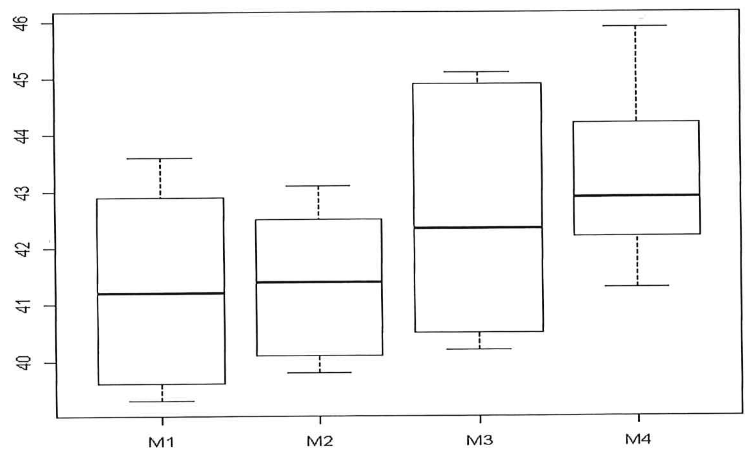
The machine example

Four different machines M_1, M_2, M_3 and M_4 are being considered for the assembling of a particular product. It was decided that six different operators would be used in an experiment to compare the machines. The response is the amount of time required to assemble a product. The operation of the machines requires physical dexterity, and it was anticipated that there would be a difference among the operators in the speed with which they operated the machines. This difference causes undesired variability in the measured response which would be of interest to eliminate. Therefore, it was decided to perform the experiment as a randomized block design, with operators as blocks. The experiments will now be randomized within blocks. The observed data are given below:

	Machines			
Operators	M_1	M_2	M_3	M_4
1	42.5	39.8	40.2	41.3
2	39.3	40.1	40.5	42.2
3	39.6	40.5	41.3	43.5
4	39.9	42.3	43.4	44.2
5	42.9	42.5	44.9	45.9
6	43.6	43.1	45.1	42.3

Boxplot maskiner



La $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_k \neq 0$ $\qquad$ H_1 at least one α_i different from zero.

Under H_0:

~~$\frac{SS_T}{\sigma^2} = \sum_{i=1}^{k}\sum_{j=1}^{b}\frac{(Y_{ij}-\bar{Y}_{..})^2}{\sigma^2} \sim \chi^2(n-1) \quad n = k\cdot b$~~

$$\frac{SS_A}{\sigma^2} = \sum_{i=1}^{k}\sum_{j=1}^{b}\frac{(\bar{Y}_{i.}-\bar{Y}_{..})^2}{\sigma^2} \sim \chi^2(k-1)$$

Det kan visast at $\frac{SS_E}{\sigma^2} = \sum_{i=1}^{k}\sum_{j=1}^{b}\frac{(Y_{ij}-\bar{Y}_{i.}-\bar{Y}_{.j}+\bar{Y}_{..})^2}{\sigma^2}$

er $\chi^2((k-1)(b-1))$ alltid

such that $F = \dfrac{\frac{SS_A}{\sigma^2(k-1)}}{\frac{SS_E}{\sigma^2(k-1)(b-1)}}$ er F fordelt med $(k-1)$ og $(k-1)(b-1)$ fridomsgrader.

