Wilcoxon Rank-Sum Test for Large Sample

$$\boxed{Z_{STAT} = \frac{T_1 - \mu_{T_1}}{\sigma_{T_1}} = \frac{T_1 - \frac{n_1(n+1)}{2}}{\sqrt{\frac{n_1 n_2 (n+1)}{12}}}}$$

The Kruskal-Wallis H-test statistic:

$$\boxed{H = \left[\frac{12}{n(n+1)} \sum_{j=1}^{c} \frac{T_j^2}{n_j}\right] - 3(n+1)}$$

Simple Linear Regression Model

$$\boxed{Y_i = \beta_0 + \beta_1 X_i + \varepsilon} \quad \boxed{\hat{Y}_i = b_0 + b_1 X}$$

Total variation is made up of two parts: SST = SSR + SSE

Coefficient of Determination

$$\boxed{r^2 = \frac{SSR}{SST} = \frac{\text{regression } \textit{sum} \text{ of squares}}{\textit{total} \text{ sum of squares}}}$$

Inference about the slope: t test with df = n -2

$$t_{STAT} = \frac{b_1 - \beta_1}{S_{b_1}}$$

F test for overall significance with df1 = k, df2 = n – k -1

$$\boxed{F_{STAT} = \frac{MSR}{MSE}}, \quad \boxed{\begin{array}{l} MSR = \dfrac{SSR}{k} \\ MSE = \dfrac{SSE}{n-k-1} \end{array}}$$

Confidence interval estimate about the slope

$$\boxed{b_1 \pm t_{\alpha/2} S_{b_1}}$$

Multiple Regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i, \quad \hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \cdots + b_k X_{ki}$$

Adjusted R square

$$\boxed{r_{adj}^2 = 1 - \left[(1 - r^2) \left(\frac{n-1}{n-k-1} \right) \right]}$$

Variance Inflation Factor

$$VIF_j = \frac{1}{1 - R_j^2}$$

Nonlinear Regression

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + \varepsilon_i$$

Logistic Regression

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1, \pi = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

Regression ANOVA table

Source	DF	Sum of Sq	Mean Sq	F Value	Pr(>F)
Regression	k	SSR	MSR	F	p-value
Residual	n-k-1	SSE	MSE		
Total	n-1	SST			