EEC 686/785
Modeling & Performance Evaluation of Computer Systems

Lecture 13

Wenbing Zhao
Department of Electrical and Computer Engineering
Cleveland State University
wenbing@ieee.org
(based on Dr. Raj Jain's lecture notes)

Outline

- Review of lecture 12
- 2^{k-p} Fractional Factorial Designs
- One factor experiment

$2^k r$ Factorial Designs with Replication

- r replications of 2^k experiments
  - $2^k r$ observations
  - Allows estimation of experimental errors
- Model: $y = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B + e$
  - e = experimental error
- Computation of effects: use sign table
- Estimation of errors:

$$e_{ij} = y_{ij} - \hat{y}_i = y_{ij} - q_0 + q_A x_{Ai} + q_B x_{Bi} + q_{AB} x_{Ai} x_{Bi}$$

$2^k r$ Factorial Designs with Replication

- Allocation of Variation:

$$SST = \sum_{ij} (y_{ij} - \bar{y}_{..})^2 = 2^2 r q_A^2 + 2^2 r q_B^2 + 2^2 r q_{AB}^2 + \sum_{i,j} e_{ij}^2$$

$$SST = SSA + SSB + SSAB + SSE$$

- Effects are random variables
  - Errors $\sim N(0, \sigma_e) \Rightarrow y \sim N(\bar{y}_{..}, \sigma_e)$
  - q_0 is normal with variance $\sigma_e^2 / (2^2 r)$

Confidence Intervals for Effects

- Variance of errors: $s_e^2 = \frac{1}{2^2(r-1)}\sum_{ij} e_e^2 = \frac{SSE}{2^2(r-1)} \triangleq MSE$
- Estimated variance of q_0: $s_{q_0}^2 = s_e^2/(2^2 r)$
- Similarly, $s_{q_A} = s_{q_B} = s_{q_{AB}} = s_e \big/ \sqrt{2^2 r}$
- Confidence intervals (CI) for the effects:

$$q_i \mp t_{[1-\alpha/2;2^2(r-1)]} s_{q_i}$$

- CI does not include a zero => significant

Multiplicative Models for 2^2r Experiments

- **Additive model**. Not valid if effects do not add
- E.g., execution time of workloads
  Execution time $y_{ij} = v_i \times w_j$
  The two effects **multiply**. Logarithm => additive model: $\log(y_{ij}) = \log(v_i) + \log(w_j)$
- **Correct model**: $y'_{ij} = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B + e_{ij}$
  where, $y'_{ij} = \log(y_{ij})$
- Taking an antilog of additive effects *q*. to get the multiplicative effects *u*.

2^{k-p} Fractional Factorial Designs

- Large number of factors
  - Large number of experiments
  - Full factorial design too expensive
  - Use a fractional factorial design
- 2^{k-p} design allows analyzing *k* factors with only 2^{k-p} experiments
  - 2^{k-1} design requires only half as many experiments
  - 2^{k-2} design requires only one quarter of the experiments

2^{7-4} Design

- Study 7 factors with only experiments!

Expt 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

2^{7-4} Design

- Full factorial design is easy to analyze due to orthogonality of sign vectors
- Fractional factorial designs also use orthogonal vectors
  - The sum of each column is zero:
    $$\sum_i x_{ij} = 0 \quad \forall j, \quad j\text{th variable, } i\text{th experiment}$$
  - The sum of the products of any two columns is zero
    $$\sum_i x_{ij} x_{il} = 0 \quad \forall j \neq l$$
  - The sum of the squares of each column is $2^{7-4} = 8$
    $$\sum_i x_{ij}^2 = 8 \quad \forall j$$

2^{7-4} Design

- Model:

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

- Effects can be computed using inner products

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

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

2^{7-4} 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

- Factors A through G explains 37.26%, 4.74%, 43.40%, 6.75%, 0%, 8.06%, and 0.03% of variation, respectively
  => Use only factors C and A for further experimentation

Preparing Sign Table for 2^{k-p} Design

- Prepare a sign table for a full factorial design with k-p factors
- Mark the first column I
- Mark the next k-p columns with the k-p factors
- Of the $(2^{k-p} - k + p - 1)$ columns on the right, choose p columns and mark them with the p factors which were not chosen in step 1

