Computer Systems Performance Analysis and Benchmarking (37-235)

Analytic Modeling

Simulation

Measurements / Benchmarking

Lecture/Assignments/Projects:

Dipl. Inf. Ing. Christian Kurmann

Textbook:

Raj Jain, “The Art of Computer Systems Performance Analysis”, 1991 Wiley & Sons, New York

Topic of Today:

- **2^{k-p} Factorial Design**
- **One Factor Experiments**
- **Two Factor Full Factorial Design (no replications)**

Experimental Design

Basic Goal:

- Find out which factor contributes how much.
- Do this before and during the analysis and not just after the analysis.

Advanced Goal:

- If you have to reduce experiments do this in a clever way.

2^{k-p} Fractional Factorial Design

allows analysis of k two-level factors with only 2^{k-p} experiments, p suitably chosen.

Look at an Example first. A 2^7 design would require 128 experiments.

Example of a 2^{7-4} design

TABLE 19.1 A 2^{7-4} Experimental Design

Experiment No.	*A*	*B*	*C*	*D*	*E*	*F*	*G*
1	−1	−1	−1	1	1	1	−1
2	1	−1	−1	−1	−1	1	1
3	−1	1	−1	−1	1	−1	1
4	1	1	−1	1	−1	−1	−1
5	−1	−1	1	1	−1	−1	1
6	1	−1	1	−1	1	−1	−1
7	−1	1	1	−1	−1	1	−1
8	1	1	1	1	1	1	1

Properties:

Orthogonality of sign vectors

1. The sum of each column is zero:

$$\sum_i x_{ij} = 0 \qquad \forall j$$

   where x_{ij} represents the level of the jth variable in the ith experiment.

2. The sum of the products of any two columns is zero:

$$\sum_i x_{ij} x_{il} = 0 \qquad \forall j \neq l$$

3. The sum of the squares of each column is 2^{7-4}, that is, 8:

$$\sum_i x_{ij}^2 = 8 \qquad \forall j$$

Computing Effects

- Orthogonality allows to compute effects and contribution to y’s variation.
- Estimate y
- Effective repsonses y_i

$$y = q_0 + q_A x_A + q_B x_B + q_C x_C + q_D x_D + q_E x_E + q_F x_F + q_G x_G$$

Using the orthogonality property of the factor levels chosen, it can be shown that

$$q_A = \sum_i y_i x_{Ai} = \frac{-y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8}{8}$$

Similarly,

$$q_B = \sum_i y_i x_{Bi} = \frac{-y_1 - y_2 + y_3 + y_4 - y_5 - y_6 + y_7 + y_8}{8}$$

and so on.

- Thus the effects are computed like in the full factorial design.
- Standard deviations of effects and confidence intervals are similar for a full factorial design with k-p factors.
- Just replace 2^k by 2^{k-p} in the formulas.

Example:

TABLE 19.2 Data for a Seven-Factor Experimental Design

I	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*y*
1	−1	−1	−1	1	1	1	−1	20
1	1	−1	−1	−1	−1	1	1	35
1	−1	1	−1	−1	1	−1	1	7
1	1	1	−1	1	−1	−1	−1	42
1	−1	−1	1	1	−1	−1	1	36
1	1	−1	1	−1	1	−1	−1	50
1	−1	1	1	−1	−1	1	−1	45
1	1	1	1	1	1	1	1	82
317	101	35	109	43	1	47	3	Total
39.62	12.62	4.37	13.62	5.37	0.125	5.87	0.37	Total/8

- The effect of factors A-G on variation:

  37.26, 4.74, 43.40, 6.75, 0, 8.06, 0.03%
- A and C explain almost everything

No interactions just Factors

- proceed if interactions are negligible
- saves lot of work

Preparing the sign table

1. Choose $k-p$ factors and prepare a complete sign table for a full factorial design with $k-p$ factors. This will result in a table of 2^{k-p} rows and 2^{k-p} columns. The first column will be marked I and consists of all 1's. The next $k-p$ columns will be marked with the $k-p$ factors that were chosen. The remaining columns are simply products of these factors.
2. Of the $2^{k-p}-k+p-1$ columns on the right, choose p columns and mark them with the p factors that were not chosen in step 1.

