Z-test (σ is known): $z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$; SE: $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$; z ~N (0, 1); $(1 - \alpha)\ CI = \bar{X} \pm z_{\frac{\alpha}{2}} S_{\bar{X}}$

One-sample t-test (σ is unknown): $t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$; $S_{\bar{X}} = \frac{S}{\sqrt{n}}$; t ~$t_{n-1,\frac{\alpha}{2}}$; $(1 - \alpha)\ CI = \bar{X} \pm t_{n-1,\frac{\alpha}{2}} S_{\bar{X}}$

Independent Samples t Test: $t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}}$; $S_{\bar{X}_1 - \bar{X}_2} = S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$; $S_p = \sqrt{\frac{{S_1}^2(n_1-1)+{S_2}^2(n_2-1)}{n_1+n_2-2}}$; $df = (n_1 - 1) + (n_2 - 1)$; $(1-\alpha)\ CI = (\bar{X}_1 - \bar{X}_2) \pm t_{\frac{\alpha}{2}, n_1+n_2-2}\, S_{\bar{X}_1 - \bar{X}_2}$

Paired Samples t Test: $t = \frac{\bar{d}}{S_{\bar{d}}}$, $S_{\bar{d}} = \frac{S_d}{\sqrt{n}}$; $\{d = post - pre\ , n = \#\ of\ pairs\}$; $(1 - \alpha)\ CI = \bar{d} \pm t_{n-1,\frac{\alpha}{2}} SE_{\bar{d}}$

Effect size: $\delta = \frac{|\mu_1 - \mu_2|}{\sigma}$; 1-Sample t test: $d = \frac{|\bar{X} - \mu_0|}{S}$; Indep Smp t test: $d = \frac{|\bar{x}_1 - \bar{x}_2|}{S_p}$; Dep Smp t test: $d = \frac{\bar{d}}{S_d}$; ANOVA: $\eta^2 = \frac{SS_B}{SS_T} \rightarrow f = \sqrt{\frac{\eta^2}{1-\eta^2}}$; Correlation (coeff of determination): $r^2{}_{XY} = \frac{{s_{XY}}^2}{{s_X}^2{s_Y}^2}$; Regression: $R^2 = \frac{SSreg}{(SStotal)}$; S: 0.2, M: 0.5, L: 0.8, for ANOVA (f): S: 0.1, M: 0.25, L: 0.4

Power: 1-β $(type\ II\ error)$= p (z= $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ > $z_{\frac{\alpha}{2}}$), assuming H_1 is true. Type I error: $1 - (1 - \alpha)^C$; C: # independent tests.

One-sample test of variance: Chi^2 test: $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$; 2-tailed critical value:$p(< .025), p(> .075)$

${\chi_{df}}^2 = (df_{num})\, F_{df_{num},\infty}$; $t_{df} = \sqrt{F_{1,df_{den}}}$; $z = \sqrt{{\chi_1}^2} = \sqrt{F_{1,\infty}} = t_{\infty\ df}$

2-sample tests of variance: $F = \frac{{S_1}^2}{{S_2}^2}$; $df_1 = n_1 - 1$, $df_2 = n_2 - 1$

Using variances to answer questions about means (multiple groups): ANOVA: $= \frac{MS_B}{MS_w}$, $df_B = J - 1$, $df_w = N - J$, $df_T = N - 1$, $MS = \frac{SS}{df}$; $[SS_w = \sum_{j=1}^{J}\sum_{i=1}^{n_j}(Y_{ij} - \bar{Y}_j)^2$, $SS_B = \sum_{j=1}^{J} n_j(\bar{Y}_j - \bar{Y}_{..})^2$, $SS_T = \sum_{j=1}^{J}\sum_{i=1}^{n_j}(Y_{ij} - \bar{Y}_{..})^2$] ; $SS_T = SS_B + SS_W$

Covariance: $\sigma_{XY} = \frac{\sum(X_i - \mu_X)(Y_i - \mu_Y)}{N}$; $S_{XY} = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{n-1}$