Example: 2^{7-4} Design

- Start with 2^3 design
- Step 1: k=7, p=4 (3 factors), sign table=>
- Step 2: mark 1st column I
- Mark next 3 columns (i.e., A, B, C) with 3 factors (i.e., A, B, C)
- Mark AB, AC, BC, ABC with factors D, E, F, G

Expt 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

Example: 2^{4-1} Design

- Start with 2^3 design
- Step 1:
- Step 2:
- Step 3:
- Step 4: choose the rightmost column and mark it D => can study effects q_A, q_B, q_C, and q_D along with interactions q_{AB}, q_{AC}, and q_{BC}

Expt 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

Confounding: 2^{4-1} Design

- Confounding: only the combined influence of two or more effects can be computed

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

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

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

$$q_D = q_{ABC}$$

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

- Effects of D and ABC are confounded. Not a problem if q_{ABC} is negligible

Confounding: 2^{4-1} Design

- Confounding representation: D = ABC
- Other confoundings:

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

$$=> A = BCD$$

$$A = BCD, B = ACD, C = ABC, AB = CD,$$

$$AC = BD, BC = AD, ABC = D, and\ I = ABCD$$

- I = ABCD => confounding of ABCD with the mean

Other Fractional Factorial Designs

- A fractional factorial design is not unique
- 2^p different fractional factorial designs possible
- Confounding: I=ABD, A=BD, B=AD, C=ABCD, D=AB, AC=BCD, BC=ACD, ABC=CD
- Not as good as the previous design

Another 2^{4-1} Experimental Design

Expt 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

Algebra of Confounding

- Given just one confounding, it is possible to list all other confoundings
- Rules:
  - I is treated as unity
  - Any term with a power of 2 is erased

Algebra of Confounding

$I = ABCD$

Multiplying both sides by A:

$A = A^2BCD = BCD$

Multiplying both sides by $B, C, D,$ and AB:

$B = AB^2CD = ACD$

$C = ABC^2D = ABD$

$D = ABCD^2 = ABC$

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

and so on.

generator polynomial: $I = ABCD$

For the second design: $I = ABC$

Algebra of Confounding

- In a 2^{k-p} design, 2^p effects are confounded together
- In the design:
  - D=AB, E=AC, F=BC, G=ABC
    => I=ABD, I=ACE, I=BCF, I=ABCG
    =>I=ABD=ACE=BCF=ABCG
  - Using products of all subsets:
    I=ABD=ACE=BCF=ABCG=BCDE=ACDF=CDG
    =ABEF=BEG=AFG=DEF=ADEG=BDFG
    =CEFG=ABCDEFG
  - Other confoundings:
    A=BD=CE=ABCF=BCG=ABCDE=CDF=ACDG=BEF
    =ABEG=FG=ADEF=DEG=ABDFG=ACEFG=BCDEFG

Design Resolution

- Order of an effect = number of factors included in it
- Order of ABCD=4, order of I=0
- Order of a confounding = sum of the order of two terms
  - E.g., AB=CDE is of order 5
- Resolution of a design = minimum of orders of confoundings
- Notation: R_{III} = Resolution III = 2^{k-p}_{III}

Design Resolution

- **Example 1**:

$$I = ABCD \Rightarrow R_{IV} = \text{Resolution IV} = 2^{4-1}_{IV}$$

$$A = BCD, B = ACD, C = ABD, AB = CD,$$

$$AC = BD, BC = AD, ABC = D, \text{and } I = ABCD$$

- **Example 2**: $I = ABD \Rightarrow R_{III}$ design

Design Resolution

- **Example 3**:

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

$$= ACDF = CDG = ABEF = BEG = AFG$$

$$= DEF = ADEG = BDFG = CEFG = ABCDEFG$$

  - This is a resolution-III design
  - A design of higher resolution is considered a better design

Case Study: Latex vs. troff

Factors and Levels

	Factor	-Level	+Level
A	Program	Latex	troff-me
B	Bytes	2100	25000
C	Equations	0	10
D	Floats	0	10
E	Tables	0	10
F	Footnotes	0	10