Example

TABLE 19.3 A 2^3 Experimental Design

Experiment No.	*A*	*B*	*C*	*AB*	*AC*	*BC*	*ABC*
1	−1	−1	−1	1	1	1	−1
2	1	−1	−1	−1	−1	1	1
3	−1	1	−1	−1	1	−1	1
4	1	1	−1	1	−1	−1	−1
5	−1	−1	1	1	−1	−1	1
6	1	−1	1	−1	1	−1	−1
7	−1	1	1	−1	−1	1	−1
8	1	1	1	1	1	1	1

Substitute D,E,F,G for AB, AC,BC,ABC and obtain 2^{7-4} design.

Example 2^{4-1} design

TABLE 19.4 A 2^{4-1} Experimental Design

Experiment No.	A	B	C	AB	AC	BC	D
1	−1	−1	−1	1	1	1	−1
2	1	−1	−1	−1	−1	1	1
3	−1	1	−1	−1	1	−1	1
4	1	1	−1	1	−1	−1	−1
5	−1	−1	1	1	−1	−1	1
6	1	−1	1	−1	1	−1	−1
7	−1	1	1	−1	−1	1	−1
8	1	1	1	1	1	1	1

Concept of confounding

- Some effects overlap each other...

$$q_A = \sum_i y_i x_{Ai} = \frac{-y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8}{8}$$

Similarly, the effect of D is given by

$$q_D = \sum_i y_i x_{Di} = \frac{-y_1 + y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + y_8}{8}$$

The effect of the interaction ABC is obtained by multiplying the respective elements of columns A, B, C, and y. This gives

$$q_{ABC} = \sum_i y_i x_{Ai} x_{Bi} x_{Ci} = \frac{-y_1 + y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + y_8}{8}$$

Notice that the expression for q_{ABC} is identical to that for q_D. In fact, the expression is neither q_D nor q_{ABC}; it is the sum of the two:

$$q_D + q_{ABC} = \sum_i y_i x_{Ai} x_{Bi} x_{Ci} = \frac{-y_1 + y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + y_8}{8}$$

Confounding

D=ABC but also A=BCD because:

$$q_A = q_{BCD} = \sum_i y_i x_{Ai} = \frac{-y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8}{8}$$

Complete list:

$$A = BCD, \quad B = ACD, \quad C = ABD, \quad AB = CD$$
$$AC = BD, \quad BC = AD, \quad ABC = D, \quad I = ABCD$$

Alternate design:

TABLE 19.5 Another 2^{4-1} Experimental Design

Experiment No.	A	B	C	D	AC	BC	ABC
1	−1	−1	−1	1	1	1	−1
2	1	−1	−1	−1	−1	1	1
3	−1	1	−1	−1	1	−1	1
4	1	1	−1	1	−1	−1	−1
5	−1	−1	1	1	−1	−1	1
6	1	−1	1	−1	1	−1	−1
7	−1	1	1	−1	−1	1	−1
8	1	1	1	1	1	1	1

List of confoundings:

$$I = ABD, \quad A = BD, \quad B = AD, \quad C = ABCD$$
$$D = AB, \quad AC = BCD, \quad BC = ACD, \quad ABC = CD.$$

Algebra of Cofounding

- Ring or Field stuctures with generator similar to finite fields over GF[2].
- Given one confounding, it is possible to list all others by multiplying:

Rules

1. The mean I is treated as unity. For example, I multiplied by A is A.
2. Any term with a power of 2 is erased. For example, AB^2C is the same as AC.

Let us illustrate this with the first design, which has

$$I = ABCD$$

Multiplying both sides by A, we get

$$A = A^2BCD = BCD$$

Multiplying both sides by B, C, D, and AB, we get

$$B = AB^2CD = ACD$$

$$C = ABC^2D = ABD$$

$$D = ABCD^2 = ABC$$

$$AB = A^2B^2CD = CD$$

and so on.

