Experimental Design and Biostatistics (TAMS38)
Lecture 3 – Pairwise comparisons

Martin Singull

Department of Mathematics
Mathematical Statistics
Linköping University, Sweden

LINKÖPING
UNIVERSITY

Content

- Residual analysis
- Simultaneous confidence level
- Linear combination of μ_i
- Pairwise comparisons
  - t-intervals
  - Scheffé's method
  - Tukeys method
  - Tukey-Kramers method
- Examples - Strontium concentrations, Sheet Metal Manufacturing

Residual analysis

We return to the case of fixed effects and observations y_{il} from the random variables

$$Y_{il} = \mu + \tau_i + \varepsilon_{il} = \mu_i + \varepsilon_{il}$$

for $i = 1, \ldots, a$, $l = 1 \ldots, n_i$, where μ, τ_i and μ_i denotes fix unknown parameters and the random variable $\varepsilon_{il} \sim N(0, \sigma)$.

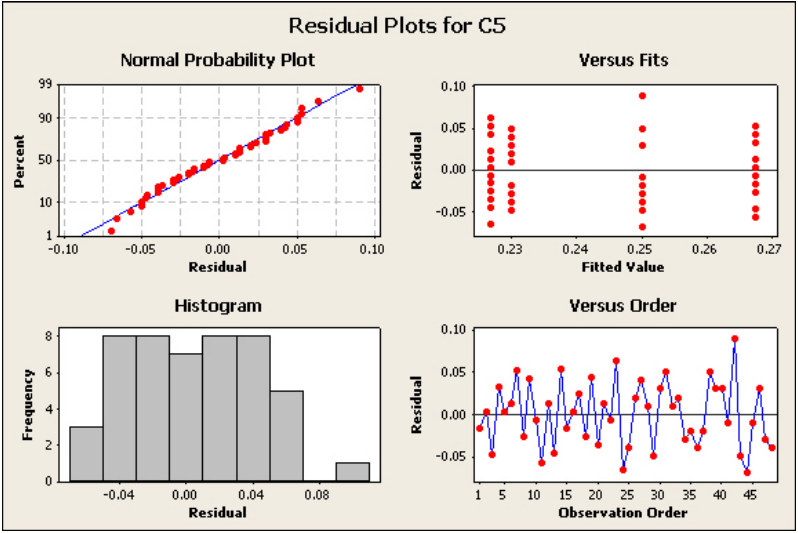
Fitted value: The estimated expected value for each level $\widehat{\mu}_i = \bar{y}_{i\cdot}$ is called *fitted value*.

Residuals: The estimated errors $e_{il} = y_{il} - \bar{y}_{i\cdot} = y_{il} - \widehat{\mu}_i$ are called *residuals* and they estimate ε_{il}.

Analyzing the residuals we can determine whether the conditions in the model seems to be fulfilled.

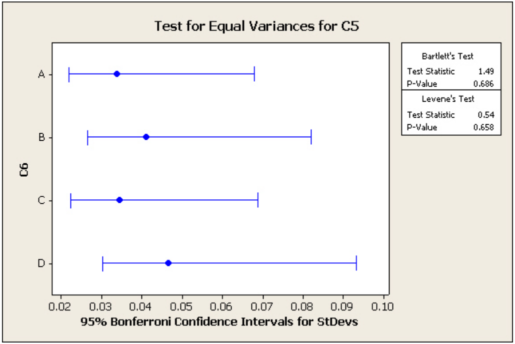
Example - Sheet Metal Manufacturing (from Lecture 1)

A steel industry make sheet iron with a thin layer of tin...

