Key Formulas

Rate	$\text{Rate} = \frac{\text{Number in subset}}{\text{Total number}} \times \text{Constant (e.g., 1,000)}$	Confidence interval around a sample mean with large samples	$\bar{X} \pm z_\alpha(\sigma_{\bar{X}}) = z_\alpha\left(\frac{s}{\sqrt{n-1}}\right)$
Proportion	$\text{Proportion} = \frac{\text{Number in subset of sample}}{\text{Total number in sample}} = \frac{f}{n}$	Confidence interval around a sample mean with small samples	$\bar{X} \pm t_\alpha(\sigma_{\bar{X}}) = \bar{X} \pm t_\alpha\left(\frac{s_x}{\sqrt{n}}\right)$
Percent	$\text{Percent} = \frac{f}{n} \times 100 = \text{Proportion} \times 100$	Confidence interval around a sample proportion with large samples	$\hat{p} \pm z_a(s_p) = z_a\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
Sample Mean	$\text{Sample mean} = \bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$	To find a z score for hypothesis testing one sample mean	$z_{\text{obt}} = \frac{\bar{X}-\mu}{s/\sqrt{n}}$
Variation ratio	$\text{VR} = 1 - \frac{f_{\text{modal}}}{n}$	To find a t score for hypothesis testing one sample mean:	$t_{\text{obt}} = \frac{\bar{X}-\mu}{s/\sqrt{n}}$
Range	Highest x_i score – Lowest x_i score	To find a z score for hypothesis testing for one sample proportion:	$z = \frac{\hat{p}-P}{\sigma_{\hat{p}}}$
Interquartile range	$\text{IQR} = xQ_3 - xQ_1$	Computational formula for chi-square statistic (equation 9–3):	$\chi^2 = \sum_{i=1}^{k}\left(\frac{f_o^2}{f_e}\right) - n$
Variance of a sample	$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{X})^2}{n-1}$	Phi coefficient (equation 9–4):	$\phi = \sqrt{\frac{\chi^2_{\text{obt}}}{n}}$
Standard deviation of a sample	$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{X})^2}{n-1}}$	Contingency coefficient:	$C = \sqrt{\frac{\chi^2_{obt}}{n+\chi^2_{obt}}}$
Computational formula for sample variance with ungrouped data – Take the Square Root for Standard Deviation	$s^2 = \frac{\Sigma(x_i^2) - \frac{(\Sigma x_i)^2}{n}}{n-1}$	Cramer's V:	$V = \sqrt{\frac{x^2_{obt}}{n(k-1)}}$

(Continued)

(Continued)

The probability of success—the binomial coefficient	$P(r) = \binom{n}{r} p^r q^{n-r}$ $P(r) = \left(\frac{n!}{r!(n-r)!}\right) p^r q^{n-r}$	Computational formula for Lambda:	$\lambda = \frac{(\Sigma f_i) - f_d}{n - f_d}$
The mathematical formula for the normal distribution	$f(x) = \frac{1}{s\sqrt{2p}} e^{\frac{-(x-m)^2}{2s^2}}$	Pooled variance difference between means *t* test	$t_{\text{obt}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} \sqrt{\frac{n_1+n_2}{n_1 n_2}}}$
Formula for converting a raw score into a *z* score	$z = \frac{x - \bar{X}}{s}$	Separate variance difference between means *t* test :	$t_{\text{obt}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1-1} + \frac{s_2^2}{n_2-1}}}$

