Problem 1 2^{6-2}_{III} - design

1. Significant effects.

ANOVA

factor	effect $l_j \cdot 10^2$	$SS \cdot 10^4$	F_{obs}	sign.
A	−0.5	1	6.80	
B	2.5	25	170.1	xx
C	1	4	27.2	x
D	1.1	4.84	32.9	x
E	2.1	17.64	120	xx
F	0.6	1.44	9.80	
AB	1.1	4.84	32.9	x
AD	0.2	0.16	1.09	
AF	0.6	1.44	9.80	
BD	−1.5	9	61.2	xx
BE	1.2	5.76	39.2	xx

$SS_{reg} = \Sigma\ 75.12$

$SS_T(\text{corr}) = 77 \cdot 10^{-4}$; Block effect = −0.006

$SS_T(\text{corr}) = SS_{Reg} + SS_{Block} + SS_E$

df: 15   11   1   3

$SS_E = 77 \cdot 10^{-4} - 75.12 \cdot 10^{-4} - 4.6^2 \cdot 10^{-6}$

$SS_E = 0.44 \cdot 10^{-4}$

$MS_E = \frac{0.44 \cdot 10^{-4}}{3} = 1.47 \cdot 10^{-5}$; compare: $s^2 = 2 \cdot 10^{-5}$

$F(0.95; 1, 3) = 10.13$; $F(0.99; 1, 3) = 34.11$

2. Optimal combination of variables.

Significant model:

$$y = M + B + C + D + E + AB + BD + BE$$

$$y = M + 0.5 \{ l_B x_B + l_C x_C + l_D x_D + l_E x_E + l_{AB} x_A x_B + l_{BD} x_B x_D + l_{BE} x_B x_E \}$$

$$y = 2.7 + \frac{10^{-3}}{2} (25 x_B + 10 x_C + 11 x_D + 21 x_E + 11 x_A x_B - 15 x_B x_D + 12 x_B x_E)$$

B:	$x_B = 1$	$\Delta y = 25 \cdot 10^{-3} \cdot 0.5$
AB:	$x_B = 1$ $x_A = 1$	$11 \cdot$
C:	$x_C = 1$	$10 \cdot$
D:	$x_D = -1$	$-11 \cdot$
BD:	$x_B = 1$ $x_D = -1$	$15 \cdot$
E:	$x_E = 1$	$21 \cdot$
BE:	$x_B = 1$ $x_E = 1$	$12 \cdot$

$$y_{opt} = 2.7 + 10^{-3} \cdot 0.5 (25 + 11 + 10 - 11 + 15 + 21 + 12)$$

$$y_{opt} = \underline{\underline{2.742}}$$

Uppgift 2

3 kontrollerbara variabler: x_1, x_2, x_3
2 icke-kontrollerbara variabler: z_1, z_2
(brusvariabler)

Responsmodell, första ordningen.

$$y(\underline{x}, \underline{z}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 +$$
$$+ \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3$$
$$+ \gamma z_1 + \gamma z_2 +$$
$$p = 15 \quad + \delta_{11} x_1 z_1 + \delta_{21} x_2 z_1 + \delta_{31} x_3 z_1$$
$$+ \delta_{12} x_1 z_2 + \delta_{22} x_2 z_2 + \delta_{32} x_3 z_2 + \varepsilon$$

1. Utför ett 2^5-faktorförsök, dvs 32 försök, t.ex. i pilotskala där z_1 och z_2 kan kontrolleras.

2. Bestäm parametrarna i respons-modellen med regression.
Skattningen av variansen, s^2, beräknas ~~från~~ med residualkvadratsumman, SS_{Res}, $s^2 = SS_{Res}/(n-p)$

3. Beräkning av variansen för prediktionen:

Modellen kan skrivas:

$$y(\underline{x}, \underline{z}) = f(\underline{x}) + h(\underline{x}, z) + \underline{\varepsilon}$$

$$E(\underline{\varepsilon}) = 0 \quad V(\underline{\varepsilon}) = \underline{I}\,\sigma^2$$

$f(\underline{x})$: endast kontrollerbara variabler

$h(\underline{x}, \underline{z})$: huvudeffekter av brusvariabler + interaktion mellan kontrollbara- och brusfaktorer.

$$h(\underline{x}, \underline{z}) = \sum_{i=1}^{2} \gamma_i z_i + \sum_{i=1}^{3} \sum_{j=1}^{2} \delta_{ij} x_i z_j$$

$E(\underline{z}) = 0$

$\underset{z}{V}(\underline{z}) = \underline{\underline{I}} \cdot \sigma_z^2$, $cov(\underline{\varepsilon}, \underline{z}) = 0$

Medel-modell:

$E_z[y(\underline{x}, \underline{z})] = f(\underline{x})$, väntevärde av både $\underline{z}$ och $\underline{\varepsilon}$ $\doteq 0$

Varians-modell:

$$\underset{z}{V}[y(\underline{x}, \underline{z})] = \underbrace{\sigma_z^2 \cdot \sum_{i=1}^{2} \left[\frac{\partial y(\underline{x}, \underline{z})}{\partial z_i} \right]^2}_{\text{variansbidraget från brusvariablerna}} + s^2$$

$\sigma_z^2 = 1$ om brusvariablerna varieras en standardavvikelse ($s_z = \pm 1$) runt medelvärdena.

