Department:	**Chemistry, biotechnology and food sciences**	
Examination in:	**STAT210**	**Experimental Design and Analysis of Variance I**
	Course code	*Course name*
Time for exam:	**Monday 2th of Sept 2013**	**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: 7 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.

The complete data and some summary statistics appear in Appendix1. Different parts of the data are used for Exercise 1 and 2. The purpose of a study was to study birth weight of newborns. Only women giving birth to a second child was included in the study. Twins and other multiple births are excluded throughout. The birth weight (recorded in the variable **Weight2**) was recorded as well as the gender of the child and information on the weight of the first child. For this study 4 hospitals were randomly chosen. The variables are summarized below:

- **Weight1:** The weight of the woman's first child coded as follows:
  - 'Low'**:** Below 3300g (<3300g)
  - 'Medium'**:** Between 3300g and 3700g
  - 'High'**:** Above 3700g (>3700g)
- **Hospital:** The hospital coded as 'H1','H2','H3' and 'H4'
- **Gender:** This variable is coded 'Girl' or 'Boy'
- **Weight2:** Birth weight of the woman's second child in gram (g)

Exercise 1

In this exercise we will only be using the variables **Weight2** and **Hospital.** You can use the output in Appendix 2 and the summary statistics in Appendix 1 for your answers.

We consider the following random effect model:

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

$i = 1, 2, 3, 4$ correspond to the a=4 hospitals 'H1', 'H2', 'H3', and 'H4',

$j = 1, 2, 3, 4, 5, 6$ correspond to the n=6 observations from each hospital,

$y_{ij} = \textbf{Weight2}$, $\tau_i \sim N(0, \sigma_\tau^2)$ and $e_{ij} \sim N(0, \sigma^2)$.

All variables τ_i and e_{ij} are independent.

a) We would like to find out if birth weight varies between hospitals.
Formulate the hypotheses, perform the test at 5% significance level, and conclude.
b) Estimate σ^2 and σ_τ^2.
c) Estimate μ. Calculate a 95% confidence interval. Interpret the answer.

Exercise 2

In this exercise we will be using the variables **Weight1, Gender** and **Weight2**. We use the following fixed effect model

$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + e_{ijk}$ where

$i = 1, 2, 3$ correspond to 'Low', 'Medium' and 'High',

$j = 1, 2$ correspond to 'Boy' and 'Girl',

$k = 1, 2, 3, 4$ correspond to replications,

$y_{ijk} = \textbf{Weight2}$ 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 a significant interaction between **Weight1** and **Gender:** Formulate the hypotheses, perform the test at 5% significance level and conclude.
b) A woman is pregnant with her second child. Her first child weighed 3800g, i.e., was classified as ‘High’. She is told at the ultra sound examination that it is a girl. Predict the weight of the child based on the above model.

c) 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 'Low','Medium', 'High',

$j = 1, 2\ldots, 8$ correspond to the eight (n=8) observations for each level of Weight1,

$y_{ij} =$ **Weight2** 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.

Find the missing numbers, indicated as (1), (2) and (3), in Appendix 4.

d) Use the output of Appendix 4 to show that there are significant differences in the weight of the second child depending on the weight of the first child. Use 5% significance level. Between which groups are there differences? Use Tukey's method and formulate a conclusion. You may use also the summary statistics at the end of Appendix 1.

e) Comment on the assumptions of the model described in c) above. You can refer to Figure 1.

Exercise 3.

a) Medical students were asked to design a pilot study involving only four birthweights to study the birth weight of the first and second child.

The medical student M1 decided to do a *two-sample design.* The design is described in Table 1 below.

Woman	BirthNo
1	1
2	1
3	2
4	2

Table 1. *Woman 1 and 2 give birth for the first time; Woman 3 and 4 for the second time. Birtweights remain to be recorded.*

The medical student M2 decided to do a *paired design*. The design is described in Table 2 below.

Woman	BirthNo
1	1
1	2
2	1
2	2

Table 2. *Two women who have each given birth to two children will be recruited and the birth weights will be obtained.*

Do you prefer the design suggested by medical student M1 or M2? Give reasons for your answer; write at most five sentences. You can introduce further assumptions if you like.

b) Consider the design suggested by the medical student M2. Let

$$D_i = \text{birthweight child 2 - birthweight child 1}, \; i = 1, \ldots, n.$$

Assume that $D_1, D_2, ..., D_n$ are independent with known standard deviation $\sigma = 100$. The difference between the weight of the second and first child is estimated by the mean, i.e., $\overline{D}$. The sample size (n) should be large enough to meet the requirement $SD(\overline{D}) \leq 10$, i.e., the standard deviation of the mean (which in this case is sometimes also called the standard error of the mean) should be at most 10.

How large must n be? Comment briefly on your answer.