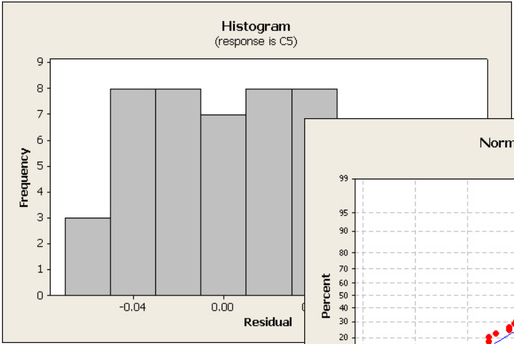


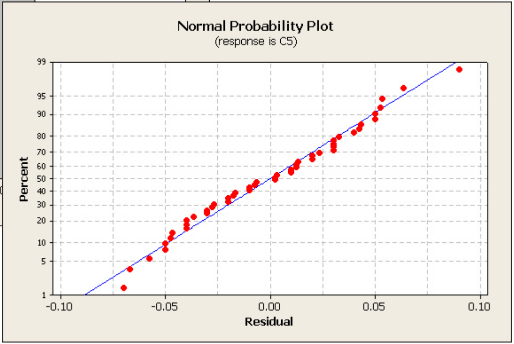
On Lecture 1 we saw how to compare two variances using F-test. If you make many such comparisons, one must have a low level of significance in the individual tests to get an acceptable simultaneous level.

There are also comprehensive test of equal variances: **Bartletts test** and **Levenes test**, see the book. These are also available in Minitab.



As we want to assess the normal distribution assumption is met, we can make the histogram or Normal probability plot.





Normal probability plot

Let $x_1, \ldots, x_N$ be observations from $N(\mu,\ \sigma)$. Rank them in order of size $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(N)}$. One can show that

$$E\left[\Phi\left(\frac{X_{(j)} - \mu}{\sigma}\right)\right] = \frac{j}{N+1}$$

Normal probability plot base on the idea that

$$\Phi\left(\frac{x_{(j)} - \mu}{\sigma}\right) \approx \frac{j}{N+1} \ \Leftrightarrow\ \Phi^{-1}\left(\frac{j}{N+1}\right) \approx \frac{x_{(j)} - \mu}{\sigma},$$

i.e., that the relationship between $x_{(j)}$ and $\Phi^{-1}\left(\frac{j}{N+1}\right)$ is approximately linear.

Instead of "normal score" $= j/(N+1)$ one can use, as it is done in the book, $(j-0.5)/N$ or, as it is in Minitab, $(j-3/8)/(N+1/4)$.

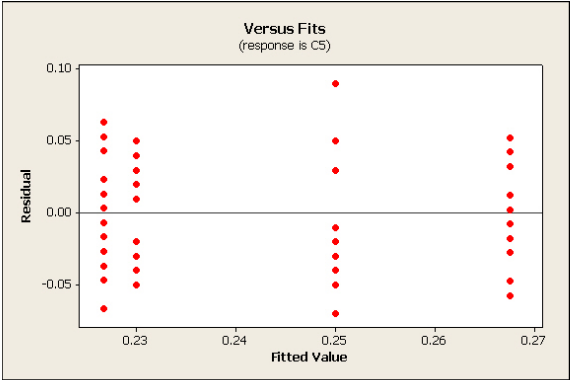
In a normal probability plot the points

$$\left(x_{(j)},\ \Phi^{-1}\left(\frac{j-0.5}{N}\right)\right)$$

(or other normal score) are plotted and if the random variable $X_j \sim N(\mu,\ \sigma)$ points should be **approximately on a straight line**.

Residual analysis, cont.

The residuals may also plotted against $\widehat{\mu}_i$, i.e., against "fitted value".



If one gets a cone shape, i.e., larger residuals for large estimated expectations, it may be appropriate to logarithm values (or use some other transformation) before doing analysis.

Transformations

- Sometimes it is quite obvious that the assumption of constant variance is not met.
- It is not unusual that there is a link between μ_i and σ_i.
- One can sometimes find a suitable transformation by studying the relationship between $\widehat{\mu}_i$ and s_i, see also Section 3.4.3 in the book and Chapter 15.1, Box-Cox-method.

Simultaneous confidence level

One often use **One-way ANOVA** to compare a treatments.

The observed values y_{il} are observations of random variable

$$Y_{il} = \mu + \tau_i + \varepsilon_{il} = \mu_i + \varepsilon_{il},$$

where μ_i is the characteristic of treatment i.

