English
Contact during exam
 John Tyssedal
tlf. 73593534

## EXAM IN  TMA4255 DESIGN OF EXPERIMENTS AND APPLIED STATISTICAL METHODS.

Monday  May 22. 2006
Time: 09.00 – 13.00

Helping aids: All printed and handwritten material. All calculators allowed.

Examination results are due June 13.

## Problem 1

A course in economics was taught to two groups of students, one in a classroom situation and the other by TV. There were 24 students in each group. These students were paired into subgroups of size 2 according to cumulative grade point averages and background in economics, and then assigned to the courses by a flip of coin. This was repeated 24 times. At the end of the course each class was given the same final examination and each student was given a score from 1 to 100.  The aim of the experiment was to find out if there was any difference in the teaching methods.

Let $X_i$, $i = 1, 2, \ldots, 24$ be the score of a student in subgroup $i$ for those who had classroom teaching and let $Y_i$, $i = 1, 2, \ldots, 24$ be the score in subgroup $i$ for those who had TV teaching. Let $D_i$ be $X_i - Y_i$. The observed values for $D_i$ were:

$d_i$ : 14 -4 -6 -2 -1 18 6 12 8 -4 13 7 2 6 21 7 -2 11 -3 -14 -2 17 -4 -5

Output from an analysis with MINITAB is given below.

## One-Sample T: C1

```
Test of mu = 0 vs not = 0


Variable            N     Mean     StDev   SE Mean        95% CI           T      P
Scoredifference    24   3,95833   9,02402  1,84202   (0,14782; 7,76884)  2,15  0,042
```
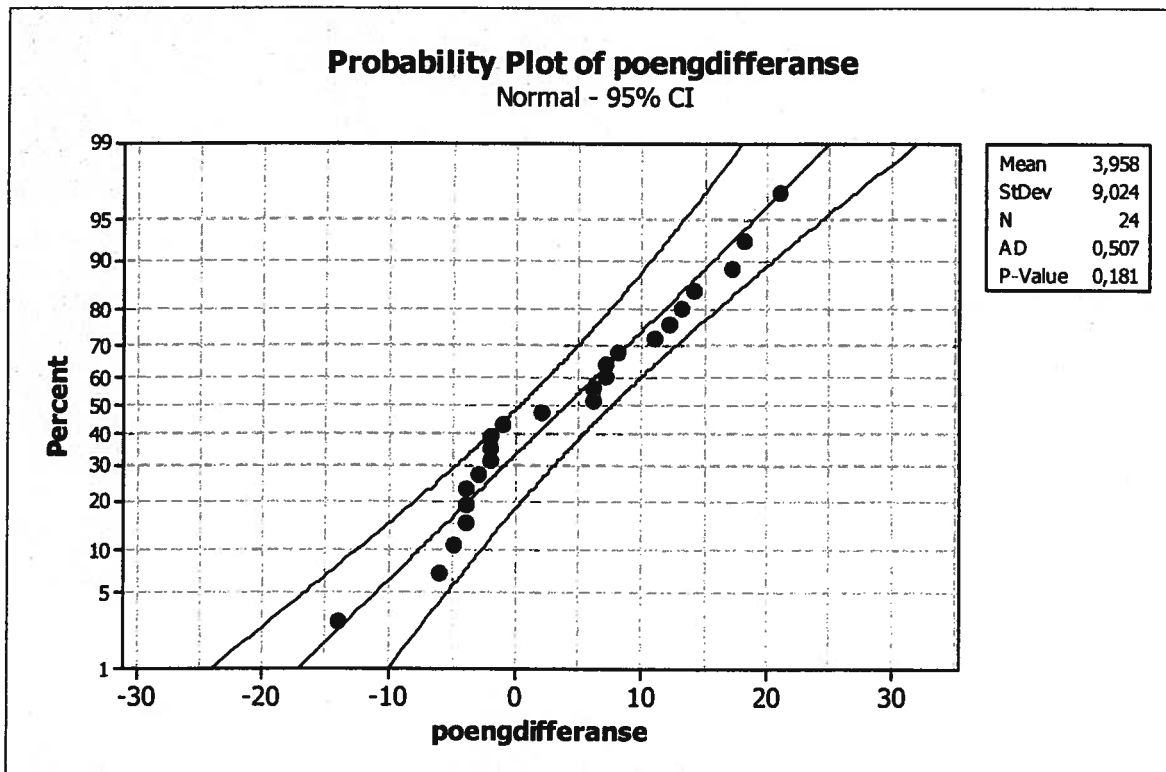
a) What type of experiment is performed in this investigation? What assumptions is needed for the analysis above? State the hypothesis test. Write down the expression for the test statistics. What is the P-value in this experiment? What is the conclusion when a 5% level of significance is used.

