Multiple comparisons - subsequent inferences for two-way ANOVA

- the kinds of inferences to be made after the F tests of a two-way ANOVA depend on the results
- if none of the F tests lead to rejection of the null hypothesis, then you have concluded that none of the means are different and no further comparisons are required

Significant interactions

- if it has been determined that the interactions are significant, then the effect of each factor depends on the level of the other factor
- another way of saying this is that each cell of the table has a possibly different mean μ_{ij}
- there are IJ such means and $r = \begin{pmatrix} IJ \\ 2 \end{pmatrix}$ possible comparisons among them
- to control the overall type 1 error rate at α, the Bonferroni correction uses $\alpha_* = \alpha/r$ for each comparison
- confidence intervals have the form

$$\bar{y}_{ij.} - \bar{y}_{kl.} \pm t_{\alpha_*/2, IJ(K-1)} \sqrt{MSE}\sqrt{2/K}$$

- tests of $H_0 : \mu_{ij} = \mu_{kl}$ versus $H_a : \mu_{ij} \neq \mu_{kl}$ are based on the statistic

$$t_{ij,kl} = \frac{\bar{y}_{ij.} - \bar{y}_{kl.}}{\sqrt{MSE}\sqrt{2/K}}$$

- the null hypothesis is rejected when $P < \alpha_*$ or when

$$|t_{ij,kl}| > t_{\alpha_*/2, IJ(K-1)}$$

- it is often easiest to rearrange this expression and reject H_0 when

$$|\bar{y}_{ij.} - \bar{y}_{kl.}| > t_{\alpha_*/2, IJ(K-1)} \sqrt{MSE}\sqrt{2/K}$$

- the right hand side remains constant so it is just a matter of looking at the differences between any two cell means

Example: The following data are the lifetimes (in hours) of four different designs of an airplane wing subjected to three different kinds of continuous vibrations.

	Design			
	1	2	3	4
Vibration 1	876	1156	1234	825
	913	1219	1181	797
Vibration 2	1413	1876	1591	1083
	1290	1710	1649	1161
Vibration 3	1291	2115	1650	1148
	1412	1963	1712	1262

- the ANOVA table is

Source	SS	DF	MS	F	P
Vibration	1346145	2	673073	139.7	
Design	1457096	3	485699	100.79	
Interaction	138403	6	23067	4.79	.01
Error	57824	12	4819		
Total	2999469	23			

- the test of H_0: *no interactions* versus H_a: *there are interactions* gives a p-value of .01, so we reject H_0 testing at the $\alpha = .05$ level.
- to determine which combinations of vibration and design give significantly different results at we use the cell means

	Design			
	1	2	3	4
Vibration 1	894.5	1187.5	1207.5	811
Vibration 2	1351.5	1793	1620	1122
Vibration 3	1351.5	2039	1681	1205

- the cell means are significantly different if their absolute difference is greater than

$$t_{\alpha_*/2,IJ(K-1)}\sqrt{MSE}\sqrt{2/K}$$

- with $I = 3$, $J = 4$ and $K = 2$ there are $r = \binom{12}{2} = 66$ possible comparisons
- in this case $MSE = 4819$, $\alpha_* = .05/66 = 0.000758$ and $t_{\alpha_*/2,12} = 4.48$, from minitab or R
- so the difference required for significance is

$$4.48 * \sqrt{4819} * \sqrt{2/2} = 310.73$$

- one way to do these comparisons efficiently is to rank them from smallest to largest, using the notation (i, j) to represent the ith vibration and jth design

(i,j)	mean	mean + 310.73
(1,4)	811	1121.73
(1,1)	894.5	1205.23
(2,4)	1122	1432.73
(1,2)	1187.5	1498.23
(3,4)	1205	1515.73
(1,3)	1207.5	1518.23
(2,1)	1351.5	1662.23
(3,1)	1351.5	1662.23
(2,3)	1620	1930.73
(3,3)	1681	1991.73
(2,2)	1793	2103.73
(3,2)	2039	