	Factor	Effect	% 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	0.70%

Design: 2^{6-1} with I=BCDEF

Case Study: Latex vs. troff

- **Conclusions**
  - Over 90% of the variation is due to: bytes, program, and equations and a second order interaction
  - Text file sizes were significantly different making its effect more than that of the programs
  - High percentage of variation explained by the "program × equation" interaction => Choice of the text formatting program depends upon the number of equations in the text. troff not as good for equations

CPU Time

Program	# of Equations	
	-1(0)	1(10)
-1(Latex)	-9.7	-9.1
1(Troff)	-5.3	24.1

Case Study: Latex vs. troff

- **Conclusions** (continued)
  - Low "program × bytes" interaction => Changing the file size affects both programs in a similar manner
  - In next phase, reduce range of file sizes. Alternately, increase the number of levels of file sizes

Case Study: Scheduler Design

- Three classes of jobs: word processing, data processing, and background data processing
  - Design: 2^{5-1} with $I=ABCDE$

Factors and Levels in the Scheduler Design Study

Symbol	Factor	Level -1	Level 1
A	Preemption	No	Yes
B	Time Slice	Small	Large
C	Queue Assignment	One Queue	Two Queues
D	Requeueing	Two Queues	Five Queues
E	Fairness	Off	On

Case Study: Scheduler Design

- Measured throughputs
  - T_W: for word processing
  - T_I: for interactive data processing
  - T_B: for batch data processing

No.	A	B	C	D	E	T_W	T_I	T_B
1	−1	−1	−1	−1	1	15.0	25.0	15.2
2	1	−1	−1	−1	−1	11.0	41.0	3.0
3	−1	1	−1	−1	−1	25.0	36.0	21.0
4	1	1	−1	−1	1	10.0	15.7	8.6
5	−1	−1	1	−1	−1	14.0	63.9	7.5
6	1	−1	1	−1	1	10.0	13.2	7.5
7	−1	1	1	−1	1	28.0	36.3	20.2
8	1	1	1	−1	−1	11.0	23.0	3.0
9	−1	−1	−1	1	−1	14.0	66.1	6.4
10	1	−1	−1	1	1	10.0	9.1	8.4
11	−1	1	−1	1	1	27.0	34.6	15.7
12	1	1	−1	1	−1	11.0	23.0	3.0
13	−1	−1	1	1	1	14.0	26.0	12.0
14	1	−1	1	1	−1	11.0	38.0	2.0
15	−1	1	1	1	−1	25.0	35.0	17.2
16	1	1	1	1	1	11.0	22.0	2.0

Case Study: Scheduler Design

- Effects and variation explained

Confounded Effects 1	2	T_W Estimate	Perc. Var.	T_I Estimate	Perc. Var.	T_B Estimate	Perc. Var.
I	ABCDE	15.44		31.74		9.54	
A	BCDE	-4.81	55.5%	-8.62	31.0%	-4.86	58.8%
B	ACDE	3.06	22.5%	-3.54	5.2%	1.79	8.0%
C	ABDE	0.06	0.0%	0.43	0.1%	-0.62	1.0%
D	ABCE	-0.06	0.0%	-0.02	0.0%	-1.21	3.6%
AB	CDE	-2.94	20.7%	1.34	0.8%	-2.33	13.5%
AC	BDE	0.06	0.0%	0.49	0.1%	-0.44	0.5%
AD	BCE	0.19	0.1%	-0.08	0.0%	0.37	0.3%
BC	ADE	0.19	0.1%	0.44	0.1%	-0.12	0.0%
BD	ACE	0.06	0.0%	0.47	0.1%	-0.66	1.1%
CD	ABE	-0.19	0.1%	-1.91	1.5%	0.58	0.8%
DE	ABC	-0.06	0.0%	0.21	0.0%	-0.47	0.5%
CE	ABD	0.06	0.0%	1.21	0.6%	-0.16	0.1%
BE	ACD	0.31	0.2%	7.96	26.4%	-1.37	4.7%
AE	BCD	-0.56	0.8%	0.88	0.3%	0.28	0.2%
E	ABCD	0.19	0.1%	-9.01	33.8%	1.66	6.8%

