Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Contact during exam:
Bo Lindqvist
Tel. 975 89 418

ENGLISH

EXAM IN TMA4255
EXPERIMENTAL DESIGN AND APPLIED STATISTICAL METHODS

Friday 5 December 2008

Time: 09:00–13:00

Aids: All printed and handwritten aids. Special calculator permitted.

Grading: 5 January 2009

Problem 1

The yield of a chemical process was studied in a pilot experiment. The following factors were considered:

Factor		Factor levels	
		-1	1
A	Amount of active compound	4 mol	5 mol
B	Acidity, pH	6	7
C	Reaction time	2 hours	4 hours
D	Filtering (first pass)	none	after 1/2 hour
E	Filtering (second pass)	none	after 1 hour

A fractional 2^{5-2} experiment was performed, based on a full 2^3-experiment with factors A, B, C and generators $D = AB$ and $E = AC$.

The response Y was defined as yield measured relative to a theoretical maximum. The responses $Y_1, \ldots, Y_8$ of the 8 experiments are assumed to be independent and normally distributed with the same variance σ^2.

The MINITAB-output on the next page shows the design and the 8 responses, in addition to the alias structure and the estimated effects. You may use this output when you solve this Problem.

```
Data Display

Row  StdOrder   A   B   C   D   E     Y
  1         1  -1  -1  -1   1   1  69,3
  2         2   1  -1  -1  -1  -1  70,8
  3         3  -1   1  -1  -1   1  71,3
  4         4   1   1  -1   1  -1  73,2
  5         5  -1  -1   1   1  -1  77,5
  6         6   1  -1   1  -1   1  79,3
  7         7  -1   1   1  -1  -1  88,9
  8         8   1   1   1   1   1  91,2

Alias Structure

I + ABD + ACE + BCDE

A + BD + CE + ABCDE
B + AD + CDE + ABCE
C + AE + BDE + ABCD
D + AB + BCE + ACDE
E + AC + BCD + ABDE
BC + DE + ABE + ACD
BE + CD + ABC + ADE

Estimated Effects and Coefficients for Y (coded units)

Term        Effect     Coef
Constant            77,6875
A           1,8750   0,9375
B           6,9250   3,4625
C          13,0750   6,5375
D           0,2250   0,1125
E           0,1750   0,0875
B*C         4,7250   2,3625
B*E         0,0250   0,0125
```

a) An estimate of the main effect of factor A is wanted. Explain how this estimate is computed from the given responses. Give a comment in light of the given alias structure.

Demonstrate how the two-factor interaction BC can be computed from the given responses and give a comment. What is in general the interpretation of a two-factor interaction?

What are the defining relations and what is the resolution of the design in this Problem? What is the practical interpretation of the resolution?

It was assumed in the analysis that the factors D and E do not interact with each other or with the other factors in the experiment, i.e. that all interactions where D or E participate can be set to 0. This assumption is made for the rest of the Problem.

b) Which main effects can now be estimated in an unconfounded manner, and which main effects are still confounded with other effects?

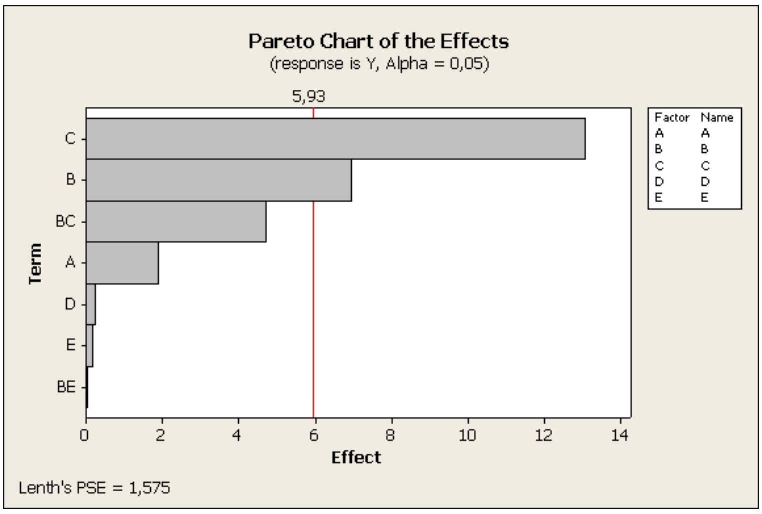
Find the variance of an estimated effect, expressed by σ^2.

Based on earlier experience, the value of σ^2 was assumed known, $\sigma^2 = 2$. Use this to select the main effects and interactions that are significant when the significance level is chosen to be 5%.

What is the lowest significance level that would lead to A having a significant effect?

c) The Pareto-chart shown below uses Lenth's PSE as an estimate of the standard deviation of the estimated effects.

