

Experimental Design and Statistical Methods



Workshop

RANDOMIZED COMPLETE BLOCK DESIGNS and LATIN SQUARES

Jesús Piedrafita Arilla

jesus.piedrafita@uab.cat

Departament de Ciència Animal i dels Aliments



Items

- Randomized complete block design
 - Concept of blocking
- Latin Squares design
- Basic commands
 - lm

Blocking: An example with penicillin yield

Remember that blocking is one of the three principles of experimental design. It has as a goal to reduce nuisance effects and to make “signal” stronger in relation to residual variation.

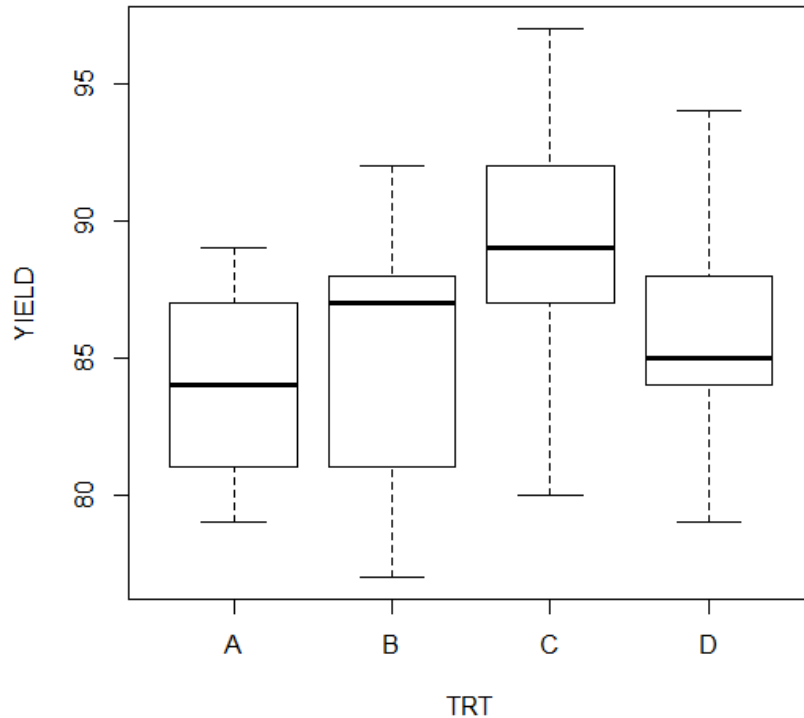
We illustrate this with an experiment to compare 4 processes, A, B, C, D, for the production of penicillin. These are the **treatments**. The raw material, corn steep liquor, is quite variable and can only be made in blends (**blocks**) sufficient for 4 runs.

| | A | B | C | D |
|---------|----|----|----|----|
| Blend 1 | 89 | 88 | 97 | 94 |
| Blend 2 | 84 | 77 | 92 | 79 |
| Blend 3 | 81 | 87 | 87 | 85 |
| Blend 4 | 87 | 92 | 89 | 84 |
| Blend 5 | 79 | 81 | 80 | 88 |

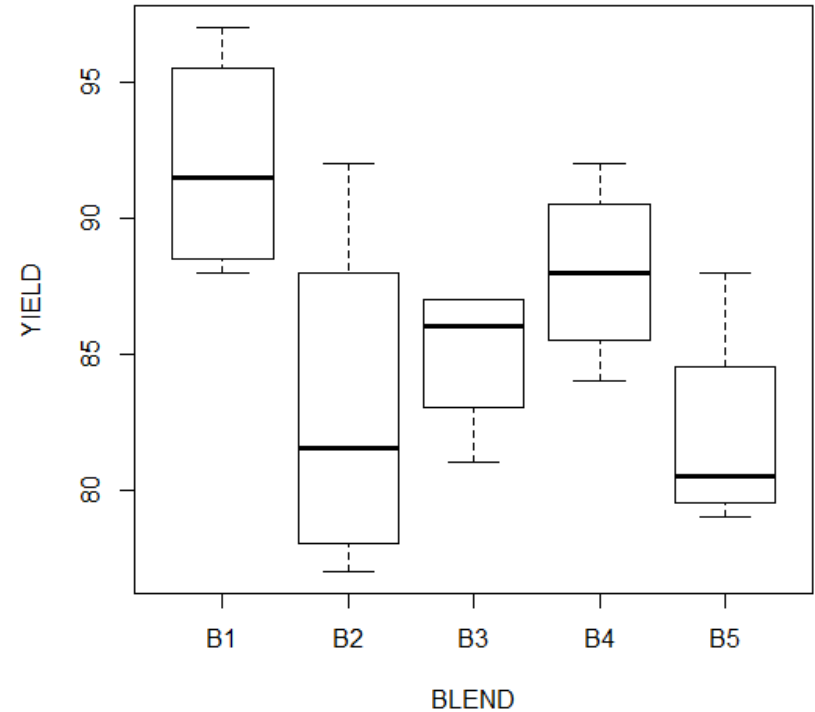
We can analyse the data ignoring the blend effect, as a CRD (one way anova), but the data suggest to analyse them by a **Randomized complete block design (RCBD)**.