The polynomial $I = ABCD$ used to generate all confoundings for this design is called the **generator polynomial** for this design. Similarly, the generator polynomial for the second design is $I = ABC$.

Example of a Generator for 2^{7-4}:

$$D = AB, \quad E = AC, \quad F = BC, \quad G = ABC$$

Multiplying each of the four equations by their left-hand sides, we get

$$I = ABD, \quad I = ACE, \quad I = BCF, \quad I = ABCG$$

Or equivalently,

$$I = ABD = ACE = BCF = ABCG$$

The product of any subset of the preceding terms is also equal to I. Thus the complete generator polynomial is

$$I = ABD = ACE = BCF = ABCG = BCDE = ACDF$$
$$= CDG = ABEF = BEG = AFG = DEF = ADEG$$
$$= BDFG = ABDG = CEFG = ABCDEFG$$

Other confoundings for the design can be obtained by multiplying this equation by $A, B, \ldots$. For example,

$$A = BD = CE = ABCF = BCG = ABCDE = CDF$$
$$= ACDG = BEF = ABEG = FG = ADEF = DEG$$
$$= ABDFG = BDG = ACEFG = BCDEFG$$

Design resolution:

- Degree of that generator polynomial: Min over all confounding expressions: Order of confounded effects left plus sum of order confounded effects right.
- denoted by Roman letters: 2_{III}^{5-1}
- higher resolution is considered better

Latex vs. troff

Case Study 19.1 The CPU time taken by two text-formatting programs, called LaTeX and troff, was measured using synthetic files of various sizes and complexity levels. Six factors, each with two levels were chosen for the study. The first two factors were the text-formatting programs and size of the files. The remaining four factors were number of equations, floats, tables, and footnotes in the file. The assignment of factors and their levels is shown in Table 19.6. A 2^{6-1} fractional factorial design with the generator polynomial $I = BCDEF$ was used. The largest effects and interactions, computed by the sign table method, are shown in Table 19.7. The following conclusions can be reached from these results:

1. Over 90% of the variation is explained by the three factors Bytes, Program, and Equations and a second-order interaction.
2. The text file sizes (in bytes) in these experiments were significantly different, making the effect more than that of the text-formatting programs being compared.
3. The high percentage of variation explained by the "program × Equation" interaction indicates that the choice of the text-formatting program depends upon the number of equations in the text. If we consider only the programs and equations in isolation, the relative amount of CPU time for various combinations is shown in Table 19.8. This shows that troff takes too much CPU time if there are equations in the text.
4. The "Program × Bytes" interaction is low. This indicates that changing the file size affects both programs in a similar manner.
5. If possible, the experiments should be redone with a reduced range of file sizes so that the programs rather than the workload come out as the most significant factor. Alternately, the number of levels of file sizes should be increased. □

TABLE 19.6 Factors and Levels for Text Formatting Programs

Symbol	Factor	Level −1	Level +1
A	Program	LaTeX	troff
B	Bytes	2100	25,000
C	Equations	0	10
D	Floats	0	10
E	Tables	0	10
F	Footnotes	0	10

TABLE 19.7 Effects for the Text Formatting Programs

Symbol	Factor	Effect	Percentage of Variation
B	Bytes	12.0	39.4
A	Program	9.4	24.4
C	Equations	7.5	15.6
AC	Program × Equations	7.2	14.4
E	Tables	3.5	3.4
F	Footnotes	1.6	0.7

TABLE 19.8 CPU Time for Various Program × Equation Combinations

	Number of Equations	
Program	−1(0)	1(10)
−1(LaTeX)	−9.7	−9.1
1(troff)	−5.3	24.1

One Factor Experiments

- Used when more than two levels (to compare several alternatives of a single categorical variable)