Reject H_0 if. $F_{obs} \geq F_{\alpha,(k-1),(k-1)(b-1)}$

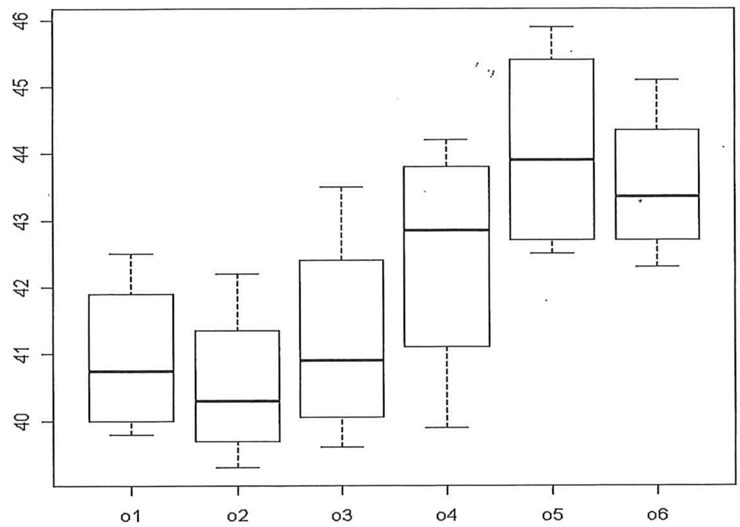
Variansanalyse tabellen

Kjelder til variasjon	SS	DF	MS	F
Behandlingar	$SS_A = b\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{y}_{..})^2$	$k-1$	$SSA/k-1$	$\frac{SSA}{k-1}\Big/\frac{SSE}{(k-1)(b-1)}$
Blokker	$SS_B = k\sum_{j=1}^{k}(\bar{y}_{.j}-\bar{y}_{..})^2$	$b-1$	$SSB/b-1$	
Feil	$SS_E = \sum_{i=1}^{k}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})^2$	$(k-1)(b-1)$	$\frac{SS_E}{(k-1)(b-1)}$	
Totalt	$SS_T = \sum_{i=1}^{k}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{..})^2$	$bk-1$		

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
Machine M1   41.3000     0.7413   55.72   <2e-16 ***
Machine M2   41.3833     0.7413   55.83   <2e-16 ***
Machine M3   42.5667     0.7413   57.42   <2e-16 ***
Machine M4   43.2333     0.7413   58.33   <2e-16 ***
```

```
           Df Sum Sq Mean Sq F value Pr(>F)
Machines    3 15.925  5.3082  1.6101 0.2186
Residuals  20 65.935  3.2968
```

Boxplot operators




```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
Operator o1  40.9500     0.7432   55.10   <2e-16 ***
Operator o2  40.5250     0.7432   54.52   <2e-16 ***
Operator o3  41.2250     0.7432   55.47   <2e-16 ***
Operator o4  42.4500     0.7432   57.12   <2e-16 ***
Operator o5  44.0500     0.7432   59.27   <2e-16 ***
Operator o6  43.5250     0.7432   58.56   <2e-16 ***
```