A couple of box-plots

```
> plot(YIELD~TRT)
```



```
> plot(YIELD~BLEND)
```



Treatment seems not to have effect on the yield of penicillin, although block (blend) does.

First analysis (CRD) in R

We ignore the blend effect and analyse the data with a one-way ANOVA.

```
> PENICILLIN.AOV<-aov(YIELD~TRT, PENICILLIN)
> anova(PENICILLIN.AOV)
> summary.lm(aov(YIELD~TRT))
```

Analysis of Variance Table

Response: YIELD

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) | |
|-----------|----|--------|---------|---------|--------|-----------------|
| TRT | 3 | 70 | 23.333 | 0.7619 | 0.5318 | Not significant |
| Residuals | 16 | 490 | 30.625 | | | |

...

Residual standard error: 5.534 on 16 degrees of freedom

Multiple R-squared: 0.125, Adjusted R-squared: -0.03906

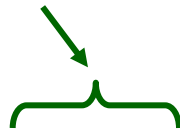
F-statistic: 0.7619 on 3 and 16 DF, p-value: 0.5318

Second analysis: Randomized Complete Block Design

In this case we take into account the effect of the blend (φ): a **BLOCK**. The model is extended as follows:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

In CRD



$$y_{ijk} = \mu + \tau_i + \varphi_j + \varepsilon'_{ijk}$$

In RCBD. The error term of the previous model is partitioned into the blend effect and a new error (probably lower \Rightarrow lower mean square error)

Furthermore, the blend effect can be considered a **random effect**, as we are not interested in the effect of each particular blend, but we have taken these blends at random, only for practical reasons. Blocking aims at reducing **noise** (i.e., the residual mean square error).

RCBD in R - (1) -

```
> PENICILLIN.AOV.RB<-aov(YIELD~TRT+BLEND, PENICILLIN)
> anova(PENICILLIN.AOV.RB)
> summary.lm(aov(YIELD~TRT+BLEND))
```

Analysis of Variance Table

Response: YIELD

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|--------|---------|---------|-----------|
| TRT | 3 | 70 | 23.333 | 1.2389 | 0.33866 |
| BLEND | 4 | 264 | 66.000 | 3.5044 | 0.04075 * |
| Residuals | 12 | 226 | 18.833 | | |

...

Residual standard error: 4.34 on 12 degrees of freedom

Multiple R-squared: 0.5964 Adjusted R-squared: 0.361

F-statistic: 2.534 on 7 and 12 DF, p-value: 0.07535

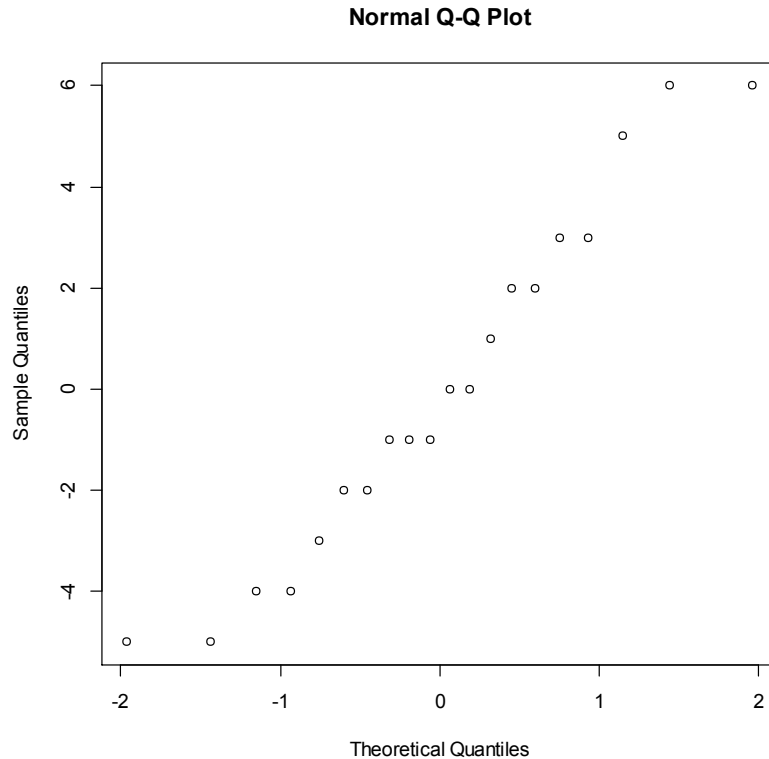
The model explains now 59.64% of the total variability, and is statistically significant.