If one, with the F-test, show that μ_i-values with high probability are different, one often wants to establish which treatment is best. Then you do pairwise comparisons between different μ_i.

There will be

$$\binom{a}{2} = \frac{a(a-1)}{2}$$

possible intervals. $\Rightarrow$ Control the simultaneous confidence level.

Assume that we have $I_1, \ldots, I_{10}$ which are **independent** confidence intervals for $\theta_1, \ldots, \theta_{10}$ each with confidence level 95%.

The **simultaneous confidence level** for $I_1, \ldots, I_{10}$ is then

$$P(\theta_k \in I_k \text{ for } k = 1, \ldots, 10) \underset{indep.}{=} 0.95^{10} \approx 0.60$$

Low probability!

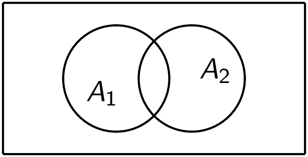
Remark: In our applications, the intervals usually are <u>not</u> independent.

Let I_1 and I_2 be two **dependent** confidence intervals for θ_1 and θ_2, each on confidence level $1 - \alpha$.

Then the simultaneous confidence level is

$$P(\theta_1 \in I_1 \text{ och } \theta_2 \in I_2) = 1 - P(\overbrace{\theta_1 \notin I_1}^{A_1} \text{ eller } \overbrace{\theta_2 \notin I_2}^{A_2})$$
$$= 1 - P(A_1 \cup A_2) \geq 1 - (P(A_1) + P(A_2)) = 1 - 2\alpha$$

according to Bonferronis inequality: $P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$



Let $I_1, \ldots, I_r$ be confidence intervals for $\theta_1, \ldots, \theta_r$ each with confidence level $1-\alpha$. Then $I_1, \ldots, I_r$ have simultaneous confidence level

$$1 - \alpha_{sim} = P(\theta_k \in I_k \text{ for } k = 1, \ldots, r) \geq 1 - r\alpha$$

according to Bonferronis inequality.

The overall risk to some interval miss its parameter $\alpha_{sim} \leq r\alpha$. So we choose $\alpha = \alpha_{sim}/r$.

Example We should do five confidence intervals and want to have simultaneous confidence level at least 95%. Then, we choose so individual confidence level $1 - \frac{0.05}{5} = 0.99$.

Linear combinations of μ_i

We have estimation variables

$$\widehat{\mu}_i = \overline{Y}_{i\cdot} \sim N\left(\mu_i, \frac{\sigma}{\sqrt{n_i}}\right),$$

and σ^2-estimator

$$s^2 = \frac{SS_E}{N-a}$$

with $N-a$ degrees of freedom, where $N = \sum_{i=1}^{a} n_i$.

We are often interested in studying **linear contrasts**, i.e., $\sum_{i=1}^{a} c_i \mu_i$ where $\sum_{i=1}^{a} c_i = 0$.

We have point estimator $\sum_{i=1}^{a} c_i \bar{y}_{i\cdot}$ with corresponding random variable

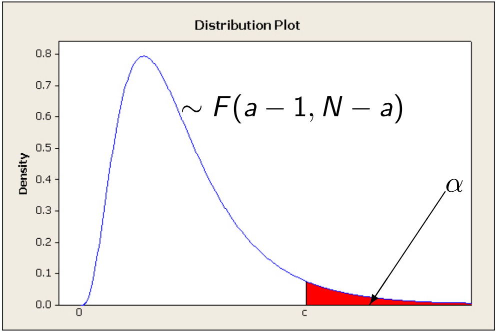
$$\sum_{i=1}^{a} c_i \overline{Y}_{i\cdot} \sim N\left(\sum_{i=1}^{a} c_i \mu_i, \sqrt{\sum_{i=1}^{a} c_i^2 \frac{\sigma^2}{n_i}}\right).$$

Scheffé's method

Using Scheffé's method we will get the intervals

$$I_{\sum_{i=1}^{a} c_i\mu_i} = \left(\sum_{i=1}^{a} c_i \bar{y}_{i\cdot} \mp s\sqrt{\sum_{i=1}^{a} \frac{c_i^2}{n_i}} \sqrt{(a-1)F_\alpha(a-1, N-a)} \right)$$

with simultaneous confidence level at least $1-\alpha$.



