Applied Statistical Methods
Factorial experiments, One Way ANOVA
Multiple comparisons

Pairwise Comparisons in One-Factor Experiments

- **In a one-way ANOVA, an *F* test is used to test the null hypothesis that all the treatment means are equal.**

- **If this hypothesis is rejected, we can conclude that the treatment means are not all the same. But the test does not tell us which ones are different from the rest.**

- **Case 1. An experimenter has in mind two specific treatments, *i* and *j*, and wants to study the difference $\mu_i - \mu_j$.**

  - Use Fisher's least significant difference (LSD) method to construct confidence intervals for $\mu_i - \mu_j$ or to test the null hypothesis that $\mu_i - \mu_j = 0$.

- **Case 2. The experimenter may want to determine all the pairs of means that can be concluded to differ from each other.**

  - Use a multiple comparisons method, such as the Tukey method.

Fisher's Least Significant Difference Method for Confidence Intervals

The Fisher's least significant difference confidence interval, at level 100(1 – α)%, for the difference $\mu_i - \mu_j$ is

$$\bar{X}_{i.} - \bar{X}_{j.} \pm t_{N-I,\alpha/2}\sqrt{\mathrm{MSE}\left(\frac{1}{J_i} + \frac{1}{J_j}\right)}$$

Weld example cont.

In the weld experiment, hardness measurements were made for five welds from each of four fluxes, A, B, C, and D. The sample mean hardness values are 253.8, 263.2, 271.0, and 262.0, respectively.

Before the experiment was performed, the carbon contents of the fluxes were measured. Flux B had the lowest carbon content and flux C had the highest. The experimenter is therefore particularly interested in comparing the hardness obtained with these two fluxes.

Find a 95% confidence interval for the difference in mean hardness between welds produced with flux B and those produced with flux C. Can we conclude that the two means differ?

Fisher's LSD Confidence Intervals

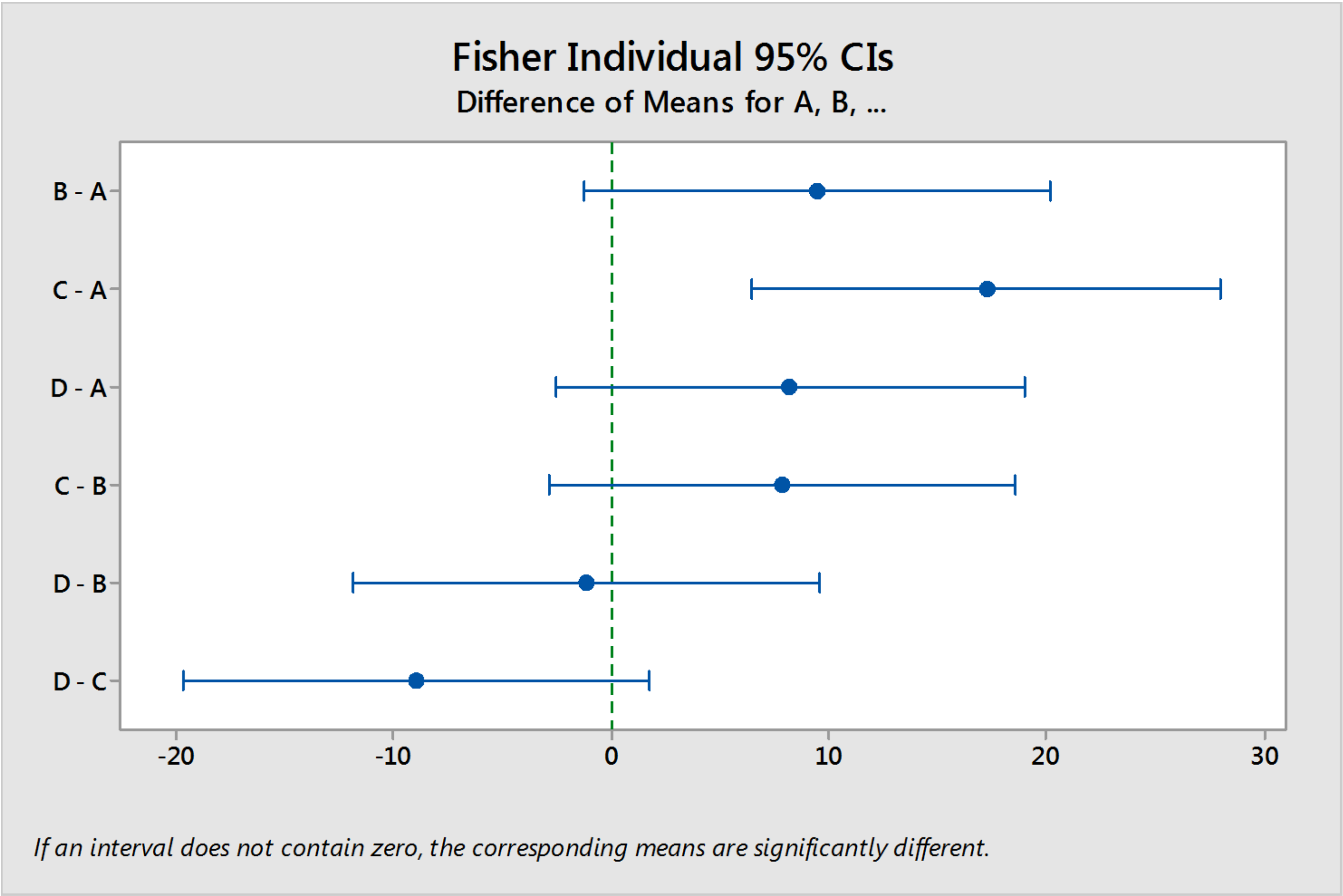
```
Fisher 95% Individual Confidence Intervals
All Pairwise Comparisons

Simultaneous confidence level = 81.11%

A subtracted from:

    Lower   Center   Upper  ------+---------+---------+---------+---
B  -1.324    9.400  20.124                 (-------*-------)
C   6.476   17.200  27.924                     (-------*-------)
D  -2.524    8.200  18.924                (-------*-------)
                            ------+---------+---------+---------+---
                                -12         0        12        24

B subtracted from:

     Lower   Center   Upper  ------+---------+---------+---------+---
C   -2.924    7.800  18.524                (-------*------)
D  -11.924   -1.200   9.524        (-------*-------)
                             ------+---------+---------+---------+---
                                 -12         0        12        24

C subtracted from:

      Lower   Center   Upper  ------+---------+---------+---------+---
D   -19.724   -9.000   1.724  (-------*------)
                              ------+---------+---------+---------+---
                                  -12         0        12        24
```