Statistical model

- r observations [i]
- a alternatives [j]
- ar measurements [i,j] = r x a matrix
- y_{ij} response of ith entry in jth column
- α_j effects
- e_{ij} errors

The model used in single-factor designs is

$$y_{ij} = \mu + \alpha_j + e_{ij} \qquad (20.1)$$

Here, y_{ij} is the ith response (or observation) with the factor at level j (that is, the jth alternative), μ is the mean response, α_j is the effect of alternative j, and e_{ij} is the error term. The effects are computed so that they add up to zero:

$$\sum \alpha_j = 0$$

Computation of the effects:

If we substitute the observed responses in the model Equation we obtain ar equations. Adding these equations, we get

$$\sum_{i=1}^{r}\sum_{j=1}^{a} y_{ij} = ar\mu + r\sum_{j=1}^{a}\alpha_j + \sum_{i=1}^{r}\sum_{j=1}^{a} e_{ij}$$

Since the effects α_j add up to zero (by design) and we want the mean error to be zero, the preceding equation becomes

$$\sum_{i=1}^{r}\sum_{j=1}^{a} y_{ij} = ar\mu + 0 + 0$$

The model parameter μ is therefore given by

$$\mu = \frac{1}{ar}\sum_{i=1}^{r}\sum_{j=1}^{a} y_{ij}$$

The quantity on the right-hand side is the so-called **grand mean** of all ar responses. It is denoted by $\bar{y}_{..}$.

$$\bar{y}_{.j} = \frac{1}{r}\sum_{i=1}^{r} y_{ij}$$

Substituting $\mu + \alpha_j + e_{ij}$ for y_{ij}, we obtain

$$\begin{aligned}\bar{y}_{.j} &= \frac{1}{r}\sum_{i=1}^{r}(\mu + \alpha_j + e_{ij}) \\ &= \frac{1}{r}\left(r\mu + r\alpha_j + \sum_{i=1}^{r} e_{ij}\right) \\ &= \mu + \alpha_j\end{aligned}$$

Here we have assumed that the error terms for r observations belonging to each alternative add up to zero. The parameter α_j can thus be estimated as follows:

$$\alpha_j = \bar{y}_{.j} - \mu = \bar{y}_{.j} - \bar{y}_{..}$$

Example:

Example 20.1 In a code size comparison study, the number of bytes required to code a workload on three different processors R, V, and Z was measured five times each (each time a different programmer was asked to code the same workload).

TABLE 20.1 Data from a Code Size Comparison Study

R	V	Z
144	101	130
120	144	180
176	211	141
288	288	374
144	72	302

TABLE 20.2 Analysis of the Code Size Comparison Study

	R	V	Z	
	144	101	130	
	120	144	180	
	176	211	141	
	288	288	374	
	144	72	302	
Column sum	$\sum y_{.1} = 872$	$\sum y_{.2} = 816$	$\sum y_{.3} = 1127$	$\sum y_{..} = 2815$
Column mean	$\bar{y}_{.1} = 174.4$	$\bar{y}_{.2} = 163.2$	$\bar{y}_{.3} = 225.4$	$\mu = \bar{y}_{..} = 187.7$
Column effect	$\alpha_1 = \bar{y}_{.1} - \bar{y}_{..} = -13.3$	$\alpha_2 = \bar{y}_{.2} - \bar{y}_{..} = -24.5$	$\alpha_3 = \bar{y}_{.3} - \bar{y}_{..} = 37.7$	

Results: levels a=3, replications r=5

- Average 187.7 bytes of storage
- Effects for R, V, Z -13.3, -24.5, +37.7

Errors

Once the model parameters have been computed, we can estimate the response for each of a alternatives. The estimated response for the jth alternative is given by

$$\hat{y}_j = \mu + \alpha_j$$

The difference between the measured and the estimated response represents experimental error. If we compute experimental errors in each of the ar observations, the mean error should come out zero since the parameter values μ and α_j were computed assuming the sum of errors for each column was zero. The variance of the errors can be estimated from the Sum of Squared Errors (SSE):

