NTNU
Noregs teknisk-naturvitskaplege universitet

Fakultet for informasjonsteknologi, matematikk og elektroteknikk
Institutt for matematiske fag

English

Contact during exam
John Tyssedal tlf. 73593534

EXAM IN TMA4255 DESIGN OF EXPERIMENTS AND APPLIED STATISTICAL METHODS.

August 9 2004
Time: 09.00 – 14.00

Helping aids: All printed and handwritten material is permitted. Defined simple calculator is allowed.

Examination results are due September 1.

Problem 1.

For wire used for certain types of cranes it is important that the wearing strength is as large as possible. Two types of wires are under investigation. One from factory A that has been used up till now and one from a new factory B. In the investigation seven randomly selected wires from factory A are tested on a machine and the time in hours until the wires break (X_i), are measured.
The results are: 4.5 2.2 6.6 3.8 7.9 3.0 1.8

For eight randomly selected wires from factory B the similar measured times (Y_j) are:
7.0 5.9 5.1 8.4 7.2 2.7 7.5 8.2.

One wants to find out if there are reasons to conclude that the time until the wires break in general is shorter for wires from factory A than they are for wires from factory B.

a) State this as a hypothesis test. An analysis performed with MINITAB is shown below:

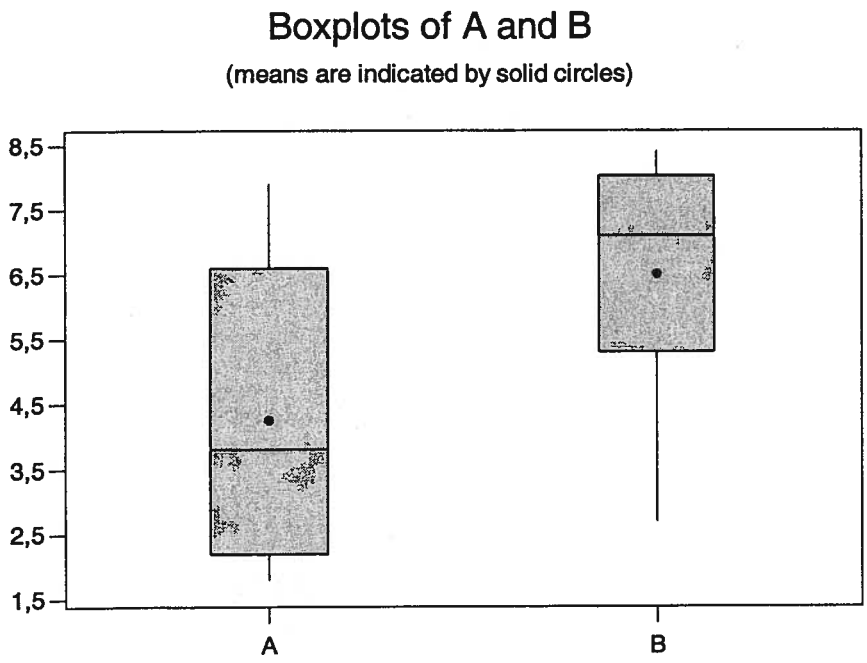
Two-Sample T-Test and CI: A; B

```
Two-sample T for A vs B

   N      Mean      StDev   SE Mean
A  7      4,26       2,27      0,86
B  8      6,50       1,89      0,67

Difference = mu A - mu B
Estimate for difference:  -2,24
95% upper bound for difference: -0,34
T-Test of difference = 0 (vs <): T-Value = -2,09  P-Value = 0,028  DF = 13
Both use Pooled StDev = 2,07
```

Give the assumptions for the test above. Write down the expression for the test statistics. Explain the P-value in this situation and give the conclusions for the test. Use a 5% level of significance.

The test above can also be performed by means of a Wilcoxon two-sample test (Mann-Whitney test).

b) Perform the test. What is the conclusion when the same level of significance as in 1a) is used? Study the box-plot given below and judge if there is any reason to chose one test instead of the other for analyzing these data.



Problem 2.

A person can choose between four alternative routes for driving to work. To get some knowledge about expected amount of time needed to drive the different routes he wrote down the time needed to drive to work each day for four weeks such that each route was tried out once on each working day. For each day the order in which the four routes are driven is randomly selected. The observed results are given in the table below.