We are 95% confident that any given confidence interval contains the true difference in means.

"Center" are the differences between pairs of treatment means

"Lower" and "Upper" are the lower and upper bounds of the confidence interval.

Fisher's Individual CIs



The mean weld hardness does not differ significantly between fluxes B and C, on a significance level 0.05, because the Fisher 95% CI contains zero.

Simultaneous Tests

- **The simultaneous confidence level of 81.11% in the output indicates, that we are only 81.11% confident that *all* the confidence intervals contain their true mean differences.**

- **When several confidence intervals or hypothesis tests are to be considered simultaneously, multiple comparison methods are used to produce simultaneous confidence intervals or simultaneous hypothesis tests.**

- **If level $100(1-\alpha)\%$ simultaneous confidence intervals are constructed for differences between every pair of means, then we are confident at the $100(1-\alpha)\%$ level that *every* confidence interval contains the true difference.**

- **If simultaneous hypothesis tests are conducted for all null hypotheses of the form H_0: $\mu_i - \mu_j = 0$, then we may reject, at level α, every null hypothesis whose *P*-value is less than α.**

Tukey's method for simultaneous confidence intervals

- **The Tukey method is based on a distribution called the Studentized range distribution.**
- **The Studentized range distribution has two values for degrees of freedom, which for the Tukey method are *I* and *N – I*.**
- **The Tukey method uses the $1-\alpha$ quantile of the Studentized range distribution with *I* and *N – I* degrees of freedom, this quantity is denoted $q_{I,\,N-I\,,\alpha}$**
- **An alternative to the Tukey's method is the Bonferroni Method (not covered in this class).**

Tukey Method for Simultaneous Confidence Intervals

- **The Tukey level 100(1 − α)% simultaneous confidence intervals for all differences $\mu_i - \mu_j$ are**

$$\overline{X}_{i.} - \overline{X}_{j.} \pm q_{I,N-I,\alpha} \sqrt{\frac{\text{MSE}}{2}\left(\frac{1}{J_i} + \frac{1}{J_j}\right)}$$

- **We are 100(1 − α)% confident that the Tukey confidence intervals contain the true value of the difference $\mu_i - \mu_j$ for every *i* and *j*.**

Tukey Simultaneous Confidence Intervals

We are 95% confident that every one of these confidence intervals contains the true difference in treatment means.