Case Study: Scheduler Design

- **Conclusions**
  - For work processing throughput (T_W): A (preemption), B (time slice), and AB are important
  - For interactive jobs: E (fairness), A (preemption), BE, and B (time slice)
  - For background jobs: A (preemption), AB, B (time slice), E (fairness)
  - May use different policies for different classes of workloads
  - Factors C (queue assignment) or any of its interaction do not have any significant impact on the throughput

Case Study: Scheduler Design

- **Conclusions** (continued):
  - Factor D (requeuing) is not effective
  - Preemption (A) impacts all workloads significantly
  - Time slide (B) impacts less than preemption
  - Fairness (E) is important for interactive jobs and slightly important for background jobs

One Factor Experiments

- Used to compare alternatives of a single **categorical** variable
  - For example, several processors, several caching schemes
- Model: $y_{ij} = \mu + \alpha_j + e_{ij}$

r = number of replications $\quad y_{ij}$ = ith response with jth alternative

μ = mean response $\quad \alpha_j$ = effect of alternative j

e_{ij} = error term

$$\sum \alpha_j = 0$$

Computation of Effects

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

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

$$\bar{y}_{.j} = \frac{1}{r}\sum_{i=1}^{r} y_{ij} = \frac{1}{r}\sum_{i=1}^{r}(\mu + \alpha_j + e_{ij})$$

$$= \frac{1}{r}\left(r\mu + r\alpha_j + \sum_{i=1}^{r} e_{ij}\right) = \mu + \alpha_j + 0$$

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

Example: Code Size Comparison

- Entries in a row are unrelated. Otherwise, need a two factor analysis

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

Analysis of Code Size Comparison Data

	R	V	Z	
	144	101	130	
	120	144	180	
	176	211	141	
	288	288	374	
	144	72	302	
Col Sum	$\Sigma y_{.1} = 872$	$\Sigma y_{.2} = 816$	$\Sigma y_{.3} = 1127$	$\Sigma y_{..} = 2815$
Col Mean	$\bar{y}_{.1} = 174.4$	$\bar{y}_{.2} = 163.2$	$\bar{y}_{.3} = 225.4$	$\mu = \bar{y}_{..} = 187.7$
Col 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	

Interpretation

- Average processor requires 187.7 bytes of storage
- The effects of the processors R, V, and Z are -13.3, -24.5, and 37.7, respectively. That is
  - R requires 13.3 bytes less than an average processor
  - V requires 24.5 bytes less than an average processor
  - Z requires 37.7 bytes more than an average processor

Estimating Experimental Errors

- Estimated response for *j*th alternative

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

- Error:

$$e_{ij} = y_j - \hat{y}_j$$

- Sum of squared errors (SSE):

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

Example

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

$$+ \begin{bmatrix} -13.3 & -24.5 & 37.7 \\ -13.3 & -24.5 & 37.7 \\ -13.3 & -24.5 & 37.7 \\ -13.3 & -24.5 & 37.7 \\ -13.3 & -24.5 & 37.7 \end{bmatrix} + \begin{bmatrix} -30.4 & -62.2 & -95.4 \\ -54.4 & -19.2 & -45.4 \\ 1.6 & 47.8 & -84.4 \\ 113.6 & 124.8 & 148.6 \\ -30.4 & -91.2 & 76.6 \end{bmatrix}$$

$$SSE = (-30.4)^2 + (-54.4)^2 + \cdots + (76.6)^2 = 94365.20$$

Allocation of Variation

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

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

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

$$SSY = SS0 + SSA + SSE$$

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

Total variation of y (SST): $SST = \sum_{i,j} (y_{i,j} - \bar{y}_{..})^2$

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

Example

$$SSY = 144^2 + 120^2 + \cdots + 302^2 = 633639$$

$$SS0 = ar\mu^2 = 3 \times 5 \times (187.7)^2 = 528281.7$$

$$SSA = r\sum_j \alpha_j^2 = 5[(-13.3)^2 + (-24.5)^2 + (37.6)^2] = 10992.1$$

$$SST = SSY - SS0 = 633639.0 - 528281.7 = 105357.3$$

$$SSE = SST - SSA = 105357.3 - 10992.1 = 94365.2$$

- Percent variation explained by processors = $100 \times 10992.13/105357.3 = 10.4\%$
  89.6% of variation in code size is due to experimental errors (programmer differences). Is 10.4% statistically significant?