Thore Egeland   Torfinn Torp

Appendix 1.

x2013

	Weight1	Gender	Hospital	Weight2
1	Low	Girl	H1	3200
2	Medium	Girl	H1	3580
3	High	Girl	H1	3750
4	Low	Boy	H1	3440
5	Medium	Boy	H1	3950
6	High	Boy	H1	3970
7	Low	Girl	H2	3440
8	Medium	Girl	H2	3790
9	High	Girl	H2	3700
10	Low	Boy	H2	3520
11	Medium	Boy	H2	3890
12	High	Boy	H2	3980
13	Low	Girl	H3	3410
14	Medium	Girl	H3	3390
15	High	Girl	H3	3870
16	Low	Boy	H3	3340
17	Medium	Boy	H3	3720
18	High	Boy	H3	3780
19	Low	Girl	H4	3120
20	Medium	Girl	H4	3330
21	High	Girl	H4	3540
22	Low	Boy	H4	3060
23	Medium	Boy	H4	3260
24	High	Boy	H4	3940

```
> summary(x2013)
   Weight1     Gender    Hospital     Weight2
 High  :8    Boy :12   H1:6      Min.   :3060
 Low   :8    Girl:12   H2:6      1st Qu.:3378
 Medium:8              H3:6      Median :3560
                       H4:6      Mean   :3582
                                 3rd Qu.:3810
                                 Max.   :3980

Mean of Weight2 depending on Weight1
          mean n
High   3816.25 8
Low    3316.25 8
Medium 3613.75 8
```

Appendix 2

```
> Anova(LinearModel.7, type="III")
Analysis of variance (unrestricted model)
Response: Weight2
              Sum Sq Df   Mean Sq F value Pr(>F)
Hospital   397812.50  3 132604.17    1.81 0.1780
Residuals 1466183.33 20  73309.17       -      -
```

Appendix 3

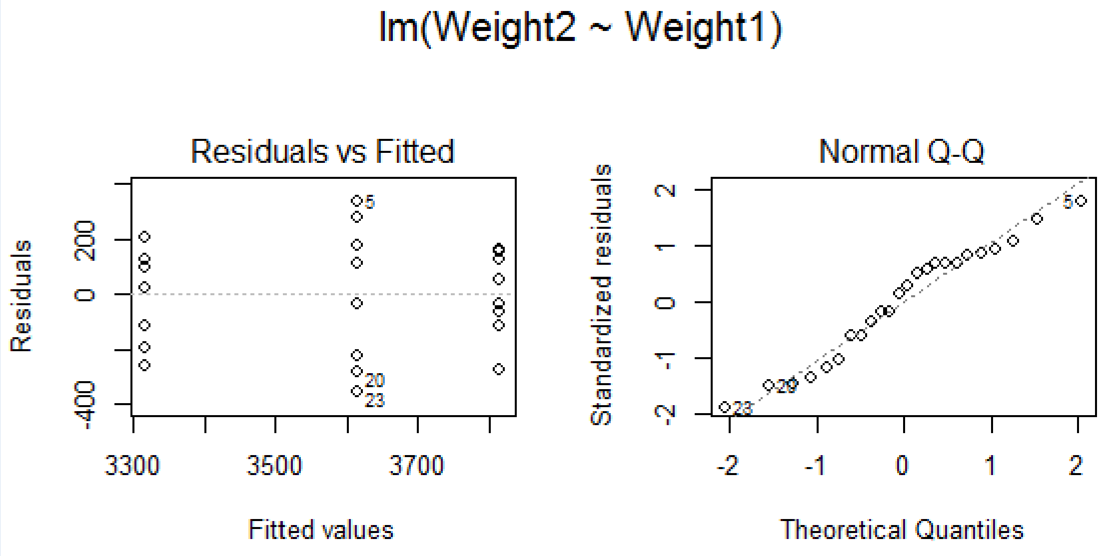
```
Coefficients:
                           Estimate
(Intercept)                 3582.08
Weight1(High)                234.17
Weight1(Low)                -265.83
Gender(Boy)                   72.08
Weight1(High):Gender(Boy)     29.17
Weight1(Low):Gender(Boy)     -48.33

               Df  Sum Sq Mean Sq F value    Pr(>F)
Weight1         2 1012033  506017 13.0337 0.0003165 ***
Gender          1  124704  124704  3.2121 0.0899221 .
Weight1:Gender  2   28433   14217  0.3662 0.6984214
Residuals      18  698825   38824
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix 4

```
          Df  Sum Sq Mean Sq F value    Pr(>F)
Weight1  (1) 1012033  506017     (2)  0.000269
Residuals 21    (3)    40570
```

Figure 1


lm(Weight2 ~ Weight1)
Residuals vs Fitted
Residuals
-400
0
200
3300
3500
3700
Fitted values
5
20
23
Normal Q-Q
Standardized residuals
-2
-1
0
1
2
-2
-1
0
1
2
Theoretical Quantiles
23
20
5