Note that estimator $\sum_{i=1}^{a} c_i \overline{Y}_{i\cdot}$ have a estimated standard deviation

$$d\left(\sum_{i=1}^{a} c_i \overline{Y}_{i\cdot}\right) = s\sqrt{\sum_{i=1}^{a} \frac{c_i^2}{n_i}}$$

which directly influences length of the interval.

Learn about the advantages of the book. Disadvantage: long intervals.

Remark: Normally you have to choose the interesting contrasts **before** one see measurements/observations, but it is not needed for the Scheffé's method.

If one is only interested in the **single** interval, one can to t-interval in the classical way.

Pairwise comparisons

One is often interested in the pairwise comparisons, i.e., construct confidence intervals for the differences $\mu_i - \mu_j$.

Point estimator: $\widehat{\mu}_i - \widehat{\mu}_j = \bar{y}_{i\cdot} - \bar{y}_{j\cdot}$ observation from

$$\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot} \sim N\left(\mu_i - \mu_j, \sqrt{\frac{\sigma^2}{n_i} + \frac{\sigma^2}{n_j}}\right) \text{ since}$$

$$\overline{Y}_{i\cdot} = \frac{1}{n_i} \sum_{k=1}^{n_i} Y_{ik} \text{ and } Y_{ik} \sim N(\mu_i, \sigma) \Rightarrow \overline{Y}_{i\cdot} \sim N\left(\mu_i, \frac{\sigma}{\sqrt{n_i}}\right),$$

$$E\left(\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot}\right) = E\left(\overline{Y}_{i\cdot}\right) - E\left(\overline{Y}_{j\cdot}\right) = \mu_i - \mu_j \quad \text{and}$$

$$Var\left(\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot}\right) = Var\left(\overline{Y}_{i\cdot}\right) + Var\left(\overline{Y}_{j\cdot}\right) = \frac{\sigma^2}{n_i} + \frac{\sigma^2}{n_j}.$$

Method 1: t-interval

All pairwise differences $\mu_i - \mu_j$, i.e., $\binom{a}{2}$ intervals.

We have as before σ^2-estimate $s^2 = \dfrac{SS_E}{N-a}$.

Method 1: t-interval with Bonferroni-estimation of confidence level.

Consider the statistic $$\frac{\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot} - (\mu_i - \mu_j)}{S\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \sim t(N-a).$$

$$\Rightarrow \quad I_{\mu_i - \mu_j} = \left(\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \mp t_{1-\frac{\alpha}{2\binom{a}{2}}}(N-a)s\sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \right).$$

Then, the simultaneous confidence level is $\geq 1 - \alpha$.

Method 2: Scheffé's

We can also adjust Scheffé's method to the pairwise comparisons.

Method 2: Scheffé's method give the intervals

$$I_{\mu_i-\mu_j} = \left(\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \mp s\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}\sqrt{(a-1)F_\alpha(a-1, N-a)}\right)$$

with simultaneous confidence level $\geq 1-\alpha$.

Note that for the method 1 och method 2 we have

$$d\left(\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot}\right) = s\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}$$

which directly influences the length of the interval.

Tukey-distribution

We turn now to discuss *Tukeys method for pairwise comparisons*. Then we must assume that $n_1 = n_2 = \ldots = n_a = n$, i.e., balanced design. The method is based on

Tukey-distribution (Studentized range). Let $\xi_1, \ldots, \xi_a$ and η be independent r.v. such that $\xi_i \sim N(0, 1)$ for $i = 1, \ldots, a$ and $\eta \sim \chi^2(f)$. Put

$$\zeta = \frac{\max_{i<j} |\xi_i - \xi_j|}{\sqrt{\eta/f}}.$$

Then, ζ have Tukey-distribution with parameters a and f.

Compare with Gossets theorem and t-distribution.