There were some doubts about the realism in the assumptions for the test above and it was therefore suggested to perform a Wilcoxon signed-rank test on these data.

b) Use the first twelve obervations of differences in scores to perform a Wilcoxon signed- rank test. State the hypothesis test. What is the conclusion when a 5% level of significance is used?
A normal probability plot together with a MINITAB output from Wilcoxon signed-rank test based on all the 24 observations are given below.



**Probability Plot of poengdifferanse**
Normal - 95% CI

| Mean | 3,958 |
| StDev | 9,024 |
| N | 24 |
| AD | 0,507 |
| P-Value | 0,181 |

**Wilcoxon Signed Rank Test: C1**

```
Test of median = 0,000000 versus median not = 0,000000

                         N
                        for      Wilcoxon              Estimated
                  N     Test     Statistic      P      Median
Scoredifference   24    24         216,0      0,061    4,000
```

Evaluate if you think there are reasons to conclude that the two teaching methods give different result.


## Problem 2

A fisherman conducted a $2^5$ experiment in order to determine the effects of 5 factors on the number of fish he catches in a four-hour period. The 5 factors studied A, B, C, D and E, and their levels are given in the table below.

| Faktorer | Lavt nivå ( - ) | Høyt nivå ( + ) |
|---|---|---|
| A: Fishing location | Pier | Boat |
| B: Bait type | Worms | Night crawlers |
| C: Time of day | Day | Night |
| D: Weather | No rain | Rain |
| E: Hook size | Small | Large |

The table below shows the result from a $2^{5-2}$ fraction of the experiment written down in standard form. Y is the number of fish.

| A | B | C | D | E | y |
|---|---|---|---|---|---|
| - | - | - | + | + | 10 |
| + | - | - | - | - | 26 |
| - | + | - | - | + | 26 |
| + | + | - | + | - | 43 |
| - | - | + | + | - | 30 |
| + | - | + | - | + | 20 |
| - | + | + | - | - | 32 |
| + | + | + | + | + | 53 |

a) Calculate an estimate of the main effect of factor A and the two factor interaction AD based on this fraction. What is the defining relation? What is the estimate of the main effect of factor B?

b) What is the resolution of the experiment? Explain your answer. Write down the alias structure that shows how main effects are confounded with two factor interactions (neglect effects of order 3 and higher). Is it possible to augment this experiment with eight new runs such that the 16 runs thus obtained constitute a half fractions of a $2^5$ design with resolution 5? Explain your answer.

An analysis performed with MINITAB on the complete $2^5$ experiment is shown below.

Estimated effects and coefficients for the number of fish.

```
Term           Effect      Coef
Constant                 29,875
A              10,625     5,313
B              15,875     7,937
C               0,000     0,000
D               7,375     3,687
E              -6,375    -3,188
A*B             0,250     0,125
A*C             0,125     0,062
A*D             0,500     0,250
A*E             7,500     3,750
B*C             0,375     0,188
B*D             0,500     0,250
B*E             7,500     3,750
C*D            -0,375    -0,187
C*E             0,125     0,062
D*E            -0,250    -0,125
A*B*C           0,500     0,250
A*B*D          -0,375    -0,187
A*B*E           0,375     0,188
A*C*D          -0,000    -0,000
A*C*E           0,250     0,125
A*D*E          -0,125    -0,063
B*C*D           0,000     0,000
B*C*E          -0,250    -0,125
B*D*E           0,125     0,063
C*D*E           0,250     0,125
A*B*C*D        -0,125    -0,063
A*B*C*E         0,375     0,187
A*B*D*E         0,750     0,375
A*C*D*E         0,625     0,313
B*C*D*E         0,375     0,187
A*B*C*D*E      -0,250    -0,125
```

S = *

Analysis of Variance for C30 (coded units)

| Source | DF | Seq SS | Adj SS | Adj MS | F | P |
|---|---|---|---|---|---|---|
| Main Effects | 5 | 3679,50 | 3679,50 | 735,900 | * | * |
| 2-Way Interactions | 10 | 907,50 | 907,50 | 90,750 | * | * |
| 3-Way Interactions | 10 | 6,00 | 6,00 | 0,600 | * | * |
| 4-Way Interactions | 5 | 10,00 | 10,00 | 2,000 | * | * |
| 5-Way Interactions | 1 | 0,50 | 0,50 | 0,500 | * | * |
| Residual Error | 0 | * | * | * | | |
| Total | 31 | 4603,50 | | | | |

c) Assume interactions of order three and higher has expectation 0 and use the output to calculate an estimate for the variance of the effects. Decide what effects that are significant. Write down the estimated model and decide on what is the best strategy when he fishes from pier or from a boat. What is the best way of fishing if he only has worms as bait type?