Pearson product moment correlation coefficient: $r_{XY} = \hat{\rho}_{XY} = corr(X, Y) = \frac{s_{XY}}{s_X s_Y} = \frac{\sum(Z_X Z_Y)}{n-1}$; S: M: L=0.1: 0.3: 0.5 ; $z = \frac{x - \bar{x}}{s}$; Estimated standard error of the correlation: $S_r = \sqrt{\frac{1-r^2}{n-2}}$

Correlation: test statistic for 1 sample: $t = \frac{r}{S_r} = r\sqrt{\frac{n-2}{1-r^2}}$; $\nu = n - 2$; $(1 - \alpha)CI = r\ t_{n-1,\frac{\alpha}{2}} S_r$ or: $z = \frac{z_r - z_\rho}{\sigma_{z_r}}$; $\sigma_{z_r} = \frac{1}{\sqrt{n-3}}$; for 2 independent samples: $Z_r = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}} = \frac{(n_1-3)z_1 + (n_2-3)z_2}{n_1+n_2-6}$

Coefficient of determination: The proportion of variance that X and Y share. The amount of variance in Y that is explainable by X (or vice versa): $r^2{}_{xy} = \frac{S^2{}_{xy}}{S^2{}_x S^2{}_y}$

Regression: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_i$; unstandard reg coeff: $\hat{\beta}_1 = \frac{cov(X,Y)}{var(X)} = \frac{S_{XY}}{{S_X}^2} = r\left(\frac{S_y}{S_x}\right)$; $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$

Regression with standardized coefficients: $Z_{\hat{Y}} = \hat{\beta}_1{}^* Z_{X1} + \hat{\beta}_2{}^* Z_{X2}$; $\hat{\beta}^* = \hat{\beta} \times \frac{s_x}{s_y} = r_{xy}$

Regression variability: $\sum(Y_i - \bar{Y})^2 = \sum(\hat{Y} - \bar{Y})^2 + \sum(Y_i - \hat{Y})^2$=> $SS_t = SS_{reg} + SS_{residual}$

Coefficient of determination (explained variation): $R^2 = \frac{SS_{regression}}{SS_{total}} = 1 - \frac{SS_{error}}{SS_{total}}$

Coeff of multiple determination: $R^2_{Y.12} = \frac{SS_{regression}}{SS_{total}} = \sqrt{\frac{r^2_{Y1} + r^2_{Y2} - 2r_{Y1}r_{Y2}r_{12}}{1 - r^2_{Y12}}}$

Adjusted R^2(variation explained by only IVs that affect the DV): $R^2_{adj} = 1 - (\frac{(1-R^2)(n-1)}{n-k-1})$

$F = \frac{MS_{reg}}{MS_{res}} = \frac{SS_{reg}/df_{reg}}{SS_{res}/df_{res}} = \frac{\frac{SS_{reg}}{k}}{\frac{SS_{res}}{n-k-1}} = \frac{\frac{R^2}{k}}{\frac{1-R^2}{n-k-1}}$; k: # predictors, n: # subjects

Standard error of estimate (a measure of accuracy of prediction by the model: SD of residuals) :

$S^2{}_{Y|X} = \frac{SS_{res}}{n-k-1} = \frac{\sum(Y-\hat{Y})^2}{n-k-1} = MS_{res}$; $SEE = S_{Y|X} = S_Y\sqrt{\frac{k}{n-k-1}(1-r^2)} = \sqrt{\frac{(1-r^2)S_Y}{n-k-1}} = \sqrt{MS_{res}}$

Standard error of slope (variation in slope due to sampling error): $SE_{\hat{\beta}_j} = \frac{SEE}{\sqrt{\sum(X-\bar{X})^2}} =$

$\frac{S_Y}{S_{X_j}}\sqrt{\frac{1-R^2}{(n-k-1)(1-R_j^2)}}$ (used in t-test for hypothesis testing of slope: $t = \frac{\hat{\beta}_j - \hat{\beta}_{j(H_0)}}{SE_{\hat{\beta}_j}} = \frac{\hat{\beta}_j - 0}{SE_{\hat{\beta}_j}}$)

