Population Mean: $\mu = \frac{\sum_{i=1}^{N} x_i}{N}$

Sample Mean: $\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$

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

Sample Variance: $s^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{X})^2}{n-1} = \frac{1}{n-1}[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}]$

Population s.d.: $\sigma = \sqrt{\sigma^2}$

Sample s.d.: $s = \sqrt{s^2}$

Population coefficient of variation: $CV = \frac{\sigma}{\mu}$

Sample coefficient of variation: $cv = \frac{s}{\overline{X}}$

Population covariance: $\sigma_{XY} = \frac{\sum_{i=1}^{N} (x_i - \mu_X)(y_i - \mu_Y)}{N}$

Sample covariance: $s_{XY} = \frac{\sum_{i=1}^{n} (x_i - \overline{X})(y_i - \overline{Y})}{n-1}$

Population coefficient of correlation: $\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$

Sample covariance (shortcut): $s_{XY} = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}\right]$

Sample coefficient of correlation: $r = \frac{s_{XY}}{s_X s_Y}$

Conditional Probability:

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

Complement Rules:

$$P(A^C) = 1 - P(A) \qquad P(A^C \mid B) = 1 - P(A \mid B)$$

Multiplication Rule: $P(A \text{ and } B) = P(A \mid B)P(B)$

Addition Rule:

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

Expected Value: $E[X] = \mu = \sum_{\text{all x}} xp(x)$

Variance: $V[X] = E[(X-\mu)^2] = \sigma^2 = \sum_{\text{all } x} (x-\mu)^2 p(x)$

Covariance: $E[(X-\mu_X)(Y-\mu_Y)] = \sigma_{XY} = \sum_{\text{all } x} \sum_{\text{all } y} (x-\mu_X)(y-\mu_Y)\, p(x,y)$

Laws of Expected Value:

E[c] = c
E[X+c] = E[X] + c
E[cX] = cE[X]
E[a+bX+cY] = a + bE[X] + cE[Y]

Laws of Variance:

V[c] = 0
V[X+c] = V[X]
V[cX] = c^2V[X]
V[a+bX+cY] = b^2V[X] + c^2V[Y] + 2bcCOV[X,Y]

Laws of Covariance:

COV[X, c] = 0
COV[a+bX, c+dY] = bdCOV[X,Y]

Combinatorial formula: $C_x^n = \dfrac{n!}{x!(n-x)!}$

Binomial probability: $p(x) = \dfrac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$

$for\ x = 0, 1, 2, \ldots, \text{n}$

Mean of Binomial: np **Variance of Binomial:** $np(1-p)$

Uniform density function: $f(x) = \dfrac{1}{b-a} \quad where \ \ a \le x \le b$

Mean of Uniform: $\dfrac{a+b}{2}$

Variance of Uniform: $\dfrac{(b-a)^2}{12}$

<u>Inference about population mean and population proportion:</u>

<u>Sample Mean:</u>

$$\mu_{\overline{X}} = E[\overline{X}] = \mu_X$$

$$\sigma_{\overline{X}}^2 = V[\overline{X}] = \frac{\sigma_X^2}{n}$$

$$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}$$

<u>Sample Proportion:</u>

$$E[\hat{P}] = p$$

$$\sigma_{\hat{P}}^2 = V[\hat{P}] = \frac{p(1-p)}{n}$$

$$\sigma_{\hat{P}} = \sqrt{\frac{p(1-p)}{n}}$$

z statistic: $z = \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}}$ **Confidence interval estimator of μ when σ^2 is known:** $\overline{X} \pm z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$

Sample size to estimate $\mu \pm \tau$ when σ^2 is known: $n = \left(\dfrac{z_{\alpha/2}\sigma}{\tau} \right)^2$

t statistic: $t = \dfrac{\overline{X} - \mu}{s / \sqrt{n}}$ **Confidence interval estimator of μ when σ^2 is unknown:** $\overline{X} \pm t_{\alpha/2} \dfrac{s}{\sqrt{n}}$

<u>Inference about *p* (population proportion):</u>

Test statistic: $z = \dfrac{\hat{p} - p}{\sqrt{p(1-p)/n}}$ **CI estimator:** $\hat{p} \pm z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}$

Sample size to estimate $p \pm \tau$: $n = \left(\dfrac{z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})}}{\tau} \right)^2$

<u>SIMPLE REGRESSION:</u>

Least squares line / linear regression line: $\hat{y} = b_0 + b_1 x \quad b_1 = \dfrac{\text{cov}(x, y)}{s_x^2} = r \dfrac{s_y}{s_x} \quad b_0 = \overline{y} - b_1 \overline{x} \quad \hat{z}_Y = r z_X$

Coefficient of Determination: $R^2 = \frac{[\text{cov}(x,y)]^2}{s_x^2 s_y^2} = (r)^2$

Standard error of estimate:

$$s_e = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}} \sqrt{\frac{SSE}{n-2}} \qquad s_e^{\,2} = \frac{SSE}{n-2}$$

Standard error of least squares slope estimate:

$$s_{b_1} = \frac{s_e}{\sqrt{(n-1)s_x^2}} = se(b_1)$$

Test statistic & confidence interval estimator for β_1:

$$t_{(n-2)} = \frac{b_1 - \beta_1}{se(b_1)} \qquad b_1 \pm t_{\alpha/2} se(b_1) \qquad \nu = n-2$$

Prediction Interval for y for a given value of x (x_g):

$$\hat{y} \pm t_{\alpha/2} s_e \sqrt{1 + \frac{1}{n} + \frac{(x_g - \overline{X})^2}{(n-1)s_x^2}} \qquad \nu = n-2$$

Confidence Interval for expected value of y given x (x_g):

$$\hat{y} \pm t_{\alpha/2} s_e \sqrt{\frac{1}{n} + \frac{(x_g - \overline{X})^2}{(n-1)s_x^2}} \qquad \nu = n-2$$

<u>SIMPLE & MULTIPLE REGRESSION:</u>

$$SST = \sum_{i=1}^{n} (y_i - \overline{y})^2 \qquad SSR = \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2 \qquad SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \qquad SST = SSR + SSE$$

Coefficient of Determination: $R^2 = \frac{SSR}{SST}$ $\quad R^2 = 1 - \frac{SSE}{SST}$ $\quad R^2 = 1 - \frac{SSE}{\sum_{i=1}^{n} (y_i - \overline{y})^2}$

<u>MULTIPLE REGRESSION:</u> $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \varepsilon_i$

Standard error of estimate:

$$s_e = \sqrt{\frac{SSE}{n-k-1}} \qquad s_e^{\,2} = \frac{\sum_{i=1}^{n} (e_i - 0)^2}{n-k-1} = \frac{SSE}{n-k-1} \qquad Adj.\,R^2 = 1 - \frac{SSE/(n-k-1)}{\sum (y_i - \overline{y})^2 / (n-1)}$$

Test statistic & confidence interval estimator for β_j:

$$t = \frac{b_j - \beta_j}{s_{b_j}} \qquad b_j \pm t_{\alpha/2} s_{b_j} \qquad v = n-k-1$$

Test of overall statistical significance of linear regression model:

$$F = \frac{R^2 / k}{(1-R^2)/(n-k-1)} \qquad F = \frac{(SST - SSE)/k}{SSE/(n-k-1)} \qquad F = \frac{SSR/k}{SSE/(n-k-1)}$$

Numerator degrees of freedom: k

Denominator degrees of freedom: $n-k-1$