Analysis of Variance (ANOVA)

- Importance $\neq$ significance
- **Important** => explains a high percent of variation
- **Significance** => high contribution to the variation compared to that by errors
- Degree of freedom = number of independent values required to compute

$$SSY = SS0 + SSA + SSE$$

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

Note that the degrees of freedom also add up

F-Test

- Purpose: to check if SSA is *significantly* greater than SSE
  - Errors are normally distributed => SSE and SSA have chi-square distributions
  - The ratio (SSA/ ν_A)/ (SSE/ ν_e) has an F distribution.
    Where $\nu_A = a - 1$ = degree of freedom for SSA
    $\nu_e = a(r - 1)$ = degree of freedom for SSE
  - Computed ratio > $F_{[1-\alpha;\nu_A,\nu_e]}$
    => SSA is **significantly** higher than SSE
  - SSA/ ν_A is called mean square of A or (MSA).
    Similarly, MSE = SSE/ ν_e

ANOVA Table for One Factor Experiment

Component	Sum of Squares	%Variation	DF	Mean Square	F-Comp.	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 $= r\Sigma\alpha_i^2$	$100\left(\frac{\text{SSA}}{\text{SST}}\right)$	$a-1$	MSA $= \frac{\text{SSA}}{a-1}$	$\frac{\text{MSA}}{\text{MSE}}$	$F_{[1-\alpha;a-1,a(r-1)]}$
e	SSE=SST- SSA	$100\left(\frac{\text{SSE}}{\text{SST}}\right)$	$a(r-1)$	MSE $= \frac{\text{SSE}}{a(r-1)}$		

Example: Code Size Comparison

Component	Sum of Squares	%Variation	DF	Mean Square	F-Comp.	F-Table
y	633639.00					
$y_{..}$	528281.69					
$y - y_{..}$	105357.31	100.0%	14			
A	10992.13	10.4%	2	5496.1	0.7	2.8
Errors	94365.20	89.6%	12	7863.8		

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

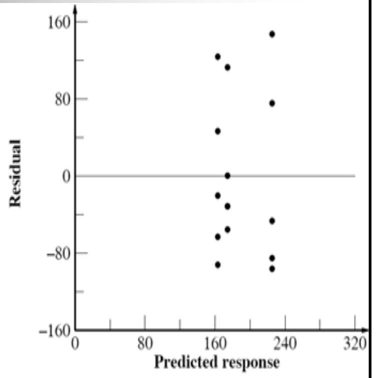
- Computed F-value < F from table
  => the variation in the code sizes is mostly due to experimental errors and not because of any significant difference among the processors

Visual Diagnostic Tests

- Assumptions:
  - Factors effects are additive
  - Errors are additive
  - Errors are independent of factor levels
  - Errors are normally distributed
  - Errors have the same variance for all factor levels
- Visual tests:
  - Normal quantile-quantile plot: Linear => Normality
  - Residuals versus predicted response:
    No trend => independence
    Scale of errors << scale of response => ignore visible trends

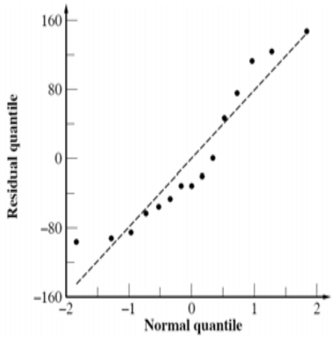
Example

- Horizontal and vertical scales similar
  - ⇒ Residuals are not small
  - ⇒ Variation due to factors is small compared to the unexplained variation
- No visible trend in the spread



Example

- S-shape => shorter tails than normal



Confidence Intervals for Effects

Parameter	Estimate	Variance
μ	$\bar{y}_{..}$	s_e^2/ar
α_j	$\bar{y}_{.j} - \bar{y}_{..}$	$s_e^2(a-1)/ar$
$\mu + \alpha_j$	$\bar{y}$	s_e^2/r
$\Sigma_{j=1}^a h_j\alpha_j$, $\Sigma_{j=1}^a h_j = 0$	$\Sigma_{j=1}^a h_j\bar{y}_{.j}$	$\Sigma_{j=1}^a s_e^2 h_j^2/ar$
s_e^2	$\frac{\sum e_{ij}^2}{a(r-1)}$	

