Summary of Formulas and Tests

Continuous data. Compare means of two populations (first row is for the mean of *one* population). Each can be either a one or two tailed test.

Assumptions	Paired	Sections	Analysis/ Test	Confidence interval	Test statistic	SE	Table	Table info needed
Normal distribution or large sample	–	6.2 – 6.5	t	$\mu = \bar{y} \pm t \cdot SE_{\bar{Y}}$	$t_s = \frac{\bar{y} - \bar{y}_{expected}}{SE_{\bar{Y}}}$	$\frac{s}{\sqrt{n}} = \sqrt{\frac{s^2}{n}}$	4	$df = n - 1$
Normal distribution or large samples	No	6.6 – 6.7 7.1 – 7.2 7.5	t	$\mu_1 - \mu_2 = (\bar{y}_1 - \bar{y}_2) \pm t \cdot SE_{\bar{Y}_1 - \bar{Y}_2}$	$t_s = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{SE_{\bar{Y}_1 - \bar{Y}_2}}$	$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{SE_1^2 + SE_2^2}$	4	df from (6.7.1)
None	No	7.10	Wilcoxon-Mann-Whitney	–	$U_s = \max(K_1, K_2)$	–	6	n and n' where $n \geq n'$
Normal distribution or large samples	Yes	8.2	Paired t	$\mu_D = \bar{d} \pm t \cdot SE_{\bar{D}}$	$t_s = \frac{\bar{d} - 0}{SE_{\bar{D}}}$	$\frac{s_D}{\sqrt{n_D}} = \sqrt{\frac{s_D^2}{n_D}}$	4	$df = n_D - 1$
None	Yes	8.4	Sign	–	$B_s = \max(N_+, N_-)$	–	7	n_D
None	Yes	8.5	Signed-Rank	–	$W_s = \max(W_+, W_-)$	–	8	n_D

Continuous data. Determine whether there is linear correlation between two variables, finding the slope for the line that fits the data.

Test whether $\rho = 0$	–	12.2	t	Not covered–optional	$t_s = r\sqrt{\frac{n-2}{1-r^2}}$	–	4	$df = n - 2$
Test whether $\beta_1 = 0$	–	12.2 – 12.5	t	$\beta_1 = b_1 \pm t \cdot SE_{b_1}$	$t_s = \frac{b_1 - 0}{SE_{b_1}}$	$\frac{s_e}{s_x\sqrt{n-1}}$	4	$df = n - 2$

Also: $b_1 = r \cdot \left(\frac{s_y}{x_x}\right)$, $s_e = \sqrt{\frac{SS(resid)}{n-2}}$

Categorical data. Confidence interval for a single population proportion or for the difference between two population proportions. All use Table 4 ($df = \infty$).

Assumptions	Section	Confidence interval	$\tilde{p}$	$SE_{\tilde{p}}$
One population 95% confidence	9.2	$p = \tilde{p} \pm 1.96 \cdot SE_{\tilde{p}}$	$\dfrac{y+2}{n+4}$	$\sqrt{\dfrac{\tilde{p}(1-\tilde{p})}{n+4}}$
One population Any confidence	9.3	$p = \tilde{p} \pm z_{\alpha/2} \cdot SE_{\tilde{p}}$	$\dfrac{y+0.5(z_{\alpha/2}^2)}{n+z_{\alpha/2}^2}$	$\sqrt{\dfrac{\tilde{p}(1-\tilde{p})}{n+z_{\alpha/2}^2}}$
Two populations 95% confidence	10.7	$p_1 - p_2 = (\tilde{p}_1 - \tilde{p}_2) \pm 1.96 \cdot SE_{\tilde{P}_1 - \tilde{P}_2}$	$\dfrac{y+1}{n+2}$	$\sqrt{\dfrac{\tilde{p}_1(1-\tilde{p}_1)}{n_1+2} + \dfrac{\tilde{p}_2(1-\tilde{p}_2)}{n_2+2}}$

Not covered in book: we could also compute $t_s = \frac{\hat{p}-p}{\sqrt{n\,p\,(1-p)}}$ (or use $\tilde{p}$ rather than $\hat{p}$), and then use Table 4 ($df = \infty$) to find P.