Uppgift 3

Faktorer: gasflöde (rad)
vätskeflöde (kolumn)

Respons (y): värmeöverföringskoefficient

Försöksdata och medelvärden: se nästa sida.

Summor:

$$SS_T = \sum\sum (y_{ij} - \bar{y}_{..})^2 = \sum\sum y_{ij}^2 - \frac{y_{..}^2}{N}$$

$SS_T = 373248$

$$SS(\text{gasflöde}) = b\sum(\bar{y}_{i.} - \bar{y}_{..})^2 = \frac{1}{b}\sum y_{i.}^2 - \frac{y_{..}^2}{N}$$

$SS(\text{gasflöde}) = 324082$

$$SS(\text{vätskeflöde}) = a\sum(\bar{y}_{.j} - \bar{y}_{..})^2 = \frac{1}{a}\sum y_{.j}^2 - \frac{y_{..}^2}{N}$$

$SS(\text{vätskeflöde}) = 39934$

$SS_E = SS_T - SS(\text{gasflöde}) - SS(\text{vätskeflöde}) = 9232$

df: $(a-1)(b-1)$ $N-1$ $a-1$ $b-1$

Tentamen 07-01-13 **Uppgift 3**

Försöksdata och beräknade medelvärden

→ j $b=4$

i ↓ $a=4$

A (gasflöde)	B (vätskeflöde) 1(190)	2(250)	3(300)	4(400)	$y_{i\cdot}$	$\bar{y}_{i\cdot}$	$\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}$
1(200)	200	226	240	261	927	231.75	−171.0625
2(400)	278	312	330	381	1301	325.25	−77.5625
3(700)	369	416	462	517	1764	441.00	39.1875
4(1100)	500	575	645	733	2453	613.25	210.4375
$y_{\cdot j}$	1347	1529	1677	1892			
$\bar{y}_{\cdot j}$	336.75	382.25	419.25	473.00			
$\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot}$	−66.0625	−20.5625	16.4375	70.1875			

$\sum\sum y_{ij} = y_{\cdot\cdot} = 6445$

$\bar{y}_{\cdot\cdot} = 402.81$

ANOVA

variation	SS	df	MS	F_{obs}
gasflöde	324082	3	108027	105.31 x
vätskeflöde	39934	3	13311	12.98 x
fel	9232	9	1026	
total	373248	15		

$F_{stat} = F(0.99, 3, 9) = 6.99$

A: gas- och vätskeflöden har signifikant effekt på värmeöverföringskoeff.

B: Signifikant skillnad mellan vätskeflödena testas med "least significant difference" - metoden.

$LSD = t_{stat} \sqrt{s^2(1/a + \frac{1}{a})}$

$t_{stat} = t(0.995, 9) = 3.2498$

$LSD = 3.2498 \cdot (1026 \cdot \frac{2}{4})^{1/2} = 73.60$

skillnad: medelvärden mellan vätskeflödena:

$\Delta \bar{y}_{j,j+1} = |\bar{y}_{\cdot j} - y_{\cdot j+1}|$

: Signifikant skillnad om

$\Delta \bar{y}_{j,j+1} > LSD$.

$\Delta \bar{y}_{1,2} = 45.5$ $\Delta \bar{y}_{2,3} = 37.0$ $\Delta \bar{y}_{3,4} = 53.25$

$\Delta \bar{y}_{1,3} = \underline{82.5}$ $\Delta \bar{y}_{2,4} = \underline{90.75}$

$\Delta \bar{y}_{1,4} = \underline{\underline{136.25}}$

Signifikant skillnad mellan flödena 1 och 4, något signifikant skillnad mellan flödena 1 och 3; 2 och 4.

Alternativt test:

Bilda testvariabeln:

$$t_{obs} = \frac{|\bar{y}_{\cdot j} - \bar{y}_{\cdot j+1}|}{\sqrt{s^2\left(\frac{1}{a} + \frac{1}{a}\right)}}$$

Signifikant skillnad om

$t_{obs} > t_{stat} = 3.25$

$t_{obs\,12} = 2.01$ $t_{obs\,23} = 1.63$ $t_{obs\,34} = 2.37$

$t_{obs\,13} = \underline{3.64}$ $t_{obs\,24} = \underline{4.01}$

$t_{obs\,14} = \underline{\underline{6.02}}$

C. Analysen visar ~~att~~ på en betydande skillnad mellan flödena 1 och 4.

Problem 4

Model: $\hat{y} = -1.15 + 0.05 z_1 + 0.2 z_2 + 0.1 z_3$

Restriction: $4z_1 + 5z_2 + 6z_3 \leq 371$

starting point: $\bar{z}_1 = 35 \quad (X_1 = 0)$

$\bar{z}_2 = 15 \quad (X_2 = 0)$

$\bar{z}_3 = 11 \quad (X_3 = 0)$

Gradient: $\underline{\nabla} y = \begin{bmatrix} 0.05 \\ 0.2 \\ 0.1 \end{bmatrix}$