The *F*-statistic corresponds to fitting the full model.

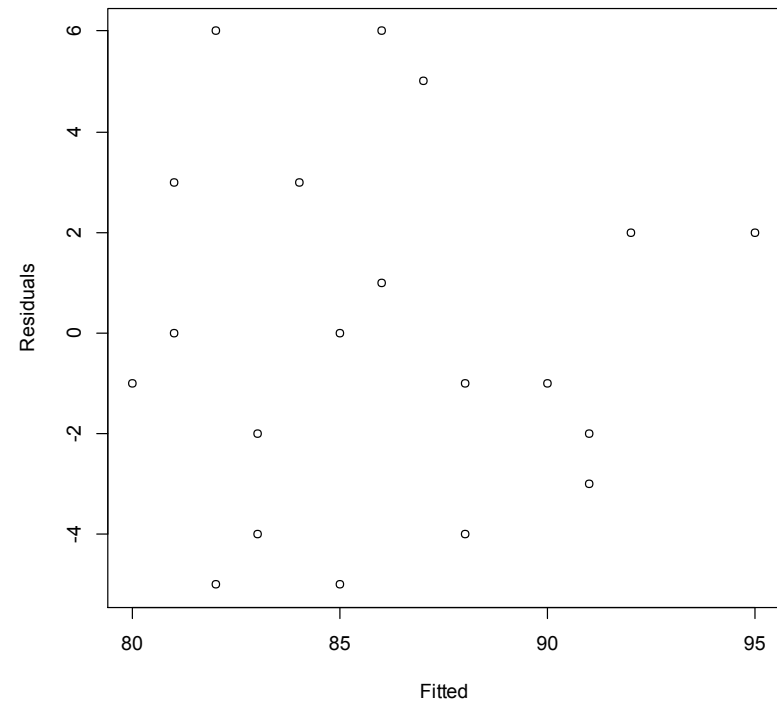
Treatment Sum Sq is equal to that in the previous model, but Error Sum Sq is much lower. In fact the sum of BLEND and Residuals Sum Sq equals Residual Sum Sq in the previous model. This is not enough to render treatment significant. Blend effect, the blocking factor, is significant.

RCBD in R - (2) -

```
> qqnorm(PENICILLIN.AOV.RB$res)
> plot(PENICILLIN.AOV.RB$fitted, PENICILLIN.AOV.RB$res,
> + xlab="Fitted", ylab="Residuals")
```



No evident deviation from normality
in the residuals



Fitted values and residuals are
independent

RCBD in R - (3) -

```
> library(multcomp)
> summary(glht(PENICILLIN.AOV.RB, linfct=mcp (TRT="Tukey")))
```

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = YIELD ~ TRT + BLEND, data = PENICILLIN)

Linear Hypotheses:

| | Estimate | Std. Error | t value | Pr(> t) |
|------------|----------|------------|---------|----------|
| B - A == 0 | 1.000 | 2.745 | 0.364 | 0.983 |
| C - A == 0 | 5.000 | 2.745 | 1.822 | 0.310 |
| D - A == 0 | 2.000 | 2.745 | 0.729 | 0.884 |
| C - B == 0 | 4.000 | 2.745 | 1.457 | 0.491 |
| D - B == 0 | 1.000 | 2.745 | 0.364 | 0.983 |
| D - C == 0 | -3.000 | 2.745 | -1.093 | 0.700 |

(Adjusted p values reported -- single-step method)

As expected, no differences among treatment means were significant.