Table					
Day\Route	Route 1	Route 2	Route 3	Route 4	Average
Monday	22	25	26	26	24.75
Tuesdag	26	27	29	28	27.50
Wednesday	25	28	33	27	28.25
Thursday	25	26	30	30	27.75
Friday	31	29	33	30	30.75
Average	25.8	27	30.2	28.2	27.80

A two-way analysis of variance was then performed on the data. Output from MINITAB follows below:

Two-way ANOVA: Time versus day; route

```
Analysis of Variance for time
Source        DF        SS        MS        F        P
day            4     73,20     18,30     8,07    0,002
route          3     52,80     17,60     7,76    0,004
Error         12     27,20      2,27
Total         19    153,20
```

a) Explain why it is reasonable that the day-effect in this case is a block-effect. What is the name of this experimental set up? Write down the model and give the assumptions the analysis is based on. State the hypothesis. What are the conclusions from this experiment? Use a 5% level of significance.

b) The person wanted to find out for which routes the driving time was significantly shorter than for the others. Perform a test to investigate that.

Problem 3.

When cement solidifies, warmth is transferred to the surroundings. To investigate how this process is affected by the amount of four chemicals x_1, x_2, x_3 and x_4 contained in the cement, thirteen measurements as shown in the table on the next page was taken. y (calories pr. gram) is the amount of warmth transferred.

Obs	y	x_1	x_2	x_3	x_4
1	78,5	7	26	6	60
2	74,3	1	29	15	52
3	104,3	11	56	8	20
4	87,6	11	31	8	47
5	95,9	7	52	6	33
6	109,2	11	55	9	22
7	102,7	3	71	17	6
8	72,5	1	31	22	44
9	93,1	2	54	18	22
10	115,9	21	47	4	26
11	83,8	1	40	23	34
12	113,3	11	66	9	12
13	109,4	10	68	8	12

The correlation matrix between the regression variables is shown below.

Correlations: x_1; x_2; x_3; x_4

	x_1	x_2	x_3
x_2	0,229		
x_3	-0,824	-0,139	
x_4	-0,245	-0,973	0,030

a) Does this correlation matrix show any properties that may be an disadvantage when more than one regression variable is to be included in the model? Explain your answer.
Below an output from a regression analysis with MINITAB is shown.

Regression Analysis: varme versus x_1; x_2; x_3; x_4

The regression equation is

varme = 62,4 + 1,55 x_1 + 0,510 x_2 + 0,102 x_3 - 0,144 x_4

Predictor	Coef	SE Coef	T	P
Constant	62,41	70,07	0,89	0,399
x_1	1,5511	0,7448	2,08	0,071
x_2	0,5102	0,7238	0,70	0,501
x_3	0,1019	0,7547	0,14	0,896
x_4	-0,1441	0,7091	-0,20	0,844

S = 2,446 R-Sq = 98,2% R-Sq(adj) = 97,4%
PRESS = 110,347 R-Sq(pred) = 95,94%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	2667,90	666,97	111,48	0,000
Residual Error	8	47,86	5,98		
Total	12	2715,76			

Source	DF	Seq SS
x_1	1	1450,08
x_2	1	1207,78
x_3	1	9,79
x_4	1	0,25

How much of the variation in the data is explained by this model? Is the regression significant? Do you find any regression variables to be significant given that the others are in the model? (When answering these two questions also state the hypotheses that are tested out and explain your answers. Choose your own significance level.)

A new regression analysis where only x_1 and x_2 are included was then performed. The output from MINITAB is shown below:

The regression equation is

varme = 52,6 + 1,47 x_1 + 0,662 x_2

Predictor	Coef	SE Coef	T	P
Constant	52,577	2,286	23,00	0,000
x_1	1,4683	0,1213	12,10	0,000
x_2	0,66225	0,04585	14,44	0,000

S = 2,406 R-Sq = 97,9% R-Sq(adj) = 97,4%
PRESS = 93,8825 R-Sq(pred) = 96,54%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	2657,9	1328,9	229,50	0,000
Residual Error	10	57,9	5,8		
Total	12	2715,8			

Source	DF	Seq SS
x_1	1	1450,1
x_2	1	1207,8

b) Compare the different criteria for variable selection for the two models in 3a) and 3b) Which of the models will you prefer? Calculate Mallows C_p for the two models.

Problem 4

a) In a 2^{5-2} experiment in the variables A, B, C, D and E, D=AB and E=-AC. Find the defining relations and which two-factor interactions that are confounded with the main effects. What is the resolution of the design?

b) Show how it is possible to perform another 2^{5-2} design in such a way that one can estimate the main effects unconfounded with two-factor interactions when both fractions are used. Assume that the interactions of order higher than two are negligible. Then it is possible to estimate 4 unconfounded two-factor interactions with the two fractions. Find these four two-factor interactions.

The two fractions in 3a) and 3b) may be used to construct a half-fraction of a 2^5 design. What is the defining relation of this half-fraction?