Standard error of intercept: $SE_{\hat{\beta}_0} = SEE\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{(n-1)S_X^2}}$ (used in t-test of intercept: $t = \frac{\hat{\beta}_0 - 0}{SE_{\hat{\beta}_0}}$)

Regression Assumptions: 1.* Independence (tough to test. assumed!) => violation: impacts SE of model. 2.* Linearity: Y vs X (simple regression), residuals vs each X, residuals vs $\hat{Y}$ => violation: Bias intercept & slope, change in Y not constant and depends on X => Polynomial Regression, testing of Higher Order terms (X^2). Log, Inverse, or Box-Cox transform: $Y' = Y^\lambda$ and Box-Tidwell transform: $X' = X^\lambda$. 3.* Homoscedasticity: Y vs X (simple), residuals vs each X, residuals vs $\hat{Y}$, Levene's test (not reliable for large n) => violation: Bias in SEE, inflate SE & type II error, non-normal conditional distributions => Weighted Least Square estimation. Sqrt<Log<Inverse. 4.* Normality: histogram of residuals, PP/QQ plot of residuals, skewness>1.96 $\times$ SE_{skewness}, KS test & Sharpiro-Wilk test (not reliable for n>50) => violation: Less precise slope, intercept and R^2 => Log transform for pos skewness, Square Root for pos/neg skewness. Sqrt<Log<Inverse.

Partial Correlation: $r_{YZ.X} = \frac{r_{YZ} - r_{YX}r_{ZX}}{\sqrt{1-r^2{}_{YX}}\sqrt{1-r^2{}_{ZX}}}$, $t = \frac{r_{YZ.X} - \rho}{S_{r_{YZ.X}}}$, $S_{r_{YZ.X}} = \sqrt{\frac{r^2{}_{YZ.X}}{n-3}}$, df= n-3

Semi-partial correlation (residualized correlation): $r_{Y\,(1.2)} = \frac{r_{Y1} - r_{Y2}r_{12}}{\sqrt{1-r^2{}_{12}}}$

Partial F-test ($\Delta R^2\ test$): $F = \frac{\frac{\Delta R^2}{\Delta df}}{\frac{(1-R^2_{full})}{df_{error(full)}}}$; $\Delta R^2 = R^2{}_{full} - R^2{}_{reduced}$; $\Delta df = df_{reg(full)} - df_{reg(reduced)} = k2 - k1$; $df_{error(full)} = n - k2 - 1$

CI for R^2 , if n>60 : $SE_{R^2_{model1} - R^2_{model2}} = \sqrt{SE^2_{R^2_{model1}} + SE^2_{R^2_{model2}}}$; $SE^2_{R^2} = \frac{kR^2(1-R^2)^2(n-k-1)^2}{(n^2-1)(n+k-1)}$

Diagnosing outliers in residuals: look for cases with values in excess of ±2 or ±3.

Leverage: A measure of each case's "pull" on the regression line: $h_i = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum(X-\bar{X})^2}$; Centered leverage: $h_i = \frac{(X_i-\bar{X})^2}{\sum(X-\bar{X})^2}$; Look for leverage > 2(k+1)/n or centered leverage> 2k/n.

Influence: Look for Standardized DFBeta > $\frac{3}{\sqrt{n}}$ or >1, $DFFit \neq 0$, Cook's distance d > 1.

Multicollinearity (predictors are correlated): highly significant $R^2$and non-significant reg coeffs, large SE of slopes. Tolerance $= 1 - \mathrm{R}_k{}^2$; $\mathrm{R}_k{}^2$ is the coeff of determination for the reg of the kth predictor on all other predictors. VIF $= \frac{1}{1-\mathrm{R}_k{}^2}$: how "inflated" the variance of the reg coeff is compared to what it'd be if the variable was uncorrelated with any other IV. Tolerance < 0.1 or .2775 (for R = .85) OR VIF>10 or 3.604 => multicollinearity. => Centering predictors (using $X - \bar{X}$ instead of X), dropping problematic predictors, combining correlated predictors, ridge regression.