RCBD in R - (3b) -

```
> library(agricolae)
> HSD.test(PENICILLIN.AOV.RB, "TRT")
```

```
TRT,  means
      YIELD  std.err  r Min. Max.
A      84 1.843909  5    79    89
B      85 2.664583  5    77    92
C      89 2.810694  5    80    97
D      86 2.469818  5    79    94
```

Means with the same letter are not significantly different.

```
Groups, Treatments and means
a      C      89
a      D      86
a      B      85
a      A      84
```

As expected, no differences among treatment means were significant.

RCBD in R - (4) -

The analysis of variance table corresponds to a sequential testing of models, here corresponding to the sequence:

YIELD ~ 1

YIELD ~ TRT

YIELD ~ TRT + BLEND

So here the p -value 0.339 corresponds to a comparison of the first two models in this list, while the p -value of 0.041 corresponds to the test comparing the second two models. One small point to note is that the denominator in both F -tests is the mean square from the full model, here 18.8.

Notice that if we change the order of the terms in the ANOVA, it makes no difference because of the orthogonal design (this can be checked by the student). But this does not hold when some observation is missing.

Four SS Types

- I. Fully sequential, it depends on the order in which even main effects are listed in the definition of the model.
Useful for nested hierarchical models and polynomial regression models.
- II. Method of fitting constants - SS for a given effect is adjusted for all effects defined in the model that do not contain the given effect.
To be used for main effects models and pure regression models.
- III. Weighted squares of means - SS for a given effect is adjusted for all other effects in the model, regardless of whether they contain the given effect or not. In particular A and B main effects will be adjusted for the A*B interaction.
In designs where all cells are filled, it allows a comparison of main effects even in the presence of interaction.
- IV. Goodnight's method for missing cells layouts - similar to Type III in spirit, with a different strategy for compensating for missing cells when estimating the model parameters.

For balanced designs, all SS types are equal.

R gives by default SS I. To get any particular SS you must write **type="III"** or any other in the model definition.

RCBD in R - (5) -

```
> anova(aov(YIELD~BLEND+TRT,PENICILLIN[-1,]))
```

Analysis of Variance Table

Response: YIELD

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|---------|---------|---------|-----------|
| BLEND | 4 | 266.526 | 66.632 | 3.2672 | 0.05373 . |
| TRT | 3 | 59.667 | 19.889 | 0.9752 | 0.43928 |
| Residuals | 11 | 224.333 | 20.394 | | |

Observe how to
remove an
observation from
a data-frame

```
> anova(aov(YIELD~TRT+BLEND,PENICILLIN[-1,]))
```

Analysis of Variance Table

Response: YIELD

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|---------|---------|---------|-----------|
| TRT | 3 | 91.776 | 30.592 | 1.5001 | 0.26880 |
| BLEND | 4 | 234.417 | 58.604 | 2.8736 | 0.07451 . |
| Residuals | 11 | 224.333 | 20.394 | | |

Notice that now the order does matter. If we want to test for a treatment effect, we would prefer the first table since in that version the blocking factor blend is already included when we test the treatment factor. Since the blocking factor is an unalterable feature of the chosen design, this is as it should be.

Relative advantage of RCBD over CRD

We can compare the residual variance for designs with the same sample size. The *relative efficiency* in our case is:

$$\frac{\hat{\sigma}_{CRD}^2}{\hat{\sigma}_{RCBD}^2} = \frac{30.6}{18.8} = 1.62$$

The interpretation is that a CRD would require 62% more observations to obtain the same level of precision as a RCBD. The efficiency is not guaranteed to be greater than one. Only use blocking where there is some heterogeneity in the experimental units. The decision to block is a matter of judgment prior to the experiment. There is no guarantee that it will increase precision.

Unbalanced designs

Until now we have analysed balanced designs, but some data are frequently lost along the experiment, either because some animals dead during the experiment or because some analytical problems, for example. In these cases we have more or less **unbalanced designs**.

Unbalanced designs are also usually found when we analyse field data, for example the factors influencing weaning weight in beef cattle, like age of the mother, single/twin birth, season of calving, ...

People interested in this topic can resort to the book of Milliken and Johnson (1994), *Analysis of Messy Data*, Vol 1., Chapman Press, New York.