```
          Df Sum Sq Mean Sq F value   Pr(>F)
Machine    3 15.925  5.3082  3.3388 0.047904 *
Operator   5 42.087  8.4174  5.2944 0.005328 **
Residuals 15 23.848  1.5899
```

Suppose we neglected the blocking and analysed as a one-way Anova
We have :

$SS_E^{one-way} = SS_E^{block} + SS_B$ showing that by bloking

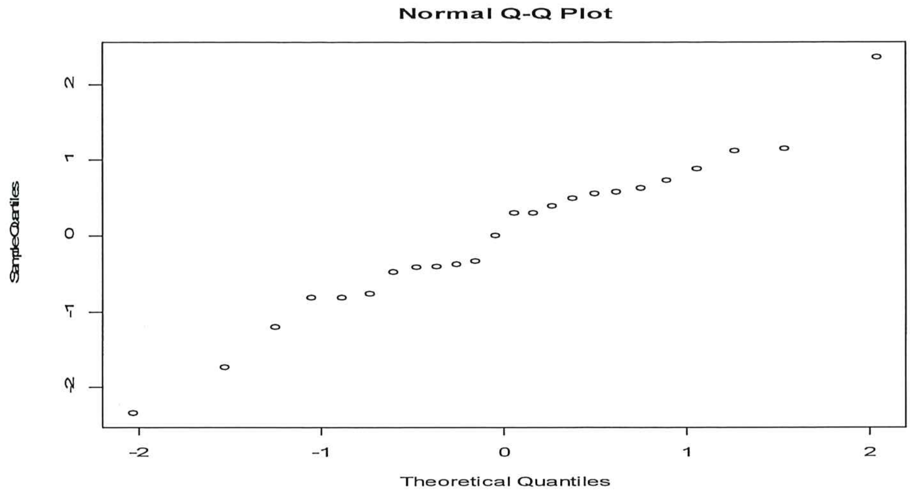
the experiment we may extract unwanted variation from the data at the expense of loosing some degrees of freedom

NB.

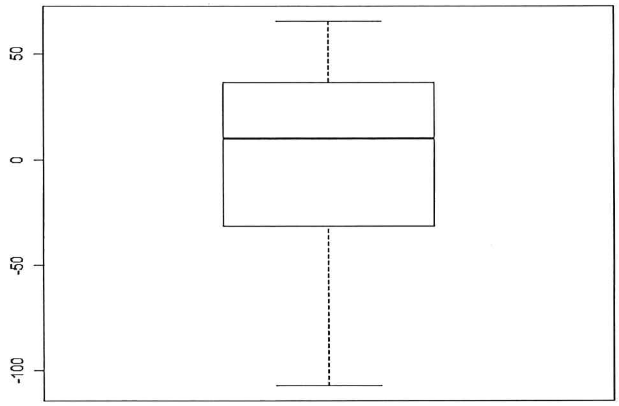
The residuals are given by $y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$

$= y_{ij} - (\underbrace{\bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..}}_{\hat{y}_{ij} = (\text{fitted model})})$.

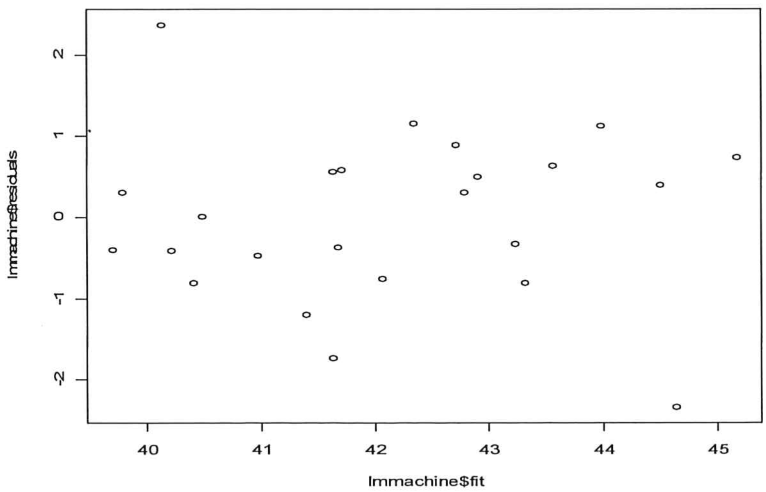
If residuals are suspected to be outliers they should be plotted against treatment and blocks.



Boxplot residuals



Residualplott



Two-factor analysis of variance with replication

Situation: We have two factors A and B. A has a levels and B has b levels. By doing replications it is possible to study interactions between the factors. The experiment should be completely randomized, i.e. all the abn experiments, where n is the number of replicates should be performed in random order.

Model: $Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$ $\Big\}$ $N(0, \sigma^2)$ and independent
$i = 1, 2, \ldots a$, $j = 1, 2, \ldots, b$
$k = 1, 2, \ldots, n$.

where α_i is the effect of the i-th level of factor A, β_j is the effect of the j-th level of factor B and $\alpha\beta_{ij}$ is the interaction between the i-th level of factor A and the j-th level of factor B.

Also $\sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \beta_j = \sum_{i=1}^{a} \alpha\beta_{ij} = \sum_{j=1}^{b} \alpha\beta_{ij} = 0$

The following hypothesis are of interest.

1. $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$ H_1: at least one is different from 0.

2. $H_0: \beta_1 = \beta_2 = \cdots = \beta_b = 0$ H_1: at least one is different from 0.

3. $H_0: \alpha\beta_{11} = \alpha\beta_{12} = \cdots = \alpha\beta_{ab} = 0$ H_1: at least one is different from 0.

Partitioning variation

We have $y_{ijk} = \bar{y}_{ij.} + y_{ijk} - \bar{y}_{ij.}$

and $\bar{y}_{ij.}$ may be partitioned the same way as y_{ij} for the unreplicated two-way anova (or randomized complete block design). Therefore.

$$\bar{y}_{ij.} = \bar{y}_{...} + \underbrace{\bar{y}_{i..} - \bar{y}_{...}}_{\hat{\alpha}_i} + \underbrace{\bar{y}_{.j.} - \bar{y}_{...}}_{\hat{\beta}_j} + \underbrace{\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}}_{\widehat{\alpha\beta}_{ij}}$$

and we get.

$$y_{ijk} = \bar{y}_{...} + \underbrace{\bar{y}_{i..} - \bar{y}_{...}} + \underbrace{\bar{y}_{.j.} - \bar{y}_{...}} + \underbrace{\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}} + \underbrace{y_{ijk} - \bar{y}_{ij.}}$$

Also $y_{ijk} - \bar{y}_{...} = \bar{y}_{ij.} - \bar{y}_{...} + y_{ijk} - \bar{y}_{ij.}$

Since $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{m}(\bar{y}_{ij.} - \bar{y}_{...})(y_{ijk} - \bar{y}_{ij.}) = 0$

We obtain

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{m}(y_{ijk} - \bar{y}_{...})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{m}(\bar{y}_{ij.} - \bar{y}_{...})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{m}(y_{ijk} - \bar{y}_{ij.})^2$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{m}(\bar{y}_{i..} - \bar{y}_{...})^2 + \sum_{i=1}^{m}\sum_{j=1}^{b}\sum_{k=1}^{m}(\bar{y}_{.j.} - \bar{y}_{...})^2 + \sum_{i=1}^{m}\sum_{j=1}^{b}\sum_{k=1}^{m}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$$

$$+ \sum_{i=1}^{a}\sum_{j=1}^{k}\sum_{k=1}^{m}(y_{ijk} - \bar{y}_{ij.})^2$$

or, $SS_T = SS_A + SS_B + SS_{AB} + SS_E$