Categorical data. Compare proportions within a population or between two populations. All use Table 9.

Expected frequencies are $e_i = \frac{(Row\ Total) \times (Column\ Total)}{Grand\ Total}$. For one population with one category, the e_i are as given in H_0.

Assumptions	k	r	Paired	Analysis/ Test	Sections	Test statistic	df	Tailed Test
Each expected value is sufficiently large, e.g. $E \geq 5$	1	≥ 2	No	Chi-square	9.4	$\chi_s^2 = \sum \dfrac{(o_i - e_i)^2}{e_i}$	$r-1$	Two only (unless $r = 2$)
	2	2	No		10.2 – 10.3	$\chi_s^2 = \sum \dfrac{(o_i - e_i)^2}{e_i}$	$(2-1)(2-1) = 1$	Can be either one or two
			Yes		10.8	$\chi_s^2 = \dfrac{(n_{12} - n_{21})^2}{n_{12} + n_{21}}$		
	≥ 2	≥ 2	No		10.5	$\chi_s^2 = \sum \dfrac{(o_i - e_i)^2}{e_i}$	$(r-1)(k-1)$	Two only (unless $r = k = 2$)

Continuous data. Compare means of three or more populations. All tests use Table 10. All tests are two-tailed (non-directional) only.
All assume approximately normal distribution and/or large samples, and approximately equal standard deviation in each population.

Analysis/ Test	Sections	Levels/Categories in each Factor/Treatment	Test statistic	df	Example
One-way ANOVA	11.2 – 11.4	I	$F_s = \frac{MS(Between\ Factor\ Levels)}{MS(Within)} = \frac{\frac{SS(Between\ Factor\ Levels)}{df(Between\ Factor\ Levels)}}{\frac{SS(Within)}{df(Within)}}$ Notes: $F_s = t_s^2$ if $I = 2$ $MS(within) = s_{pooled}^2$	Between Factor Levels/Numerator: $df = I - 1$ Within Factor Levels/Denominator: $df = n_. - I$ $= (n_1 - 1) + \cdots + (n_I - 1)$ $= (n_. - 1) - (df\ Between)$	Table 11.2.3, Example 11.4.1
One-way ANOVA, with blocking	11.6	Factor: I Blocks: J	$F_s = \frac{MS(Between\ Factor\ Levels\ or\ Blocks)}{MS(Within)} = \frac{\frac{SS(Between\ Factor\ Levels\ or\ Blocks)}{df(Between\ Factor\ Levels\ or\ Blocks)}}{\frac{SS(Within)}{df(Within)}}$	Between Factor Levels/Numerator: $df = I - 1$ Between Blocks/Numerator: $df = J - 1$ Within Factor Levels/Denominator: $df = (n_. - 1) - (sum\ of\ others)$	Table 11.6.4, Example 11.6.9
Two-way ANOVA	11.7	Factor 1: I Factor 2: J	$F_s = \frac{MS(Between\ Factor\ Levels\ or\ Interaction)}{MS(Within)} = \frac{\frac{SS(Between\ Factor\ Levels\ or\ Interaction)}{df(Between\ Factor\ Levels\ or\ Interaction)}}{\frac{SS(Within)}{df(Within)}}$ Note: First check F_s for Interaction. If not significant Interaction, then check F_s for the two Factors.	Between Factor 1 Levels/Numerator: $df = I - 1$ Between Factor 2 Levels/Numerator: $df = J - 1$ Interaction/Numerator: $df = (I - 1) \times (J - 1)$ Within/Denominator: $df = (n_. - 1) - (sum\ of\ others)$	Table 11.7.5, Example 11.7.5