APPENDIX C
Key Formulas

		Learn to use this formula on:

ANOVA (One-way Analysis of Variance)

Total sum of squares (version 1)	$SS_{total} = SS_{between\,group} + SS_{within\,group}$	p. 234
Total sum of squares (version 2)	$SS_{total} = \sum(x_i - \bar{x}_{total})^2$	p. 238
Between group sum of squares	$SS_{between\,group} = \sum n_{group}(\bar{x}_{group} - \bar{x}_{total})^2$	p. 237
Between group degrees of freedom	$df_{between\,group} = (number\ of\ groups - 1)$	p. 239
Within group sum of squares	$SS_{within\,group} = \sum(x_i - \bar{x}_{group})^2$	p. 235
Within group degrees of freedom	$df_{within\,group} = (n_{group1} - 1) + (n_{group2} - 1) + (n_{group3} - 1) + \cdots$	p. 239
F-statistic (or F-ratio)	$F = \dfrac{SS_{between\,group} / df_{between\,group}}{SS_{within\,group} / df_{within\,group}}$	p. 240

Chi-square test of independence

Chi-square statistic	$\chi^2 = \sum \dfrac{(f_o - f_e)^2}{f_e}$	p. 280
Degrees of freedom of the chi-square statistic	$df_{\chi^2} = (rows - 1)(columns - 1)$	p. 287
Expected frequencies	$f_e = \dfrac{(row\ total)(column\ total)}{overall\ total}$	p. 280

Cohen's d

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s}$$ p. 201

Correlation coefficient, Pearson's

Pearson correlation coefficient (version 1)	$r = \dfrac{SP}{\sqrt{SS_x}\sqrt{SS_y}}$	p. 320

Pearson correlation coefficient (version 2)	$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}}$$	p. 320
Sum of products	$$SP = \sum(x_i - \bar{x})(y_i - \bar{y})$$	p. 318
Sum of squares (*x*)	$$SS_x = \sum(x_i - \bar{x})^2$$	p. 319
Sum of squares (*y*)	$$SS_y = \sum(y_i - \bar{y})^2$$	
T-statistic for Pearson's correlation coefficient	$$t = r\sqrt{\frac{n-2}{1-r^2}}$$	p. 323
Degrees of freedom of the t-statistic for Pearson's correlation coefficient	$$df_t = n - 2$$	p. 327

Correlation coefficient, Spearman's rank-order

Spearman's rank-order correlation coefficient (version 1)	$$\rho = \frac{SP}{\sqrt{SS_x}\sqrt{SS_y}}$$ *Note: Ranks are used in place of values and means in the calculation of the sum of products and sum of squares.*	p. 330
Spearman's rank-order correlation coefficient (version 2)	$$\rho = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}}$$ *Note: Ranks are used in place of values and means.*	p. 330
Sum of products	$$SP = \sum(x_i - \bar{x})(y_i - \bar{y})$$ *Note: Ranks are used in place of values and means.*	p. 318
Sum of squares (*x*)	$$SS_x = \sum(x_i - \bar{x})^2$$ *Note: Ranks are used in place of values and means.*	p. 319
Sum of squares (*y*)	$$SS_y = \sum(y_i - \bar{y})^2$$ *Note: Ranks are used in place of values and means.*	p. 319
T-statistic for Spearman's rank-order correlation coefficient	$$t = \rho\sqrt{\frac{n-2}{1-\rho^2}}$$	p. 332
Degrees of freedom of the t-statistic for Spearman's rank-order correlation coefficient	$$df_t = n - 2$$	p. 335

Covariance

$$cov(x,y)=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{n-1}$$ p. 318

Cramér's V

$$V=\sqrt{\frac{\chi^2}{n(\min[rows-1],[columns-1])}}$$ p. 289

Gamma

$$\gamma=\frac{Concordant\ pairs-Discordant\ pairs}{Concordant\ pairs+Discordant\ pairs}$$ p. 272

Lambda

$$\lambda=\frac{E_1-E_2}{E_1}$$ p. 266

$$E_1=n_{overall}-modal\ frequency_{overall}$$

$$E_2=\sum\left(n_{group}-modal\ frequency_{group}\right)$$

Means (averages)

Mean (average)

$$\bar{X}=\frac{\sum X_i}{N}\quad or\quad \bar{x}=\frac{\sum x_i}{n}$$ p. 98

Standard error of a mean

$$se_{\bar{x}}=\frac{s}{\sqrt{n}}$$ p. 163

95% confidence interval for a mean

$$95\%\ CI_{\bar{x}}=\bar{x}\pm1.96\left(se_{\bar{x}}\right)$$ p. 168

Percentage

$$\%=\left(\frac{F}{N}\right)100$$ p. 31

Phi

$$\phi=\sqrt{\frac{\chi^2}{n}}$$ p. 288

Probability

Simple probability

$$probability\ (outcome\ of\ interest)=\frac{number\ of\ outcomes\ of\ interest}{number\ of\ possible\ outcomes}$$ p. 132

Joint probability

$$probability\ (outcome\ of\ interest,\ outcome\ of\ interest)$$
$$=\frac{number\ of\ outcomes\ of\ interest}{number\ of\ possible\ outcomes}\times\frac{number\ of\ outcomes\ of\ interest}{number\ of\ possible\ outcomes}$$ p. 134

