Department:	**Chemistry, biotechnology and food sciences**	
Examination in:	**STAT210**	**Experimental Design and Analysis of Variance I**
	Course code	*Course name*
Time for exam:	**Monday 3th of Sept 2012**	**14:00-17:30 (3,5 hours)**
	Day and date	*As from – to and duration of examination (hours)*
Person responsible for the course:	**Thore Egeland, 41479582**	
	Name and telephone	

C3: all types of calculators, all other aids

The exam includes: 6 pages
Number of pages incl. attachment

Each sub-question will be given the same score in the evaluation of the exam. You may answer in English or Norwegian (or "Scandinavian"). There are three exercises.

Exercises 1 and 2 below are based on the data reproduced in Appendix 1. For this study 6 school classes were randomly chosen. From each class, 2 girls and two boys were randomly randomly chosen. The purpose of the study was to study the effect of three different teaching methods. The response variable **Score** measured how well a pupil did on a test at the end of the study. The variables are summarized below:

- **Method:** This variable indicates one of three possible teaching methods and is coded: 'Method1', 'Method2', and 'Method3'.
- **Class:** This variable indicates which of six classes the pupils come from and is coded 'Class1',...,'Class6'.
- **Gender:** This variable is coded 'Girl' or 'Boy'.
- **Score:** The worst possible value is 0. The best possible value is 10.

Exercise 1

In this exercise we will only be using the variables **Class** and **Score.** We consider the following random effect model:

$$y_{ij} = \mu + \tau_i + e_{ij} \text{ where}$$
$i = 1, 2, \cdots, 6$ correspond to 'Class1', 'Class2' ..., 'Class6',
$j = 1, 2, 3, 4$ correspond to the four observations from each class,
$y_{ij} =$ **Score**, $\tau_i \sim N(0, \sigma_\tau^2)$ and $e_{ij} \sim N(0, \sigma^2)$.
All variables τ_i and e_{ij} are independent.

a) We would like find out if the scores vary significantly between classes, i.e, if we can reject

$$H_0 : \sigma_\tau^2 = 0.$$

Use the output in Appendix 2 to perform the test at 5% significance level. Formulate a conclusion.

b) Use the output of Appendix 2 to estimate σ_τ^2 and σ^2. Calculate a 95% confidence interval for σ^2 and interpret the answer.

c) Estimate the correlation between two observations from the same class.

Exercise 2

In this exercise we will be using the variables **Method, Gender** and **Score**. We use the following fixed effect model

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + e_{ijk} \text{ where}$$
$i = 1, 2, 3$ correspond to 'Method1','Method', 'Method3',
$j = 1, 2$ correspond to 'Boy', 'Girl',
$k = 1, 2, 3, 4$ correspond to replications,
$y_{ijk} =$ **Score** and $e_{ijk} \sim N(0, \sigma^2)$,

$$\sum_{i=1}^{3} \tau_i = 0, \quad \sum_{j=1}^{2} \beta_j = 0, \quad \sum_{i=1}^{3} (\tau\beta)_{ij} = \sum_{j=1}^{2} (\tau\beta)_{ij} = 0 \ .$$

The random variables e_{ijk} are assumed to be independent.

a) Use the output of Appendix 3 to determine if there is an interaction between **Method** and **Gender.** Formulate the hypotheses and a conclusion. Use 5% significance level.

b) For the remaining part of the exercise we will be using the following reduced model:

$$y_{ij} = \mu + \tau_i + e_{ij} \text{ where}$$

$i = 1, 2, 3$ correspond to the three (a=3) groups 'Method1','Method2', 'Method3',

$j = 1, 2..., 8$ correspond to the eight (n=8) observations for each method,

$y_{ij} =$ **Score** and $e_{ij} \sim N(0, \sigma^2)$,

$$\sum_{i=1}^{3} \tau_i = 0.$$

The random variables e_{ij} are assumed to be independent.

Use the output of Appendix 4 to show that there are significant differences depending on teaching method at 5% significance level. Between which methods are there significant differences? Use Appendix 4 and Tukey's test.

c) Use Appendix 5 to estimate τ_1, τ_2 and $\tau_{3.}$

d) Comment on the independence assumption in b) above, i.e., the assumption "The random variables e_{ij} are assumed to be independent". Also comment on the other standard assumptions of the model in b). You may refer to Figure 1 in your answer.