## Problem 3

The VO2 index is used as a measure of aerobic fitness. For a given person this is found measuring the maximum amount of oxygen in the inhaled air pr. unit time that is used by the individual and thereafter scaling this number by the person's weight. One wants to find out how this index depends on some explanatory variables such as "heart rate at rest", "age", percentage body fat and weight (in lbs). Collected data from 20 men are given below.

| Person | VO2 | heart rate at rest | age | %body fat | weight |
|--------|-----|--------------------|-----|-----------|--------|
| 1 | 23 | 62 | 59 | 26 | 182,0 |
| 2 | 45 | 59 | 47 | 18 | 175,0 |
| 3 | 29 | 82 | 44 | 22 | 200,0 |
| 4 | 55 | 61 | 32 | 10 | 168,5 |
| 5 | 48 | 60 | 45 | 19 | 193,0 |
| 6 | 42 | 58 | 61 | 22 | 170,0 |
| 7 | 32 | 76 | 71 | 28 | 193,0 |
| 8 | 33 | 70 | 32 | 23 | 218,0 |
| 9 | 34 | 68 | 28 | 27 | 228,0 |
| 10 | 52 | 76 | 36 | 10 | 128,0 |
| 11 | 40 | 67 | 36 | 18 | 167,0 |
| 12 | 35 | 66 | 51 | 29 | 194,0 |
| 13 | 45 | 50 | 31 | 29 | 219,0 |
| 14 | 47 | 57 | 44 | 13 | 215,0 |
| 15 | 26 | 61 | 73 | 28 | 246,0 |
| 16 | 42 | 51 | 47 | 19 | 171,0 |
| 17 | 35 | 60 | 40 | 25 | 212,0 |
| 18 | 41 | 63 | 43 | 16 | 167,0 |
| 19 | 29 | 66 | 68 | 22 | 162,0 |
| 20 | 38 | 57 | 40 | 28 | 239,0 |

Parts from an output from a regression analysis performed by MINITAB is given below.

```
Predictor          Coef   SE Coef      T       P
Constant          95,06     14,65    6,49   0,000
heart rate      -0,3623    0,1500      *    0,029
age            -0,21471         *   -2,19   0,045
%bodyfat              *     0,2882   -2,69   0,017
weight         -0,03530    0,05697   -0,62   0,545


S = *   R-Sq = *   R-Sq(adj) = 64,5%

PRESS = 713,673   R-Sq(pred) = 50,06%


Analysis of Variance

Source           DF       SS      MS      F       P
Regression        4   1028,45  257,11   9,63   0,000
Residual Error   15    400,50   26,70
Total            19   1428,95


Source           DF   Seq SS
heart rate        1   135,58
age               1   368,76
%bodyfat          1   513,86
weight            1    10,25
```

a) Fill in the numbers where there is only given a * in the output. Explain the assumptions you do and evaluate if the regression is significant on a 5% level of significance (state the hypothesis test and conclude). How much of the variation in the data is explained by the model?

Output from a new regression analysis performed by MINITAB is given below.

```
The regression equation is
VO2 = 88,8 - 0,343 heart rate - 0,195 age - 0,901 %body fat


Predictor          Coef  SE Coef      T      P
Constant          88,82    10,44   8,51  0,000
heart rate      -0,3428   0,1438  -2,38  0,030
age            -0,19496  0,09084  -2,15  0,048
%body fat       -0,9007   0,2013  -4,47  0,000


S = 5,06673    R-Sq = 71,3%    R-Sq(adj) = 65,9%

PRESS = 697,690    R-Sq(pred) = 51,17%


Analysis of Variance

Source          DF        SS      MS      F      P
Regression       3   1018,20  339,40  13,22  0,000
Residual Error  16    410,75   25,67
Total           19   1428,95


Source      DF  Seq SS
hvilepuls    1  135,58
alder        1  368,76
%kroppsfett  1  513,86
```

b) Which of the models in the outputs before 3a) and 3b) will you prefer? Explain your answer. Calculate the value of Mallows $C_p$ for the model in the output before 3b).

Use the model you evaluate to be the best one to predict the VO2 index of a man with "heart rate at rest" equal to 60, age 45, 20% body fat and a weight of 200 lbs.

c) A man trains in order to improve his VO2 index. He measures his "heart rate at rest" daily. Estimate how much he needs to reduce his "heart rate at rest" in order to improve his VO2 index by 2 when it is assumed that the percentage body fat is kept unchanged. Give an estimate of the standard deviation of your obtained estimate.