Dependent or Matched-samples difference between means *t* test:	$t_{\text{obt}} = \frac{\bar{X}_D}{s_D / \sqrt{n}}$	Other formulas necessary for Analysis of Variance and Related Statistics:	$SS_{\text{between}} = \sum_i \sum_k (\bar{X}_k - \bar{X}_{\text{grand}})^2$ Total variance: $\frac{SS_{\text{total}}}{df_{\text{total}}} = \frac{\sum_i \sum_k (x_{ik} - \bar{X}_{\text{grand}})^2}{n-1}$ Within-group variance : $\frac{SS_{\text{within}}}{df_{\text{within}}} = \frac{\sum_i \sum_k (x_{ik} - \bar{X}_k)^2}{n-k}$ Between-groups variance : $\frac{SS_{\text{between}}}{df_{\text{between}}} = \frac{\sum_i \sum_k (x_k - \bar{X}_{\text{grand}})^2}{k-1}$ $F : \frac{SS_{\text{between}}/df_{\text{between}}}{SS_{\text{within}}/df_{\text{within}}} = \frac{\text{Variance between groups}}{\text{Variance within group}}$ Tukey's Honest Significant Difference Test: Critical difference score: $\text{CD} = q\sqrt{\frac{\text{Within-group variance}}{n_k}}$ Eta squared or the correlation ratio: $\eta^2 = \frac{SS_{\text{between}}}{SS_{\text{total}}}$
Difference between proportions *z* test:	$z_{\text{obt}} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}} \sqrt{\frac{n_1+n_2}{n_1 n_2}}}$		
Computational formula for Pearson's correlation coefficient:	$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{\left[n\Sigma x^2 - (\Sigma x)^2\right]\left[n\Sigma y^2 - (\Sigma y)^2\right]}}$		
Ordinary least-squares regression line for the population:	$y = \alpha + \beta x$		
Computational formula for the slope coefficient:	$b = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2}$		
t statistic for testing null hypothesis about *b* and *r*:	$t = r\sqrt{\frac{n-2}{1-r^2}}$	Beta weights:	$b^*_{x_1} = b_{x_1}\left(\frac{s_{x_1}}{s_y}\right)$ $b^*_{x_2} = b_{x_2}\left(\frac{s_{x_2}}{s_y}\right)$

Multivariate ordinary least-squares (OLS) regression equation:	$y = a + b_1 x_1 + b_2 x_2 + \cdots b_k x_k + \varepsilon$	Partial correlation coefficients:	$r_{yx_1 . x_2} = \dfrac{r_{yx_1} - \left(r_{yx_2}\right)\left(r_{x_1 x_2}\right)}{\sqrt{1 - r^2_{yx_2}}\sqrt{1 - r^2_{x_1 x_2}}}$ $r_{yx_2 . x_1} = \dfrac{r_{yx_1 - \left(r_{yx_1}\right)\left(r_{x_1 x_2}\right)}}{\sqrt{1 - r^2_{yx_1}}\sqrt{1 - r^2_{x_1 x_2}}}$
Partial slope coefficients:	$b_1 = \left(\dfrac{s_y}{s_{x_1}}\right)\left(\dfrac{r_{yx_1} - \left(r_{yx_2}\right)\left(r_{x_1 x_2}\right)}{1 - r^2_{x_1 x_2}}\right)$ $b_2 = \left(\dfrac{s_y}{s_{x_2}}\right)\left(\dfrac{r_{yx_2} - \left(r_{yx_1}\right)\left(r_{x_1 x_2}\right)}{1 - r^2_{x_1 x_2}}\right)$	Multiple coefficient of determination, R^2:	$R^2 = r^2_{yx_1} + (r^2_{yx_2 . x_1})(1 - r^2_{yx_1})$
Total sum of squares (SS_{total}):	$SS_{total} = \sum_i \sum_k (x_{ik} - \bar{X}_{grand})^2$	Logistic regression model:	$\ln\left(\dfrac{P}{1-P}\right) = \beta_0 + \beta_1 x_1$
Within-group sum of squares (SS_{within}):	$SS_{within} = \sum_i \sum_k (x_{ik} - \bar{X}_k)^2$	Predicted probabilities from logit model:	$\hat{p} = \dfrac{e^{(b_0 + b_1 x_1)}}{1 + e^{(b_0 + b_1 x_1)}}$