Exercise 3.

The purpose of this exercise is to determine if a new treatment (called 'M1') designed to help people loose weight is better than two well known methods (called 'M2' and 'M3'). Two individuals were recruited in each group. The individuals were weighed at the beginning and the end of the study. For each individual the weight difference ('final weight'-'initial weight') was recorded. The data and some summary statistics are as follows:

Treatments	Observations		Averages
M1	0	2	1
M2	1	3	2
M3	5	7	6
			Total average=3

We will use the following fixed effect model

$y_{ij} = \mu + \tau_i + e_{ij}$ where

$i = 1, 2, 3$ correspond to 'M1','M2' and 'M3'

$j = 1, 2$ correspond to the two observations in each treatment group

$y_{ij} =$ **Weight difference**

and $e_{ij} \sim N(0, \sigma^2)$,

$\sum_{i=1}^{3} \tau_i = 0.$

The random variables e_{ij} are assumed to be independent.

a)

Calculate $SS_{Treatments}$ and SS_E defined below:

$$SS_{Treatments} = 2\sum_{i=1}^{3} (\bar{y}_{i.} - \bar{y}_{..})^2,$$

$$SS_E = \sum_{i=1}^{3} \sum_{j=1}^{2} (y_{ij} - \bar{y}_{i.})^2.$$

b) Consider the null hypothesis

$$H_0 : \tau_1 = \tau_2 = \tau_3 = 0.$$

Show that the test statistic is $F_0=7$ and perform the test. Formulate a conclusion.

c) We would like to extend on the above pilot study and design a real study. How would you design this study? You are free to state your assumptions. You *can* mention words like *randomization*, *blocking* and *replication* in your answer. Please do not write more than 10 full sentences.

Thore Egeland Torfinn Torp

Appendix 1.

	Method	Gender	Class	Score
1	Method1	Boy	Class1	6
2	Method1	Boy	Class1	8
3	Method1	Boy	Class2	7
4	Method1	Boy	Class2	7
5	Method1	Girl	Class1	7
6	Method1	Girl	Class1	10
7	Method1	Girl	Class2	10
8	Method1	Girl	Class2	4
9	Method2	Boy	Class3	8
10	Method2	Boy	Class3	7
11	Method2	Boy	Class4	5
12	Method2	Boy	Class4	7
13	Method2	Girl	Class3	5
14	Method2	Girl	Class3	8
15	Method2	Girl	Class4	4
16	Method2	Girl	Class4	8
17	Method3	Boy	Class5	7
18	Method3	Boy	Class5	5
19	Method3	Boy	Class6	5
20	Method3	Boy	Class6	7
21	Method3	Girl	Class5	3
22	Method3	Girl	Class5	5
23	Method3	Girl	Class6	2
24	Method3	Girl	Class6	1

Appendix 2

	Sum Sq	Df	Mean Sq	F value	Pr(>F)
Class	44.33	5	8.87	2.17	0.1030
Residuals	73.50	18	4.08	-	-

Appendix 3

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Method	2	38.08	19.04	6.0132	0.009998	**
Gender	1	6.00	6.00	1.8947	0.185551	
Method:Gender	2	16.75	8.38	2.6447	0.098405	.
Residuals	18	57.00	3.17			

Appendix 4

```
            Df Sum Sq Mean Sq F value  Pr(>F)
Method       2 38.083 19.0417  5.0141 0.01659 *
Residuals 21 79.750   3.7976


  Tukey multiple comparisons of means
    95% family-wise confidence level

                  diff       lwr        upr     p adj
Method2-Method1 -0.875 -3.330978  1.5809782 0.6474618
Method3-Method1 -3.000 -5.455978 -0.5440218 0.0150694
Method3-Method2 -2.125 -4.580978  0.3309782 0.0979447
```

Appendix 5

```
Mean scores in the groups and total mean
              mean    n
Method1        7.4    8
Method2        6.5    8
Method3        4.4    8
Total          6.1   24
```

Figure 1