$$\text{SSE} = \sum_{i=1}^{r} \sum_{j=1}^{a} e_{ij}^2$$

Example 20.2 Computation of errors for the code size comparison study of Example 20.1 is as follows:

$$\begin{bmatrix} 144 & 101 & 130 \\ 120 & 144 & 180 \\ 176 & 211 & 141 \\ 288 & 288 & 374 \\ 144 & 72 & 302 \end{bmatrix} = \begin{bmatrix} 187.7 & 187.7 & 187.7 \\ 187.7 & 187.7 & 187.7 \\ 187.7 & 187.7 & 187.7 \\ 187.7 & 187.7 & 187.7 \\ 187.7 & 187.7 & 187.7 \end{bmatrix} + \begin{bmatrix} -13.3 & -24.5 & 37.7 \\ -13.3 & -24.5 & 37.7 \\ -13.3 & -24.5 & 37.7 \\ -13.3 & -24.5 & 37.7 \\ -13.3 & -24.5 & 37.7 \end{bmatrix}$$

$$+ \begin{bmatrix} -30.4 & -62.2 & -95.4 \\ -54.4 & -19.2 & -45.4 \\ 1.6 & 47.8 & -84.4 \\ 113.6 & 124.8 & 148.6 \\ -30.4 & -91.2 & 76.6 \end{bmatrix}$$

Each observation has been broken into three parts: a grand mean μ, the processor effect α_j's, and the residuals. A matrix notation is used for all three parts. The sum of squares of entries in the residual matrix is

$$\text{SSE} = (-30.4)^2 + (-54.4)^2 + \cdots + (76.6)^2 = 94{,}365.20 \qquad \square$$

Allocation of Variation

$$y_{ij}^2 = \mu^2 + \alpha_j^2 + e_{ij}^2 + 2\mu\alpha_j + 2\mu e_{ij} + 2\alpha_j e_{ij}$$

Adding corresponding terms of ar such equations, we obtain

$$\sum_{i,j} y_{ij}^2 = \sum_{i,j} \mu^2 + \sum_{i,j} \alpha_j^2 + \sum_{i,j} e_{ij}^2 + \text{cross-product terms}$$

The cross-product terms all add to zero due to the constraints that the effects add to zero ($\sum \alpha_j = 0$) and that the errors for each column add to zero ($\sum e_{ij} = 0$). The preceding equation, expressed in terms of sums of squares, can be written as

$$\text{SSY} = \text{SS0} + \text{SSA} + \text{SSE}$$

where SSY is the sum of squares of y, SS0 is the sum of squares of grand means, SSA is the sum of squares of effects, and SSE is the sum of square errors. Note that SS0 and SSA can be easily computed as follows:

$$\text{SS0} = \sum_{i=1}^{r} \sum_{j=1}^{a} \mu^2 = ar\mu^2$$

$$\text{SSA} = \sum_{i=1}^{r} \sum_{j=1}^{a} \alpha_j^2 = r \sum_{j=1}^{a} \alpha_j^2$$

Thus, SSE can be calculated easily from SSY without calculating individual errors.

The total variation of y (SST) is defined as

$$\text{SST} = \sum_{i,j} (y_{i,j} - \bar{y}_{..})^2 = \sum_{i,j} y_{ij}^2 - ar\bar{y}_{..}^2 = \text{SSY} - \text{SS0} = \text{SSA} + \text{SSE}$$

The total variation can therefore be divided into two parts, SSA and SSE,

$$SSY = 144^2 + 120^2 + \cdots + 302^2 = 633{,}639$$

$$SS0 = ar\mu^2 = 3 \times 5 \times (187.7)^2 = 528{,}281.7$$

$$SSA = r \sum_j \alpha_j^2 = 5[(-13.3)^2 + (-24.5)^2 + (37.6)^2] = 10{,}992.1$$

$$SST = SSY - SS0 = 633{,}639.0 - 528{,}281.7 = 105{,}357.3$$

$$SSE = SST - SSA = 105{,}357.3 - 10{,}992.1 = 94{,}365.2$$

$$\text{Percentage of variation explained by processors} = 100 \times \frac{10{,}992.13}{105{,}357.3} = 10.4\%$$

The remaining 89.6% of the variation in code size is due to experimental errors, which in this case could be attributed to programmer differences. The issue of whether 10.4%—the processors's contribution to variation—is statistically significant is addressed in the next section. □

Analysis of Variance (ANOVA)

Until now: Factor with high variation is important. But this importance should be distinguished from significance.