Tukeyes method for pairwise comparisons of μ_i is contrstructed from the r.v.

$$\max_{i<j}\left|\frac{\overline{Y}_{i\cdot}-\overline{Y}_{j\cdot}-(\mu_i-\mu_j)}{S/\sqrt{n}}\right| = \frac{\max_{i<j}\left|\frac{\overline{Y}_{i\cdot}-\mu_i}{\sigma/\sqrt{n}}-\frac{\overline{Y}_{j\cdot}-\mu_j}{\sigma/\sqrt{n}}\right|}{S/\sigma}$$

that have Tukey-distribution and that depends only on the parameters

$$\begin{aligned} a = {} & \text{ amount of } \xi_i \quad (\text{here, amount of } \overline{Y}_{i\cdot}) \\ f = {} & \text{ degrees of freedom for } S \quad (\text{here, } N-a), \end{aligned}$$

because the r.v. $\dfrac{\overline{Y}_{i\cdot}-\mu_i}{\sigma/\sqrt{n}}$ are independent and $N(0,1)$ and even independent from the r.v. $\dfrac{S}{\sigma}=\sqrt{\dfrac{fS^2}{\sigma^2}/f}=\sqrt{\dfrac{SS_E}{\sigma^2}/f}$ where $\dfrac{SS_E}{\sigma^2}\sim\chi^2(f)$; $f=N-a$.

Method 3: Tukey

Using the Tukey-distributed statistic above and quantiles $q_\alpha(a, f)$, given in Tukey-table, we obtain

Method 3: Tukeys method gives the $\binom{a}{2}$ intervals

$$I_{\mu_i - \mu_j} = \left(\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \mp q_\alpha(a, f) \frac{s}{\sqrt{n}} \right)$$

with simultaneous confidence level **exact** $1 - \alpha$

Note that this is the estimated standard deviation for **one** estimated value, i.e.,

$$d\left(\overline{Y}_{i\cdot}\right) = \frac{s}{\sqrt{n}},$$

that directly influences length of the interval.

Results follow from

$$1-\alpha = P\left(\max_{i<j}\left|\frac{\overline{Y}_{i\cdot}-\overline{Y}_{j\cdot}-(\mu_i-\mu_j)}{S/\sqrt{n}}\right| < q_\alpha(a,f)\right)$$

$$= P\Bigg(-q_\alpha(a,f) < \frac{\overline{Y}_{i\cdot}-\overline{Y}_{j\cdot}-(\mu_i-\mu_j)}{S/\sqrt{n}} < q_\alpha(a,f)$$

$$\text{for all } (i,j) : i<j)\Bigg)$$

$$= P\Bigg(\overline{Y}_{i\cdot}-\overline{Y}_{j\cdot}-q_\alpha(a,f)\frac{S}{\sqrt{n}} < \mu_i-\mu_j < \overline{Y}_{i\cdot}-\overline{Y}_{j\cdot}+q_\alpha(a,f)\frac{S}{\sqrt{n}}$$

$$\text{for all } (i,j) : i<j\Bigg)$$

Method 3': Tukey-Kramers

Remark Tukeys method demands the same sample size for all treatments. For the samples with different sample size we need modification of the method given below.

Method 3': Tukey-Kramers method gives the intervals

$$I_{\mu_i - \mu_j} = \left(\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \mp \frac{q_\alpha(a, f)}{\sqrt{2}} \cdot s\sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \right)$$

and have simultaneous confidence level approximately $1 - \alpha$.

Note that interval length again is related to

$$d\left(\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot}\right) = s\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}.$$

Should not be used if the sample sizes are too different!

What to choose?

- *Equal sample sizes:* Choose Tukeys method if you are interested in **all** intervals and want a high simultaneous confidence level.

Tukey values are given in table for $\alpha = 10\%, 5\%, 1\%$.

Choose *t*-interval with Bonferroni-estimation of confidence level if you could have a lower simultaneous confidence level or if you are only interested in some pairwise comparisons.

Note that one have to decide which pairs to consider **before** looking at the data.

- *Different sample sizes:* Choose *t*-intervals with Bonferroni-estimation of confidence level, Scheffé interval or if the sample sizes do not differ too much Tukey-Kramers method.