Proportions

Proportion

$$p=\left(\frac{F}{N}\right)\quad or\quad p=\left(\frac{f}{n}\right)$$ p. 31

Standard error of a proportion	$se_p = \sqrt{\frac{p(1-p)}{n}}$	p. 182
95% confidence interval for a proportion	$95\%\ CI_p = p \pm 1.96(se_p)$	p. 185

R^2 (Coefficient of determination)

R^2	$R^2 = \frac{\Sigma(\hat{y}_i - \bar{y})^2}{\Sigma(y_i - \bar{y})^2}$	p. 367
Adjusted R^2	$Adjusted\ R^2 = 1 - \left[\frac{(1-R^2)(n-1)}{n-k-1}\right]$	p. 455

Rates

Rate (common base)	$\frac{F}{N} = \frac{x}{base\ of\ the\ rate}$	p. 37
Rate (standardized numerator)	$\frac{F}{N} = \frac{1}{x}$	p. 37

Ratio

$ratio = \frac{F_1}{F_2}$ p. 36

Regression, Simple linear

Simple linear regression predictions	$\hat{y} = a + bx$	p. 356
Simple linear regression with residuals	$y_i = a + bx_i + e_i$	p. 365
Standard error of the estimate	$se_{est} = \sqrt{\frac{\Sigma(y_i - \hat{y}_i)^2}{n-2}}$	p. 372
Sum of products	$SP = \Sigma(x_i - \bar{x})(y_i - \bar{y})$	p. 318
Sum of squares (x)	$SS = \Sigma(x_i - \bar{x})^2$	p. 319
Constant coefficient	$a = \bar{y} - b(\bar{x})$	p. 363
Standard error of the constant coefficient (version 1)	$se_a = se_{est}\sqrt{\frac{1}{n} + \frac{(\bar{x})^2}{SS_x}}$	p. 373
Standard error of the constant coefficient (version 2)	$se_a = se_{est}\sqrt{\frac{1}{n} + \frac{(\bar{x})^2}{\Sigma(x_i - \bar{x})^2}}$	p. 373
Slope coefficient (version 1)	$b = \frac{SP}{SS_x}$	p. 360

Slope coefficient (version 2)	$b=\frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2}$	p. 360
Standard error of a slope coefficient (version 1)	$se_b=\frac{se_{est}}{\sqrt{SS_x}}$	p. 373
Standard error of a slope coefficient (version 2)	$se_b=\frac{se_{est}}{\sqrt{\Sigma(x_i-\bar{x})^2}}$	p. 373
95% confidence intervals for regression coefficients (large samples)	$95\%CI_b=b\pm1.96(se_b)$ $95\%CI_a=b\pm1.96(se_a)$	p. 378
T-statistic for regression coefficients	$t=\frac{b}{se_b}\quad or \quad t=\frac{a}{se_a}$	p. 374
Degrees of freedom of the t-statistic for regression coefficients	$df_t=n-k-1$	p. 377

Regression, multiple linear

Multiple linear regression predictions	$\hat{y}=a+b_1x_1+b_2x_2+b_3x_3+b_4x_4+b_5x_5+\cdots$	p. 407
Multiple linear regression with residuals	$y_i=a+b_1x_{1i}+b_2x_{2i}+b_3x_{3i}+b_4x_{4i}+b_5x_{5i}+\cdots+e_i$	p. 444
Constant coefficient (two independent variables)	$a=\bar{y}-b_1\bar{x}_1-b_2\bar{x}_2$	p. 412
Partial slope coefficients (two independent variables)	$b_{x_1}=\left(\frac{r_{x_1,y}-[r_{x_1,x_2}][r_{x_2,y}]}{1-[r_{x_1,x_2}]^2}\right)\left(\frac{s_y}{s_{x_1}}\right)$ $b_{x_2}=\left(\frac{r_{x_2,y}-[r_{x_1,x_2}][r_{x_1,y}]}{1-[r_{x_1,x_2}]^2}\right)\left(\frac{s_y}{s_{x_2}}\right)$	p. 410
Standardized slope coefficient (Beta)	$\beta_x=b_x\left(\frac{s_x}{s_y}\right)$	p. 416

Standard deviation

Standard deviation (using sample data)	$s=\sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n-1}}$	p. 140
Standard deviation (using population data)	$S=\sqrt{\frac{\Sigma(X_i-\bar{X})^2}{N}}$	p. 102

Standardized weight

$$standardized\ weight = \frac{weight\ for\ each\ case}{mean\ of\ the\ weight\ variable}$$ p. 148

T-statistic (unequal variances)

T-statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$ p. 210

Degrees of freedom of the t-statistic

$$df_t = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{(n_1-1)} + \frac{(s_2^2/n_2)^2}{(n_2-1)}}$$ p. 215

Variance

Variance (using sample data)

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$$ p. 140

Variance (using population data)

$$S^2 = \frac{\Sigma(X_i - \bar{X})^2}{N}$$ p. 103

Z-score

$$Z = \frac{(X_i - \bar{X})}{S}$$ p. 105