To determine significance, comparison of the contribution of a factor to the variation with that of the errors.

- Significance Test
- Chi-square Test ?
- Incorporate distribution and degrees of freedom.

TABLE 20.3 ANOVA Table for One-Factor Experiments

Component	Sum of Squares	Percentage of Variation	Degrees of Freedom	Mean Square	F-Computed	F-Table
y	$SSY = \sum y_{ij}^2$		ar			
$\bar{y}_{..}$	$SS0 = ar\mu^2$		1			
$y - \bar{y}_{..}$	$SST = SSY - SS0$	100	$ar - 1$			
A	$SSA = r\sum \alpha_i^2$	$100\left(\frac{SSA}{SST}\right)$	$a - 1$	$MSA = \frac{SSA}{a-1}$	$\frac{MSA}{MSE}$	$F_{[1-\alpha; a-1, a(r-1)]}$
e	$SSE = SST - SSA$	$100\left(\frac{SSE}{SST}\right)$	$a(r - 1)$	$MSE = \frac{SSE}{a(r-1)}$		

$$s_e = \sqrt{MSE}$$

ANOVA Example

- Code Size Comparison 20.1

TABLE 20.4 ANOVA Table for the Code Size Comparison Study

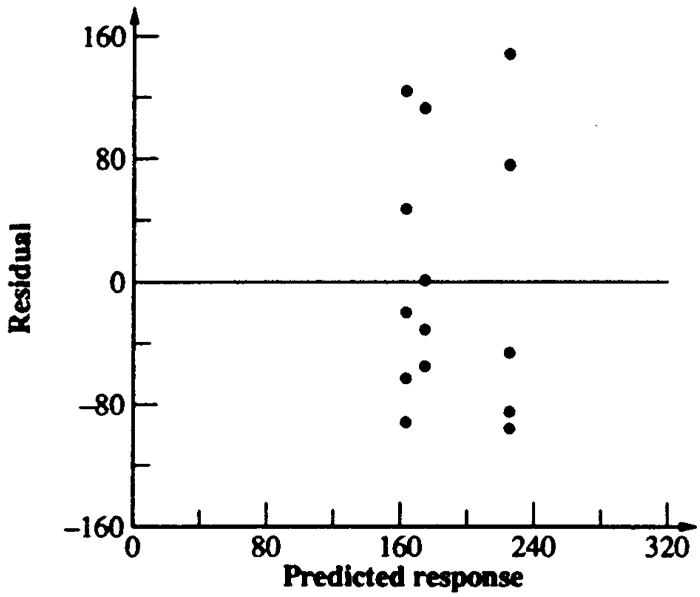
Component	Sum of Squares	Percentage of Variation	Degrees of Freedom	Mean Square	F-Computed	F-Table
y	633,639.00					
$y_{..}$	528,281.69					
$y - y_{..}$	105,357.31	100.0	14			
A	10,992.13	10.4	2	5496.1	0.7	2.8
Errors	94,365.20	89.6	12	7863.8		