Latin squares (1)

These are useful when there are **two blocking factors**.

For example, in an industrial experiment, suppose we wish to compare 4 production methods (the treatment) — A, B, C, and D. We have available 4 machines 1, 2, 3, and 4, and 4 operators, I, II, III, IV. A Latin square design is:

| | 1 | 2 | 3 | 4 |
|-----|---|---|---|---|
| I | A | B | C | D |
| II | B | D | A | C |
| III | C | A | D | B |
| IV | D | C | B | A |

The number of possible Latin squares increases rapidly with the number of treatments. It is equal to $k * k! * (k - 1)!$, where k is the number of treatments.

Ex: for $k = 4$, the number of total possible squares is: $(4) (4!) (3!) = 576$

Latin squares (2)

Often in Latin squares designs in Animal Science, rows correspond to periods and columns correspond to animals. In this particular situation, where there is a succession of treatments, the risk of carryover effects exists, and particular arrangements of the Latin square are recommended so each treatment follows each other treatment once.

| | | | |
|---|---|---|---|
| D | C | B | A |
| C | D | A | B |
| B | A | D | C |
| A | B | C | D |

| | | | |
|---|---|---|---|
| D | C | B | A |
| C | A | D | B |
| B | D | A | C |
| A | B | C | D |

- (Left) $C \rightarrow D$ twice, $A \rightarrow D$ once, $B \rightarrow D$ never
- (Right) Each trt follows each other trt once

Some general rules

- Single 2×2 Latin squares are not sufficiently powerful to detect statistical differences due to the very few degrees of freedom.
 - 3×3 Latin squares have similar problems to 2×2 squares, with the additional limitation that it is not possible to have each treatment follow every other treatment, thereby creating the risk of unequal treatment carryover if period is part of the design.
 - **4×4 designs are the most powerful**, as they can be set up to ensure that all treatments follow all other treatments in order to equalize treatment carryover effects.
- ➔ **Use only 4×4 designs (if possible)**

Latin squares – the model -

$$y_{ij(k)} = \mu + row_i + col_j + \tau_{(k)} + e_{ij(k)}$$

The parenthesis around k indicates that each k appears only once in each row and each column.

The **null hypothesis** and the **alternative hypothesis** can be stated as:

$H_0: \tau_1, \tau_2, \dots, \tau_k$, no treatment effects.

$H_1: \tau_i \neq \tau_{i'}$, for at least one pair (i, i') , there is a difference between treatments.

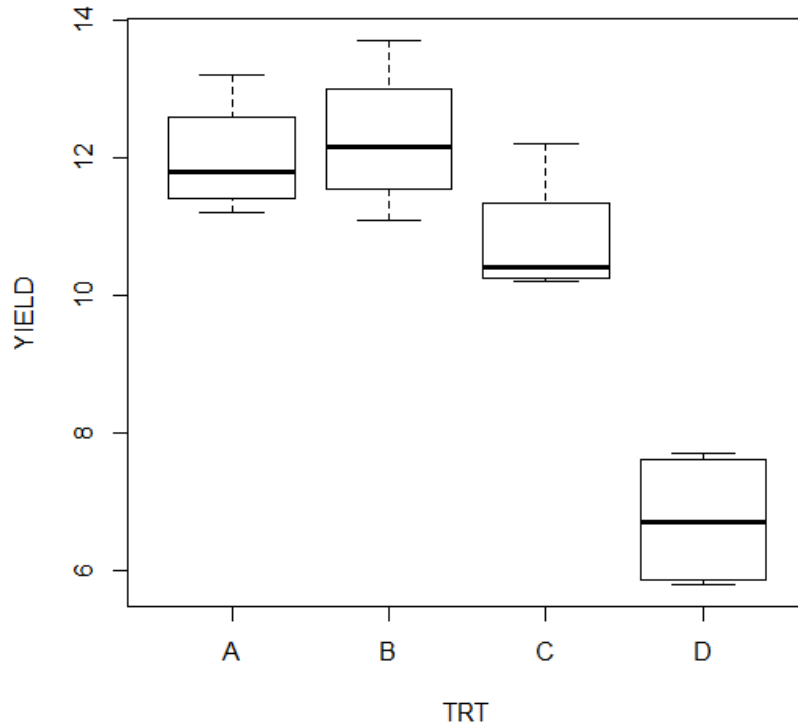
A LS numerical example

| | OPER 1 | OPER 2 | OPER 3 | OPER 4 |
|-------|---------|---------|---------|---------|
| DAY 1 | C, 10.5 | D, 7.7 | B, 12.0 | A, 13.2 |
| DAY 2 | B, 11.1 | A, 12.0 | C, 10.3 | D, 7.5 |
| DAY 3 | D, 5.8 | C, 12.2 | A, 11.2 | B, 13.7 |
| DAY 4 | A, 11.6 | B, 12.3 | D, 5.9 | C, 10.2 |