Example - Strontium concentrations from Lecture 2

We do pairwise comparisons of μ_i and choose Tukeys method with simultaneous confidence level 95%.

One-way ANOVA: C6 versus C7

```
Source  DF       SS      MS      F      P
C7       4  2193.44  548.36  56.15  0.000
Error   25   244.13    9.77
Total   29  2437.57

S = 3.125   R-Sq = 89.98%   R-Sq(adj) = 88.38%

                         Individual 95% CIs For Mean Based on
                         Pooled StDev
Level  N    Mean  StDev  ---+---------+---------+---------+------
1      6  32.083  3.205  (--*--)
2      6  40.233  2.530            (--*---)
3      6  44.083  3.081                 (--*--)
4      6  41.100  3.666             (--*---)
5      6  58.300  3.036                                 (--*--)
                         ---+---------+---------+---------+------
                         32.0      40.0      48.0      56.0

Pooled StDev = 3.125
```

We have σ^2-estimate $s^2 = \dfrac{SS_E}{25} = 9.7652$, i.e., $s = 3.125$ with degrees of freedom: 25.

$$
\begin{aligned}
I_{\mu_i - \mu_j} &= \left(\bar{y}_i - \bar{y}_j \mp q_{0.05}(5, 25) \cdot \frac{s}{\sqrt{6}}\right) \\
&= \left(\bar{y}_i - \bar{y}_j \mp 4.16 \cdot \frac{3.125}{\sqrt{6}}\right) = \left(\bar{y}_i - \bar{y}_j \mp 5.307\right).
\end{aligned}
$$

We have

$$
\bar{y}_{1\cdot} = 32.083 \quad \bar{y}_{2\cdot} = 40.233 \; \bar{y}_{4\cdot} = 41 \cdot 100 \; \bar{y}_{3\cdot} = 44.083 \quad \bar{y}_{5\cdot} = 58.300
$$

Water 1 have significantly lower strontium content than the others, and 5 have significantly higher strontium than the others.

There are no significant differences between 2, 4 and 3.

Examples - Sheet Metal Manufacturing, cont.

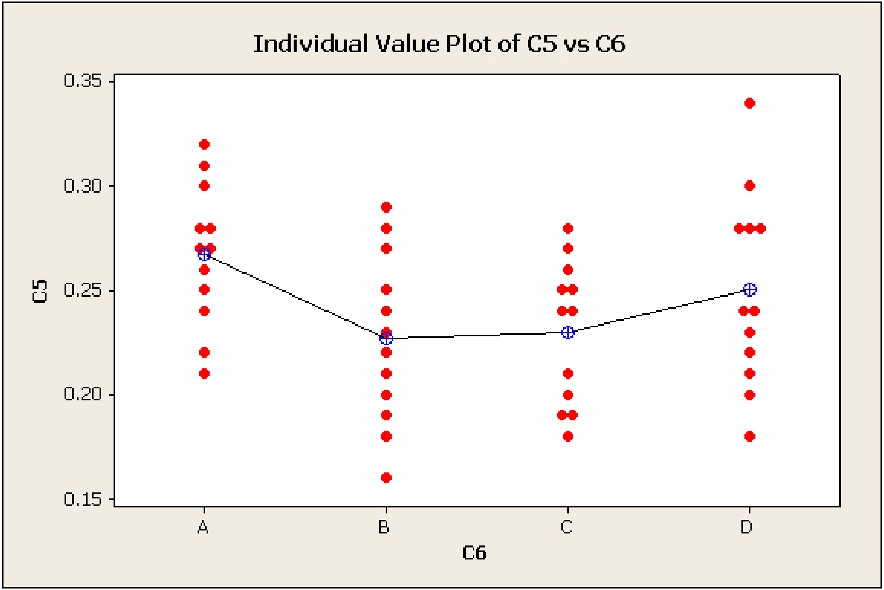
We have $\bar{y}_{1\cdot} = 0.2675$, $\bar{y}_{2\cdot} = 0.2267$, $\bar{y}_{3\cdot} = 0.2300$, $\bar{y}_{4\cdot} = 0.2500$ and

$s^2 = \dfrac{SS_E}{44} = 0.001543$, $s = 0.0393$ with 44 degrees of freedom.

