Multiple comparisons - subsequent inferences for one-way ANOVA

- if the overall F test does not show significant differences among the groups, then no further inferences are required
- if the overall test of equality of means concludes in favour of the alternative H_A : not all means are the same, then the natural question is "which of the means are difference"

  The differences between a particular means, say of the i'th and k'th populations can be tested with a t-test, using the test statistic

$$T = \frac{\bar{x}_{i.} - \bar{x}_{k.}}{\sqrt{MSE}\sqrt{\frac{1}{n_i} + \frac{1}{n_k}}}$$

- The null hypothesis here is $H_{0,ik} : \mu_i = \mu_k$ and the alternative is $H_{A,ik} : \mu_i \neq \mu_k$.
- in this expressions, MSE is the estimate of σ^2 from the analysis of variance
- the degrees of freedom for the t-test is $N - a$, which is the degrees of freedom associated with MSE in the ANOVA
- **adjustment must be made for the fact that we are doing multiple comparisons**, that is, for the fact that several tests are being done, sometimes known as **simultaneous inference**
- the simplest adjustment is the *Bonferroni correction*, which reduces the significance level for each test so that the overall probability of making at least one type I error is no larger than the level α associated with the ANOVA
- in a one-way ANOVA with a groups, there are $r = \binom{a}{2}$ natural comparisons between pairs of groups
- if you do r tests each at level α, then the probability of incorrectly rejecting at least one null hypothesis could be as large as $r\alpha$
  - for example for $r = 2$

$$P(\textit{reject at least one } H_0) =$$

$$P(\textit{reject } 1st) + P(\textit{reject } 2nd)$$

$$-P(\textit{reject both}) \leq 2\alpha$$

- to control the overall level, or *experimentwise error rate*, at α, each test should be done using $\alpha_* = \alpha/r$

- alternatively the P value should be multiplied by r
- similarly for r confidence intervals, use of α_* will give simultaneous confidence level $1-\alpha$
- These simultaneous confidence intervals for the differences of two means are of the form

$$\left(\bar{x}_{i.} - \bar{x}_{k.} - t_{\alpha*/2,N-a}\sqrt{MSE}\sqrt{\frac{1}{n_i}+\frac{1}{n_k}}, \bar{x}_{i.} - \bar{x}_{k.} + t_{\alpha*/2,N-a}\sqrt{MSE}\sqrt{\frac{1}{n_i}+\frac{1}{n_k}}\right)$$

Example: for the golf balls, the summary statistics are

i	$\bar{x}_i$	s_i^2	n_i
1	251.28	33.487	5
2	261.98	18.197	5
3	269.66	27.253	5

- the value for MSE is 26.312
- there are 3 possible pairwise comparisons between the groups
- the denominator of the test statistics is

$$\sqrt{MSE}\sqrt{\frac{1}{n_i}+\frac{1}{n_k}} = 5.1295(.6325) = 3.24$$

- the degrees of freedom are 12
- with $\alpha = .05$, $\alpha_* = .05/3 = .0167$, $\alpha_*/2 = .00833$, and $t_{.00833,12} = 2.7794$,
- found, for example, as

```
MTB > invcdf .00833;
SUBC> t 12.

Inverse Cumulative Distribution Function

Student's t distribution with 12 DF

P(X<=x)          x
    0.00833  -2.77969
```

- the test statistics are

$$t_{12} = \frac{251.18 - 261.98}{3.24} = -3.329$$

$$t_{13} = \frac{251.18 - 269.66}{3.24} = -5.697$$

and

$$t_{23} = \frac{261.18 - 269.66}{3.24} = -2.367$$

- the first two comparisons are significant at the .05 level but the third one is not
- confidence intervals for the differences in means are

$$-10.8 \pm 2.78(3.24) \quad or \quad (-19.81, -1.79)$$

$$-18.48 \pm 9.01 \quad or \quad (-27.49, -9.47)$$

and

$$-7.68 \pm 9.01 \quad or \quad (-16.69, 1.33)$$

Example: for the liver weights, the means in ascending order are

diet	B	C	A	D
n_i	8	6	7	8
mean	3.43	3.598	3.803	3.935

- the estimated standard deviation is
  $\sqrt{MSE} = .1899$
- there are 6 comparisons, so the
  appropriate table value for $\alpha = .05$ is $t_{25}^{.025/6} = 2.8649$, from MINITAB

- the pairwise differences in the means are

i/k	B	C	D
A	.373	.205	-.132
B		-.168	-.505
C		-.337	

- the absolute difference in means must exceed $t_{25}^{.025/6}\sqrt{MSE}\sqrt{\frac{1}{n_i}+\frac{1}{n_k}}$, which depends on the two sample sizes.

n_i/n_k	7	8
6	.3025	.2938
7		.2818
8		.2720

- using this table, we find that B and C, C and A and A and D are not statistically significant, the other 3 comparisons are significant