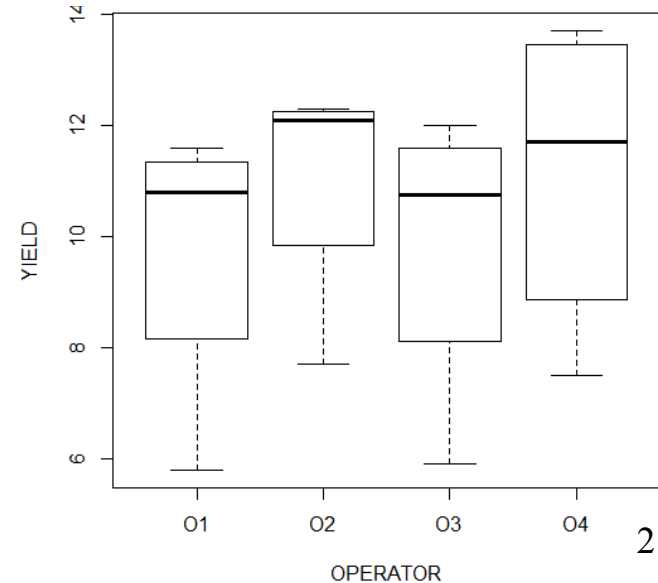
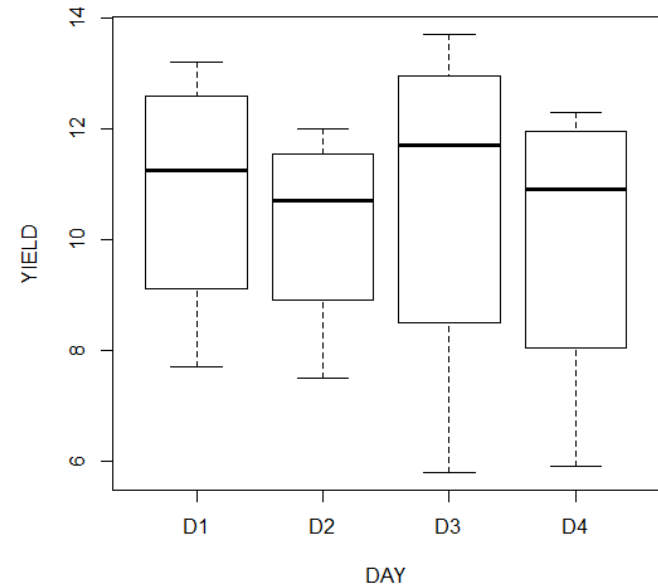
Rows correspond to days, columns to operator, and the treatments (A, B, C, D) are indicated in green capital letters. The numbers correspond to yield.

A LS numerical example –box-plots-

```
> plot(YIELD ~ DAY, LATINSQ)  
> plot(YIELD ~ OPERATOR, LATINSQ)  
> plot(YIELD ~ TRT, LATINSQ)
```



No obvious violations of normality nor homogeneity of variances.
Treatment D could be different from the other three treatments.



A LS numerical example – ANOVA 1 -

```
> LATINSQ.AOV <- lm(YIELD ~ DAY + OPERATOR + TRT, LATINSQ)
> anova(LATINSQ.AOV)
```

Analysis of Variance Table

Response: YIELD

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|--------|---------|---------|---------------|
| DAY | 3 | 1.955 | 0.6517 | 1.4375 | 0.3219 |
| OPERATOR | 3 | 6.800 | 2.2667 | 5.0000 | 0.0452 * |
| TRT | 3 | 78.925 | 26.3083 | 58.0331 | 7.987e-05 *** |
| Residuals | 6 | 2.720 | 0.4533 | | |

Observe that all Mean Sq are contrasted against the Residuals Mean Sq.

As expected, significant differences among treatment means exist. Operator effect was also significant.