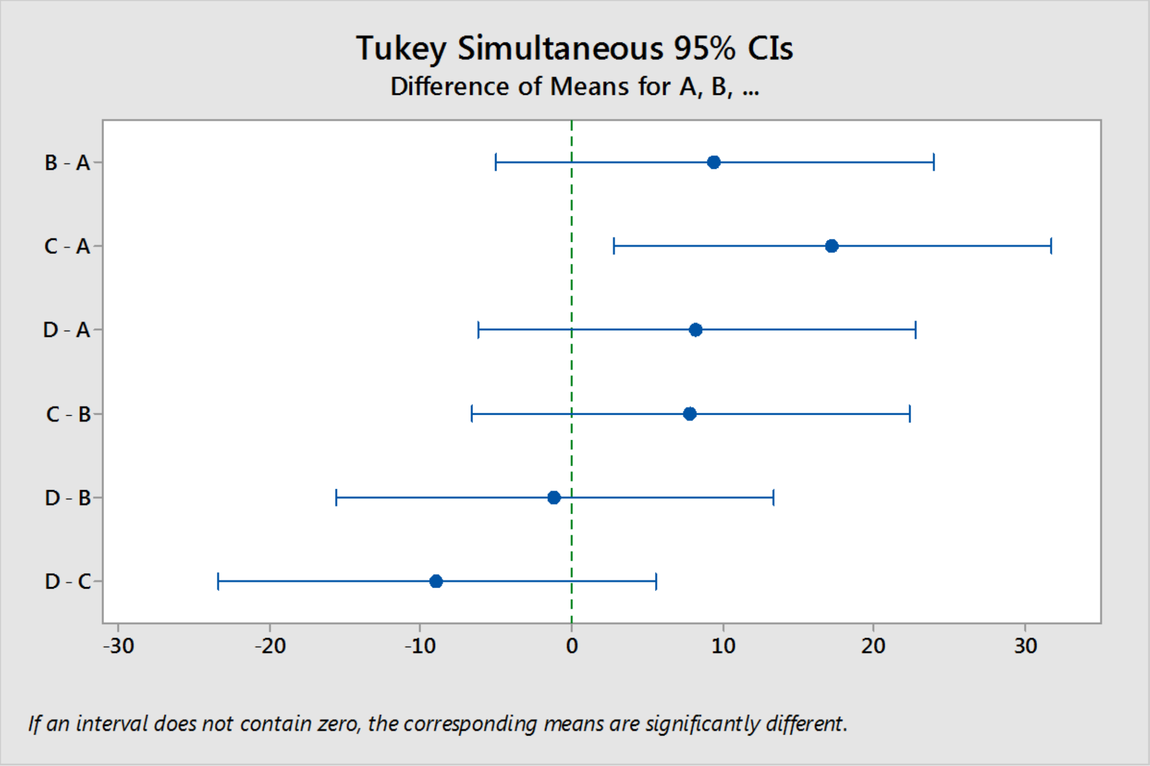
```
Tukey 95% Simultaneous Confidence Intervals
All Pairwise Comparisons

Individual confidence level = 98.87%

A subtracted from:

     Lower   Center    Upper  ------+---------+---------+---------+---
B   -5.087   9.4000   23.887              (--------*--------)

C    2.713   17.200   31.687                      (--------*--------)
D   -6.287    8.200   22.687             (--------*--------)
                              ------+---------+---------+---------+---
                                  -15         0        15        30

B subtracted from:

      Lower   Center    Upper  ------+---------+---------+---------+---
C    -6.687    7.800   22.287            (--------*--------)
D   -15.687   -1.200   13.287       (--------*--------)
                               ------+---------+---------+---------+---
                                   -15         0        15        30

C subtracted from:

      Lower   Center    Upper  ------+---------+---------+---------+---
D   -23.487   -9.000    5.487  (--------*--------)
                               ------+---------+---------+---------+---
                                   -15         0        15        30
```

The "Individual confidence level" is 98.87%, this means that we are 98.87% confident that a specific confidence interval contains its true value.

Weld Example cont.

For the weld data, which pair of fluxes, if any, can be concluded, at the 5% level, to differ in their effect on hardness?

The C and A fluxes differ significantly on a 5% significance method.



Fixed Effects Model

- **If particular treatments are chosen deliberately by the experimenter, rather than at random from a larger population of treatments, then we say that the experiment follows a fixed effects model.**

- **Example: Test four tomato fertilizers for their effect on yield. We have four treatments that we have chosen to test.**

- **In a fixed effects model, the only conclusions that can be drawn are conclusions about the treatments actually used in the experiment.**

- **Example: Test hypothesis: Ho: mean1=mean2=mean3=mean4**
  All our conclusions are only about the (4) treatments used in the experiment.

Random Effects Model

- In some experiments, treatments are chosen at random from a population of possible treatments. In this case, the experiment is said to follow a random effects model.

- Example: Determine whether or not different flavors of ice cream melt at different speeds. We test a random sample of three flavors for a large population of flavors offered to the customer by a single manufacturer.

- In a random effects model, the interest is in the whole population of possible treatments and there is no particular interest in the ones that happen to be chosen for the experiment.

- Since the treatments are a simple random sample from a population of treatments, conclusions can be drawn concerning the whole population, including treatments not actually used in the experiment.

- In the random effects model, the null hypothesis of interest is
  H_0: the treatment means are equal for every level in the population.

- Note: The methods of analysis for fixed and random effects models are essentially the same, although the conclusions to be drawn from them differ.