- the extra column shows the value required for a mean to be significantly different
- so (1,4) is not different from (1,1)
- (1,1) is not different from (2,4), (1,2) and (3,4)
- (2,4) is not different from (1,2), (3,4), (1,3), (2,1) and (3,1)
- (1,2) is not different from (3,4), (1,3), (2,1) and (3,1)
- (3,4) is not different from (1,3), (2,1) and (3,1)
- (1,3) is not different from (2,1) and (3,1)
- (2,1) is not different from (3,1) and (2,3)
- (3,1) is not different from (2,3)
- (2,3) is not different from (3,3) and (2,2)
- (3,3) is not different from (2,2)
- (2,2) is not different from (3,2)
- all other differences are significant

if the interactions were not significant

- if it is determined that the interactions are not significant then the main effects can be tested
- if both the row and column factors are significant then there are

$$r = \binom{I}{2} + \binom{J}{2}$$

  pairwise comparisons of interest

- if only the row factor or column factor is significant, $r = \binom{I}{2}$ or $r = \binom{J}{2}$ respectively
- in either case, the Bonferroni correction for multiple comparisons uses $\alpha_* = \alpha/r$
- comparisons between the rows are made using row means, so that confidence intervals have the form

$$\bar{y}_{i..} - \bar{y}_{l..} \pm t_{\alpha_*/2, IJ(K-1)} \sqrt{MSE} \sqrt{\frac{2}{JK}}$$

- (note that there are JK observations in each row)
- in comparing the rows we are making inferences about the difference in row effects $\alpha_i - \alpha_l$
- the test of $H_0 : \alpha_i - \alpha_l = 0$ versus $H_a : \alpha_i - \alpha_l \neq 0$ uses

$$t_{il} = \frac{\bar{y}_{i..} - \bar{y}_{l..}}{\sqrt{MSE}\sqrt{\frac{2}{JK}}}$$

- because the denominator is the same for all such statistics, one can simply compare the absolute difference $|\bar{y}_{i..} - \bar{y}_{l..}|$ to $t_{\alpha_*/2, IJ(K-1)} \sqrt{MSE} \sqrt{\frac{2}{JK}}$ and conclude the difference is significant if the former is larger than the latter
- similarly, comparisons between the columns are made using column means, so that confidence intervals have the form

$$\bar{y}_{.j.} - \bar{y}_{.u.} \pm t_{\alpha_*/2, IJ(K-1)} \sqrt{MSE} \sqrt{\frac{2}{IK}}$$

- (note that there are IK observations in each column)
- in comparing the columns we are making inferences about the difference in column effects $\beta_j - \beta_u$
- the test of $H_0 : \beta_j - \beta_u = 0$ versus $H_a : \beta_j - \beta_u \neq 0$ uses

$$t_{ju} = \frac{\bar{y}_{.j.} - \bar{y}_{.u.}}{\sqrt{MSE}\sqrt{\frac{2}{IK}}}$$

- because the denominator is the same for all such statistics, one can simply compare the absolute difference $|\bar{y}_{.j.} - \bar{y}_{.u.}|$ to $t_{\alpha_*/2, IJ(K-1)} \sqrt{MSE} \sqrt{\frac{2}{IK}}$, and conclude the difference is significant if the former is larger than the latter

Example: For the data on burn rates with 3 different engines and 4 different propellants, we determined that there were no interactions but that both factors were significant, when testing at level $\alpha = .05$.

- there are 3 comparisons to be made among the engines and 6 to be made among the propellants
- for an overall error rate of $\alpha = .05$, the Bonferroni correction uses $\alpha_* = .05/9 = 0.0056$
- from the output from the model with interaction, $MSE = 1.2425$, with 12 degrees of freedom (Note that even though the interaction was NOT significant, we use the MSE, and associated degrees of freedom, from the model that included interaction.
- the critical t value is $t_{\alpha_*/2,12} = 3.37$ using a computer program
- the engine means are $\bar{y}_{1..} = 30.5$, $\bar{y}_{2..} = 29.675$ and $\bar{y}_{3..} = 28.60$
- the required difference in engine means for significance is

$$\begin{aligned} t_{\alpha_*/2,IJ(K-1)}\sqrt{MSE}\sqrt{\frac{2}{JK}} &= 3.37\sqrt{1.2425}\sqrt{2/8} \\ &= 3.37(1.1147)(.5) \\ &= 1.8782 \end{aligned}$$

- using this approach we conclude that α_1 is significantly different from α_3, but that α_1 is not different from α_2 and α_2 is not different from α_3
- the propellant means are $\bar{y}_{.1.} = 31.6$, $\bar{y}_{.2.} = 29.85$, $\bar{y}_{.3.} = 28.38$ and $\bar{y}_{.4.} = 28.53$
- the required difference in propellant means for significance is

$$\begin{aligned} t_{\alpha_*/2,IJ(K-1)}\sqrt{MSE}\sqrt{\frac{2}{IK}} &= 3.37\sqrt{1.2425}\sqrt{2/6} \\ &= 2.1689 \end{aligned}$$

- examining the propellant means shows that 3 and 4 are significantly different from 1, but that none of the other comparisons are significant