Degrees of freedom for errors $= a(r-1)$

- Estimates are random variables
- For the confidence intervals, use t values at $r(a\text{-}1)$ degree of freedom
- Mean response: $\hat{y}_j = \mu + \alpha_j$
- Contrasts $\sum h_j\alpha_j$: e.g., to compare 2 alternatives, $\alpha_1 - \alpha_2$

Example: Code Size Comparison

Error variance: $S_e^2 = \frac{94365.2}{12} = 7863.8$

Std dev of errors $= \sqrt{(\text{var. of errors})} = 88.7$

Std dev of $\mu = s_e / \sqrt{ar} = 88.7/\sqrt{15} = 22.9$

Std dev of $\alpha_j = s_e / \sqrt{\{(a-1)/(ar)\}} = 88.7/\sqrt{(2/15)} = 32.4$

- For 90% confidence, $t_{[0.95;12]}=1.782$

Example: Code Size Comparison

- 90% confidence intervals:

$\mu = 197.7 \mp (1.782)(22.9) = (146.9, 228.5)$

$\alpha_1 = -13.3 \mp (1.782)(32.4) = (-71.0, 44.4)$

$\alpha_2 = -24.5 \mp (1.782)(32.4) = (-82.2, 33.2)$

$\alpha_3 = 37.6 \mp (1.782)(32.4) = (-20.0, 95.4)$

- The code size on an average processor is significantly different from zero
- Processor effects are not significant

Example: Code Size Comparison

- Using $h_1 = 1$, $h_2 = -1$, $h_3 = 0$, $(\sum h_j = 0)$:

Mean $\alpha_1 - \alpha_2 = \bar{y}_{.1} - \bar{y}_{.2} = 174.4 - 163.2 = 11.2$

Std dev of $\alpha_1 - \alpha_2 = \frac{s_e}{\sqrt{(\sum h_j^2 / ar)}} = \frac{88.7}{\sum(2/15)} = 56.1$

90% CI for $\alpha_1 - \alpha_2 = 11.2 \mp (1.782)(56.1) = (-88.7, 111.1)$

CI includes zero => one isn't superior to other

Example: Code Size Comparison

- Similarly,

90% CI for $\alpha_1 - \alpha_3$

$= (174.4 - 225.4) \mp (1.782)(56.1) = (-150.9, 48.9)$

90% CI for $\alpha_2 - \alpha_3$

$= (163.2 - 225.4) \mp (1.782)(56.1) = (-162.1, 37.7)$

Any one processor is not superior to another

Unequal Sample Sizes

- Model: $y_{ij} = \mu + \alpha_j + e_{ij}$
- By definition: $\sum_{j=1}^{a} r_j \alpha_j = 0$
- Here, r_j is the number of observations at jth level (alternative)
- Total number of observations: $N = \sum_{j=1}^{a} r_j$

Parameter Estimation

Parameter	Estimate	Variance
μ	$\bar{y}_{..}$	s_e^2/N
α_j	$\bar{y}_{.j} - \bar{y}_{..}$	$s_e^2(N - r_j)/(N r_j)$
$\mu + \alpha_j$	$\bar{y}_{.j}$	s_e^2/r_j
$\Sigma h_j \alpha_j$, $\Sigma h_j = 0$	$h_j \bar{y}_{.j}$	$s_e^2 \Sigma_{j=1}^{a}(h_j^2/r_j)$
s_e^2	$\Sigma\, e_{ij}^2/\{N - a\}$	