$\underline{z}_0 = \begin{bmatrix} 35 \\ 15 \\ 11 \end{bmatrix} \qquad \Delta \underline{z} = k \cdot \begin{bmatrix} 0.05 \\ 0.2 \\ 0.1 \end{bmatrix}$

$\underline{z} = \underline{z}_0 + \Delta \underline{z}$ is inserted in the restriction

$4(35 + 0.05k) + 5(15 + 0.2k) +$
$+ 6(11 + 0.1k) \leq 371$

$1.8k = 371 - 281 = 90$

$k = 50$

$\underline{z}_{stop} = \begin{bmatrix} 35 + 50 \cdot 0.05 \\ 15 + 50 \cdot 0.2 \\ 11 + 50 \cdot 0.1 \end{bmatrix} = \begin{bmatrix} 37.5 \\ 25 \\ 16 \end{bmatrix}$

$x_{stop} = [0.5 \quad 2 \quad 1]^T$

Uppgift 5

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{12} x_1 x_2$$

$b = [50 \;\; 10 \;\; 5 \;\; -5 \;\; -3]^T$, 11 försök

$$\underline{X} = \begin{array}{ccccc} x_0 & x_1 & x_2 & x_3 & x_1 x_2 \end{array}$$

$$\underline{X} = \begin{bmatrix} 1 & -1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Beräkningsmatris

Momentmatris, inverterad:

$$(\underline{X}^T \underline{X})^{-1} = \begin{bmatrix} 1/11 = 0.0909 & & & & 0 \\ & 1/8 = 0.125 & & & \\ & & 1/8 & & \\ & & & 1/8 & \\ 0 & & & & 1/8 \end{bmatrix} = \underline{A}$$

A: Konfidensintervall för parametrar

$$s^2 = SS_{res}/(n-p) = \frac{8}{11-5} = \frac{8}{6} = \frac{4}{3} \approx 1.33$$

Konfint: $t(0.975, n-p) \cdot (a_{ii})^{1/2} \cdot s$
$= t \cdot se(b_i) = 2.447 \cdot se(b_i)$

$b_0 = 50 \pm 0.85$

$b_1 = 10 \pm 0.999 = 1$

$b_2 = 5 \pm 1$

$b_3 \quad -5 \pm 1$

$b_{12} \quad 3 \pm 1$

B. Konfidensintervall för prediktionen i [1 -1 1].

$$\underline{x_0} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}$$

$$\hat{y}(x_0) = 50 + 10 - 5 - 5 - 3 = 47$$

$$\eta = \hat{y}(x_0) \pm t(0.975, n-p) \cdot (s^2(1 + x_0^T (X^T X) \cdot x_0)^{1/2}$$

$$x_0^T (X^T X)^{-1} \cdot x_0 =$$

$$\begin{bmatrix} 1 & 1 & -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1/11 & & & & \\ & 1/8 & & & \\ & & 1/8 & & \\ & & & 1/8 & \\ & & & & 1/8 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} =$$

$$= \begin{bmatrix} 1/11 & 1/8 & -1/8 & 1/8 & -1/8 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} =$$

$$= 1/11 + 1/8 + 1/8 + 1/8 + 1/8 = 1/11 + \frac{4}{8} = 0.5909$$

$$\eta = \hat{y}(x_0) \pm 2.447 \left[\frac{4}{3} (1 + 0.5909) \right]^{1/2}$$

$$\underline{\eta = \hat{y}(x_0) \pm 3.5639}$$

$$\underline{\underline{\eta = 47 \pm 3.6}}$$

Problem 6 07-01-13

Reasons to large confidence intervals:

1. Bad experiments

Bad experiments give large s^2, which can be caused by "outliers". These can be seen in a residual plot. Points with a standardized residual $-3 \geq \frac{e_i}{\sigma} \geq 3$ can be suspected to be outliers. Outliers can depend on experimental errors but even on real effects. Possibly the experiments that gave outliers can be repeated.

Another reason can be too few observations $\Rightarrow$ too few degrees of freedom $\Rightarrow$ large t-value.

2. Bad model

Large "lack of fit" sum of squares (SS_{LOF}) $\Rightarrow$ large s^2. Can be tested by repeated runs and following significance study.

Calculate R^2 and R^2_{adj}. Large difference $R^2 - R^2_{adj}$ indicates that there are not significant parameters in the model.

Patterns in the residual plots (e_i vs y_{mod}, x_i) show if terms is missing in the model.

3. Bad design matrix

Near singular $\mathbf{X}^T\mathbf{X}$ –matrix indicates that two or more variables covariate.

If the study of the correlation matrix $(\mathbf{X}^T\mathbf{X})^{-1}$ shows that the correlation between two parameters is high(>0.8) , then the design matrix should be changed (moved).

An simple method to study the influence of separate observation points is to calculate the diagonal elements in the hat matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$.
If $h_{ii} > 2p/n$, then the observation i is a high leverage point.