A LS numerical example – ANOVA 2 -

```
> summary(LATINSQ.AOV)
```

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) | |
|-------------|----------|------------|---------|----------|-----|
| (Intercept) | 11.7000 | 0.5323 | 21.980 | 5.79e-07 | *** |
| DAYD2 | -0.6250 | 0.4761 | -1.313 | 0.2372 | |
| DAYD3 | -0.1250 | 0.4761 | -0.263 | 0.8017 | |
| DAYD4 | -0.8500 | 0.4761 | -1.785 | 0.1245 | |
| OPERATORO2 | 1.3000 | 0.4761 | 2.731 | 0.0342 | * |
| OPERATORO3 | 0.1000 | 0.4761 | 0.210 | 0.8406 | |
| OPERATORO4 | 1.4000 | 0.4761 | 2.941 | 0.0259 | * |
| TRTB | 0.2750 | 0.4761 | 0.578 | 0.5845 | |
| TRTC | -1.2000 | 0.4761 | -2.521 | 0.0453 | * |
| TRTD | -5.2750 | 0.4761 | -11.080 | 3.22e-05 | *** |

Residual standard error: 0.6733 on 6 degrees of freedom

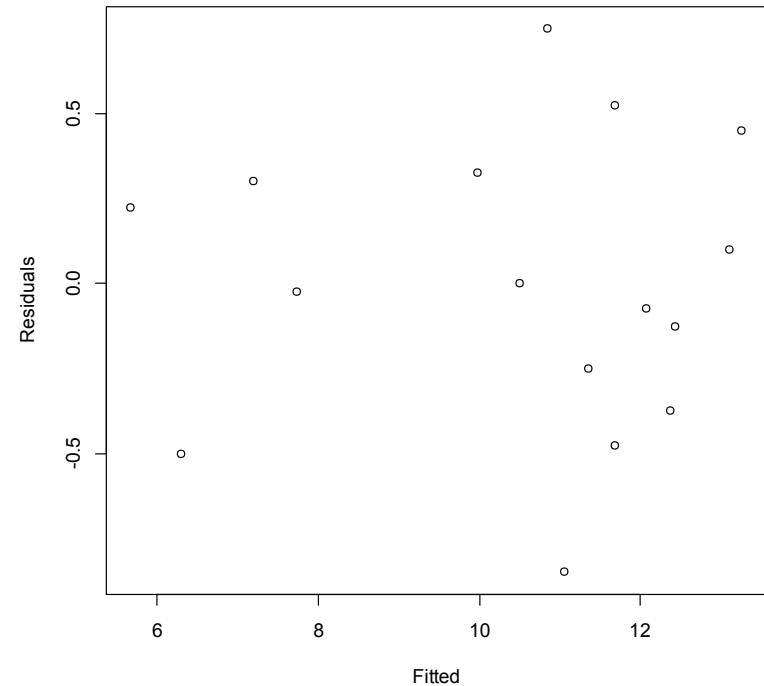
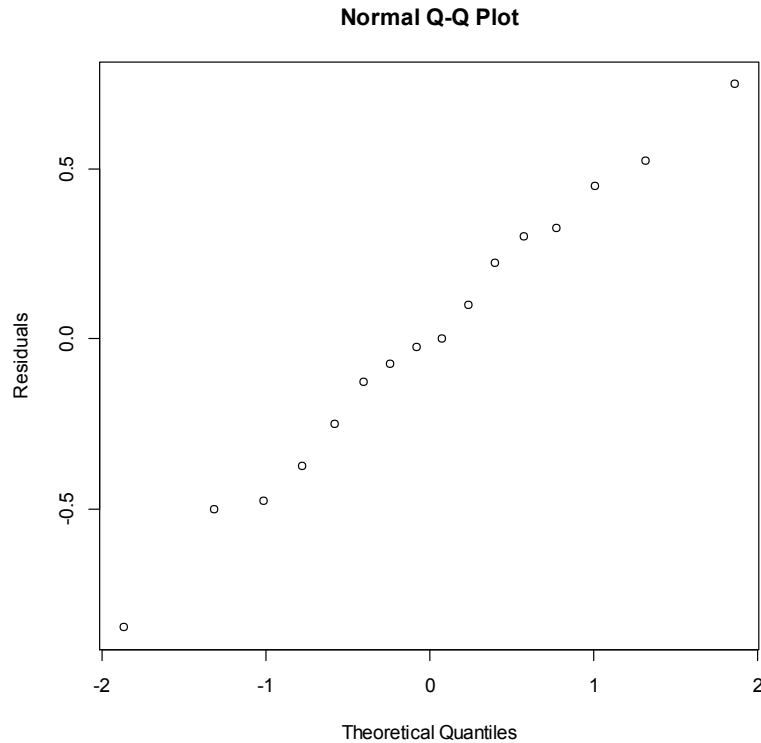
Multiple R-squared: 0.9699, Adjusted R-squared: 0.9248