We want to do pairwise comparisons between labs and we choose simultaneous confidence level to be approximately 95%.

The largest difference between $\bar{y}_{i\cdot}$ and $\bar{y}_{j\cdot}$ is for

$$\bar{y}_{1\cdot} - \bar{y}_{2\cdot} = 0.2675 - 0.2267 = 0.0408$$


Individual Value Plot of C5 vs C6
0.35
0.30
0.25
0.20
0.15
C5
A
B
C
D
C6

One-way ANOVA: C5 versus C6

```
Source  DF       SS       MS     F      P
C6       3  0.01301  0.00434  2.81  0.050
Error   44  0.06789  0.00154
Total   47  0.08090

S = 0.03928   R-Sq = 16.08%   R-Sq(adj) = 10.36%

                                 Individual 95% CIs For Mean Based on
                                 Pooled StDev
Level   N     Mean    StDev   --------+---------+---------+---------+-
A      12  0.26750  0.03388                  (--------*--------)
B      12  0.22667  0.04097   (--------*--------)
C      12  0.23000  0.03438    (--------*--------)
D      12  0.25000  0.04651           (--------*--------)
                              --------+---------+---------+---------+-
                                    0.225     0.250     0.275     0.300

Pooled StDev = 0.03928
```

Let us construct t-intervals.

There is $\binom{4}{2} = 6$ comparisons. We choose to have 99% confidence level for each of the intervals

$$I_{\mu_i-\mu_j} = \left(\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \mp 2.69 \cdot s\sqrt{\frac{1}{12} + \frac{1}{12}}\right)$$
$$= (\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \mp 1.098s) = (\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \mp 0.0432),$$

where $t = 2.69$ is given in $t(44)$-table under condition $F(t) = 0.995$.

Then, the simultaneous confidence level is $\geq 94\%$.

We can also construct Scheffé intervals as

$$I_{\mu_i-\mu_j} = \left(\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \mp s\sqrt{\frac{1}{12}+\frac{1}{12}} \cdot \sqrt{(4-1)\cdot F_{0.05}(3,44)}\right)$$
$$= (\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \mp 1.187s) = (\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \mp 0.0466),$$

where $F_{0.05}(3,44) = 2.82$ with the simultaneous confidence level $\geq 95\%$,

or Tukey-intervals

$$I_{\mu_i-\mu_j} = \left(\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \mp \frac{s}{\sqrt{12}} \cdot 3.78\right)$$
$$= (\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \mp 1.091s) = (\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \mp 0.0429)$$

with $3.78 = q_{0.05}(4,44)$ and the simultaneous confidence level is exactly 95%.

No significant differences can be found!

```
Tukey 95% Simultaneous Confidence Intervals
All Pairwise Comparisons among Levels of C6

Individual confidence level = 98.95%

C6 = A subtracted from:

C6     Lower    Center    Upper    -+---------+---------+---------+--------
B   -0.08370  -0.04083  0.00203    (----------*----------)
C   -0.08036  -0.03750  0.00536     (----------*---------)
D   -0.06036  -0.01750  0.02536          (----------*---------)
                                   -+---------+---------+---------+--------
                                 -0.080    -0.040     0.000     0.040
C6 = B subtracted from:

C6     Lower   Center    Upper    -+---------+---------+---------+--------
C   -0.03953  0.00333  0.04620              (----------*----------)
D   -0.01953  0.02333  0.06620                   (----------*----------)
                                  -+---------+---------+---------+--------
                                -0.080    -0.040     0.000     0.040
C6 = C subtracted from:

C6     Lower   Center    Upper    -+---------+---------+---------+--------
D   -0.02286  0.02000  0.06286                  (----------*----------)
                                  -+---------+---------+---------+--------
                                -0.080    -0.040     0.000     0.040
```

Linköping University - Research that makes a difference

LINKÖPING
UNIVERSITY