma}{ME} \right]^2 \quad (26)$$

One-Sample t-statistic

$$SEM = \frac{s_x}{\sqrt{n}}, t = \frac{\bar{x} - \mu}{s_x/\sqrt{n}} \quad (27)$$

$$\text{Confidence Interval} = \bar{x} \pm t^* \frac{s_x}{\sqrt{n}} \quad (28)$$

Two-Sample t-statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (29)$$

$$\text{Conf. Interval} = (\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (30)$$

Sample Proportions

$$\mu_{\hat{p}} = p, \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \quad (31)$$

$$\text{Conf. Int.} = \hat{p} \pm z^*(SE) \quad (32)$$

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad (33)$$

$$\text{sample size } n > \left[\frac{z^*}{ME} \right]^2 p^* (1-p^*) \quad (34)$$

$$\text{To test } H_0 : p = p_0, \text{ use } z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \quad (35)$$

Two-Sample Proportions

$$SE = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \quad (36)$$

$$CI = (\hat{p}_1 - \hat{p}_2) \pm z^*(SE) \quad (37)$$

$$\text{To test } H_0 : p_1 = p_2, \text{ use } \quad (38)$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (39)$$

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}, X_i = \text{successes} \quad (40)$$

Chi-Square Statistic

$$\chi^2 = \sum_{i=1}^n \frac{(o_i - e_i)^2}{e_i} \quad (41)$$

o_i = observed, e_i = expected

Central Limit Theorem

$$s_{\bar{x}} \rightarrow \frac{\sigma}{\sqrt{n}} \text{ as } n \rightarrow \infty \quad (42)$$