$$s_e = \sqrt{\text{MSE}} = \sqrt{7863.77} = 88.68$$

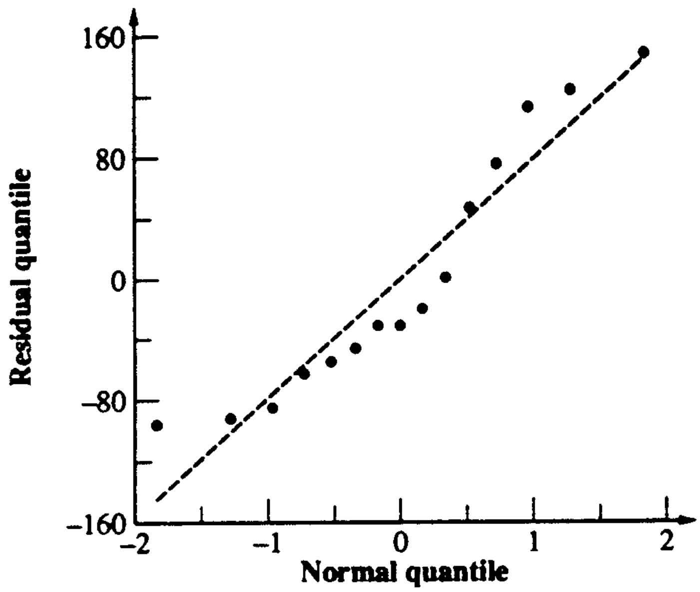
- The computed F-value is less than that from the table. Therefore the observed difference in the code sizes is mostly due to experimental errors and not to any significant difference among the processors.

Visual Test

- Are prerequistites satisfied?



Plot of the residuals versus predicted response for the code size comparison study.



Normal quantile-quantile plot for the residuals of the code size comparison study.

Confidence Intervals

Confidence Intervals for

- α_j is something above/below average with 95% confidence.
- α_i-α_j is a difference relevant with 95% confidence.

Do calculations according to receipe in the book.

1. Model: $y_{ij} = \mu + \alpha_j + e_{ij}$; the effects are computed so that $\sum_{j=1}^{a} \alpha_j = 0$.
2. Effects: $\mu = \overline{y}_{..} = \sum_{j=1}^{a} \sum_{i=1}^{r} y_{ij}$
   $\alpha_j = \overline{y}_{.j} - \overline{y}_{..} = \sum_{i=1}^{r} y_{ij} - \overline{y}_{..}, \; j = 1, 2, \ldots, a$
3. Allocation of variation: SSE can be calculated after computing other terms below:

$$\sum_{ij} y_{ij}^2 = ar\mu^2 + r\sum_{j} \alpha_j^2 + \sum_{ijk} e_{ij}^2$$

$$\text{SSY} = \text{SS0} + \text{SSA} + \text{SSE}$$

4. Degrees of freedom: $\text{SSY} = \text{SS0} + \text{SSA} + \text{SSE}$

$$ar = 1 + (a-1) + a(r-1)$$

5. Mean squares: $\text{MSA} = \text{SSA}/(a-1)$; $\text{MSE} = \text{SSE}/[a(r-1)]$
6. Analysis of variance: MSA/MSE should be greater than $F_{[1-\alpha; a-1, a(r-1)]}$.
7. Standard deviation of errors: $s_e = \sqrt{\text{MSE}}$
8. Standard deviation of parameters: $s_\mu^2 = s_e^2/ar$; $s_{\alpha_j}^2 = s_e^2(a-1)/ar$
9. Contrast of effects $\sum_{j=1}^{a} h_j \alpha_j$, where $\sum_{j=1}^{a} h_j = 0$:

$$\text{Mean} = \sum_{j=1}^{a} h_j \overline{y}_{.j}; \qquad \text{Variance} = \sum_{j=1}^{a} s_e^2 h_j^2 / ar$$

10. All confidence intervals are computed using $t_{[1-\alpha/2; a(r-1)]}$.
11. Model assumptions:
    (a) Errors are IID normal variates with zero mean.
    (b) Errors have the same variance for all factor levels.
    (c) The effect of the factor and errors are additive.
12. Visual tests:
    (a) The scatter plot of errors versus predicted responses should not have any trend.
    (b) The normal quantile-quantile plot of errors should be linear.
    (c) Spread of y values for all levels of the factor should be comparable.

If any test fails or if the ratio y_{max}/y_{min} is large, multiplicative models or transformations should be investigated.