Degrees of freedom for errors $= N - a$

Analysis of Variance

Component	Sum of Squares	%Variation	DF	Mean Square	F-Comp.	F-Table
y	SSY $= \sum y_{ij}^2$		N			
$\bar{y}_{..}$	SS0 $= N\mu^2$		1			
$y - \bar{y}_{..}$	SST=SSY-SS0	100	$N-1$			
A	SSA $= \sum_{j=1}^{a} r_j \alpha_j^2$	$100\left(\frac{SSA}{SST}\right)$	$a-1$	MSA $= \frac{SSA}{a-1}$	$\frac{MSA}{MSE}$	$F_{[1-\alpha; a-1, N-a]}$
e	SSE=SST- SSA	$100\left(\frac{SSE}{SST}\right)$	$N-a$	MSE $= \frac{SSE}{N-a}$		

Example: Code Size Comparison

	R	V	Z	
	144	101	130	
	120	144	180	
	176	211	141	
	288	288		
	144			
Column Sum	872	744	451	2067
Column Mean	174.40	186.00	150.33	172.25
Column effect	2.15	13.75	-21.92	

- All means are obtained by dividing by the number of observations added
- The column effects are 2.15, 13.75, and -21.92

Analysis of Variance

$$\begin{bmatrix} 144 & 101 & 130 \\ 120 & 144 & 180 \\ 176 & 211 & 141 \\ 288 & 288 & \\ 144 & & \end{bmatrix} = \begin{bmatrix} 172.25 & 172.25 & 172.25 \\ 172.25 & 172.25 & 172.25 \\ 172.25 & 172.25 & 172.25 \\ 172.25 & 172.25 & \\ 172.25 & & \end{bmatrix}$$

$$+ \begin{bmatrix} 2.15 & 13.75 & -21.92 \\ 2.15 & 13.75 & -21.92 \\ 2.15 & 13.75 & -21.92 \\ 2.15 & 13.75 & \\ 2.15 & & \end{bmatrix} + \begin{bmatrix} -30.4 & -85.00 & -20.33 \\ -54.4 & -42.00 & 29.67 \\ 1.6 & 25.00 & -9.33 \\ 113.6 & 102.00 & \\ -30.4 & & \end{bmatrix}$$

Example: Code Size Comparison

- Sums of Squares:

$$SSY = \sum y_{ij}^2 = 397375$$
$$SS0 = N\mu^2 = 356040.75$$
$$SSA = 5\alpha_1^2 + 4\alpha_2^2 + 3\alpha_3^2 = 2220.38$$
$$SSE = (-30.40)^2 + (-54.40)^2 + \cdots + (-9.33)^2 = 39113.87$$
$$SST = SSY - SS0 = 41334.25$$

- Degrees of freedom:

$$SSY = SS0 + SSA + SSE$$
$$N = 1 + (a-1) + N - a$$
$$12 = 1 + 2 + 9$$

ANOVA Table: Code Size Comparison

Component	Sum of Squares	%Variation	DF	Mean Square	F-Comp.	F-Table
y	397375.00					
$y_{..}$	356040.75					
$y - y_{..}$	41334.25	100.00%	11			
A	2220.38	5.37%	2	1110.19	0.26	3.01
Errors	39113.87	94.63%	9	4345.99		

$$s_e = \sqrt{\text{MSE}} = \sqrt{4345.99} = 65.92$$

- Conclusion: variation due to processors is insignificant as compared to that due to modeling errors

Derivation of Standard Deviation

- Consider the effect of processor Z:
- Since,

$$\alpha_3 = y_{.3} - y_{..}$$
$$= \frac{1}{3}(y_{13} + y_{23} + y_{33}) - \frac{1}{12}(y_{11} + y_{21} + \cdots + y_{32} + y_{42} + y_{13} + y_{23} + y_{33})$$
$$= \frac{1}{4}(y_{13} + y_{23} + y_{33}) - \frac{1}{12}(y_{11} + y_{21} + \cdots + y_{32} + y_{42})$$

Derivation of Standard Deviation

- Error in α_3 = $\sum$errors in terms on the right hand size:

$$e_{\alpha_3} = \frac{1}{4}(e_{13} + e_{23} + e_{33}) - \frac{1}{12}(e_{11} + e_{21} + \cdots + e_{32} + e_{42})$$

e_{ij}'s are normally distributed

=> α_3 is normal with $s_{\alpha_3}^2 = \frac{1}{4^2} \times 3s_e^2 + \frac{1}{12^2} \times 9s_e^2 = 1086.36$