Show, using the estimated effects in the MINITAB-output, how one gets the result PSE = 1.575.



Problem 2

A research lab will investigate the vibration which occurs when thin plates of polyethylen are exposed to wind.

The experiment uses plates of equal thickness and width, but with three different lengths. The plates are exposed to four different wind speeds, and the number of oscillations in a certain time period are measured. Two experiments are performed for each of the 12 combinations of length and wind speed.

The responses of the individual experiments are denoted Y_{ijk} and are given as one hundredth of the number of oscillations per second when plate no. k with length x_{1i} (factor level A_i) is exposed to wind of speed x_{2j} (factor level B_j); for $i = 1, 2, 3;\ j = 1, 2, 3, 4$ and $k = 1, 2$.

The result of the experiment is given in the table below. The unity for length is *inches*, while the unity for wind speed is *feet/second*.

Length	Wind speed (x_{2j})			
(x_{1i})	$B_1 = 62.5$	$B_2 = 54.6$	$B_3 = 44.3$	$B_4 = 31.3$
$A_1 = 1.50$	50.5	46.0	36.5	23.0
	50.0	45.1	37.0	24.5
$A_2 = 1.75$	47.0	41.5	33.1	22.0
	48.0	42.0	34.1	24.2
$A_3 = 2.00$	45.5	39.4	30.8	20.3
	45.1	38.8	31.0	21.6

The first analysis of the data is by two-factor analysis of variance, where the factors A and B have, respectively, 3 and 4 levels. You may in the problems below use the following output from MINITAB:

Two-way ANOVA: Y versus A; B

```
Source       DF       SS       MS        F
A             2   100,54   50,268    93,52
B             3  2145,40  715,134  1330,48
Interaction   6     8,99    1,498     2,79
Error        12     6,45    0,537
Total        23  2261,38
```

a) Write down the model for the observations $\{Y_{ijk}\}$ which takes into account a possible interaction between the length effect (A) and the wind speed effect (B).

How can you test whether there is an interaction between the two factors? Write down and motivate a test statistic for this, and then find the critical value when the significance level is set to 5%. What is the conclusion?

Are the main effects of length and wind speed significant at 5% level? Answer the question by referring to the values of the test statistics given in the MINITAB output.

b) Write down the expressions for SSE and $SS(AB)$ in the model from question (a).

Show, using results from the course, that SSE/σ^2 is chi-square distributed with 12 degrees of freedom. (*Hint:* Use that if $V_1, \ldots, V_n$ are independent observations from $N(\nu, \tau^2)$, then $\sum_{k=1}^{n}(V_k - \bar{V})^2/\sigma^2$ is chi-square distributed with $n-1$ degrees of freedom).

Write down an expression for the estimator S^2 for σ^2 based on SSE. Find the numerical value of S^2 in the MINITAB-output.

Assume now that there is no interaction between the factors A and B, i.e. that alle the parameters $(\alpha\beta)_{ij}$ in the model in question (a) equal 0. It can then be shown that $SS(AB)/\sigma^2$ is chi-square distributed with 6 degrees of freedom, and that $SS(AB)$ is independent of SSE. (You are not asked to do this).

Show how a new estimator of σ^2 can be derived by using both SSE and $SS(AB)$. Find the estimator and compute the new estimate for σ^2 with the given data.

Problem 3

The situation is the same as in Problem 2. It is now assumed that

$$Y_{ijk} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2j} + \epsilon_{ijk} \tag{1}$$

for $i = 1, 2, 3;\ j = 1, 2, 3, 4$ and $k = 1, 2$.

The output from a regression analysis using MINITAB, based on the data in Problem 2, is given on the next page. You may use this when solving the Problem.

a) Explain briefly how this model can be treated as a multiple linear regression model with 24 observations and explanatory variables x_1 = length and x_2 = wind speed.

Which assumptions are made in the model (1)?

What portion of the variation in the data is explained by the model? Is a significant amount of variation explained?

One will test the hypotheses $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$ and $H_0: \beta_2 = 0$ vs. $H_1: \beta_2 \neq 0$ in this model.

Write down the relevant test statistics and perform the two tests with significance level 1%.

To what extent do the results of this testing support the conclusion from the analysis of variance in Problem 2?

What are similarities and differences between the model in this Problem and the model of Problem 2?