F-statistic: 21.49 on 9 and 6 DF, p-value: 0.0006744

The model fits extremely well the data.

A LS numerical example – Diagnostics -

```
> plot(LATINSQ.AOV$fit,LATINSQ.AOV$res,xlab="Fitted",ylab="Residuals")  
> qqnorm(LATINSQ.AOV$res,ylab="Residuals")
```



Residuals do not deviate from normality and are independent of fitted values.

A LS numerical example – Tukey test (a) -

```
> tapply(YIELD,list(TRT),mean)
```

| | A | B | C | D |
|--|--------|--------|--------|-------|
| | 12.000 | 12.275 | 10.800 | 6.725 |

```
> library(multcomp)
```

```
> summary(glht(LATINSQ.AOV, linfct=mcp (TRT="Tukey")))
```

Multiple Comparisons of Means: Tukey Contrasts

Linear Hypotheses:

| | Estimate | Std. Error | t value | Pr(> t) | |
|------------|----------|------------|---------|----------|-----|
| B - A == 0 | 0.2750 | 0.4761 | 0.578 | 0.9353 | |
| C - A == 0 | -1.2000 | 0.4761 | -2.521 | 0.1532 | |
| D - A == 0 | -5.2750 | 0.4761 | -11.080 | <0.001 | *** |
| C - B == 0 | -1.4750 | 0.4761 | -3.098 | 0.0767 | . |
| D - B == 0 | -5.5500 | 0.4761 | -11.657 | <0.001 | *** |
| D - C == 0 | -4.0750 | 0.4761 | -8.559 | <0.001 | *** |

Treatment D differs significantly from the other treatments. The difference between C and B approached statistical significance.

A LS numerical example – Tukey test (b) -

```
> library(agricolae)
> HSD.test(LATINSQ.AOV, "TRT")
```

```
TRT, means
      YIELD    std.err r Min. Max.
A 12.000 0.4320494 4 11.2 13.2
B 12.275 0.5390965 4 11.1 13.7
C 10.800 0.4708149 4 10.2 12.2
D  6.725 0.5072393 4  5.8  7.7
```

Means with the same letter are not significantly different.

```
Groups, Treatments and means
a      B      12.28
a      A      12
a      C      10.8
b      D      6.725
```

Treatment D differs significantly from the other treatments.

Relative advantage of LS over CRD and RCBD

We can compare again the residual variance for designs with the same sample size. The *relative efficiency* in our case is:

$$\frac{\hat{\sigma}_{CRD}^2}{\hat{\sigma}_{LS}^2} = \frac{0.9562}{0.4533} = 2.11; \quad \frac{\hat{\sigma}_{RCBrow}^2}{\hat{\sigma}_{LS}^2} = \frac{1.0578}{0.4533} = 2.33; \quad \frac{\hat{\sigma}_{RCBcol}^2}{\hat{\sigma}_{LS}^2} = \frac{0.5194}{0.4533} = 1.14$$

The interpretation is that a CRD and RCBrow would require more than twice observations to obtain the same level of precision as a LS.

LS will be only a little bit more efficient (14%) than blocking for columns.

But remember that the efficiency is not always guaranteed to be greater than one.

Some comments on Latin Squares

Pros:

1. Latin squares design can be an **efficient way to use few animals** when their use is limited by cost, ethics, etc.
2. Several Latin squares can be combined in a unique analysis. This is out of the scope of this introductory course.

Disadvantages:

1. The **number of treatments must be equal to the number of rows and columns**. This restriction imposes a problem for actual experimental work, especially if there is a large number of treatments.
 2. The **design is only valid if there are no interactions among rows, columns, and treatments**. If there is any interaction no valid tests of significance are possible.
- ➔ To check this assumption there is a version of Tukey's test for non-additivity for Latin squares that is out of the aims of this course.