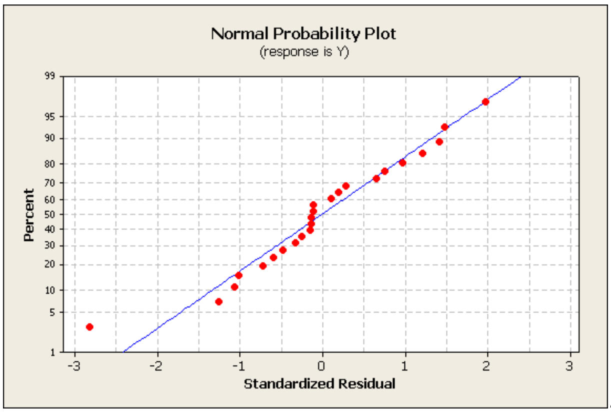
Regression Analysis: Y versus x1; x2

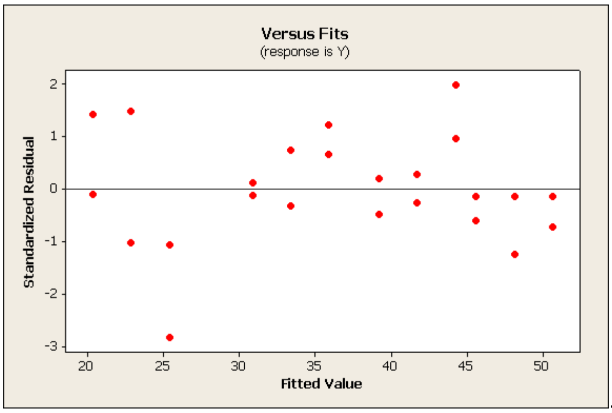
```
The regression equation is
Y = 15,1 - 10,0 x1 + 0,808 x2
```

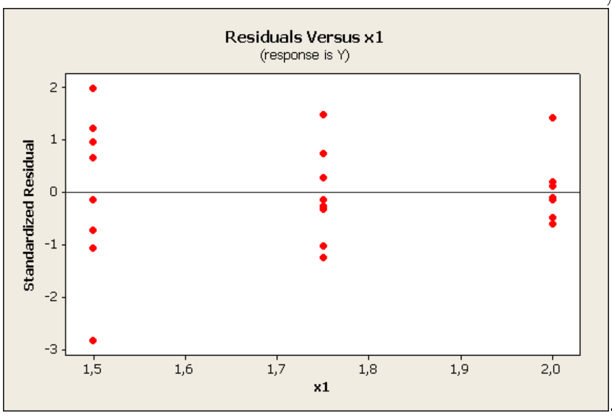
Predictor	Coef	SE Coef	T
Constant	15,140	1,848	8,19
x1	-10,0250	0,9466	-10,59
x2	0,80842	0,01653	48,89

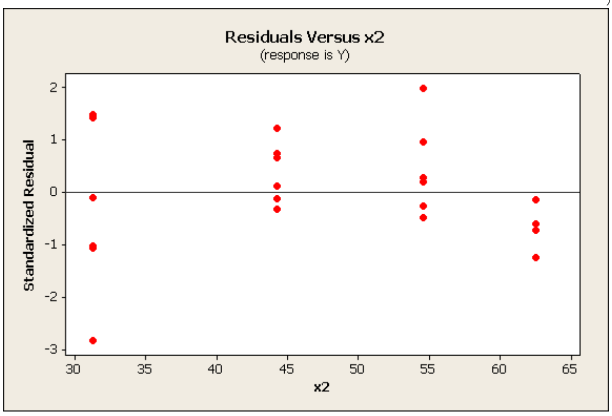
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	2242,6	1121,3	1251,40	0,000
Residual Error	21	18,8	0,9		
Total	23	2261,4			



,


,


,


b) Four different residual plots are given on the previous page.

What is here meant by standardized ("studentized") residuals?

Explain briefly what is being plotted in each diagram, and give an evaluation of the results with respect to fit between data and model.

Can you suggest simple extensions of the model which may possibly improve the fit? No new analyses are required.

Problem 4

Let X be the number of million revolutions to fatigue failure in testing of a type of ball bearings. 65 such ball bearings were tested, with results summarized in the table below.

Interval	Number of observations
$< 0, 40]$	7
$< 40, 60]$	14
$< 60, 80]$	18
$< 80, 100]$	15
$< 100, 120]$	8
$< 120, \infty]$	3

The goal was to find out whether the probability distribution of X can be assumed to be a Weibull distribution with cumulative distribution function

$$F(x) = P(X \leq x) = 1 - e^{-\left(\frac{x}{80}\right)^2} \text{ for } x > 0. \tag{2}$$

a) Use the data in the table to perform a test of the null hypothesis that X has the distribution (2). Let the significance level be 5%. What is the conclusion?

Explain briefly how you would proceed if the null hypothesis were that X is Weibull distributed, without specified parameters, and you had the same data. You are not asked to do any new computations here.

(The Weibull-distribution is generally given by a cumulative distribution function

$$F(x; \theta, \alpha) = 1 - e^{-(\frac{x}{\theta})^\alpha} \text{ for } x > 0,$$

where α, θ are positive parameters.)