EXAMINATION QUESTIONS

Department:	**IKBM**	
Examination in:	**STAT 210**	**Design of experiments and analysis of variance I**
	Course code	*Course name*
Time for exam:	Sep 1, 2014	14:00-17:30
	Day and date	*As from – to and duration of examinations (hours)*
Course responsible:	Thore Egeland 41479582	
	Name, phone	
The exam paper includes:	6	
	Number of pages incl. attachment	

Permissible aids:

C3: all types of calculators, all other written material

Each sub-question will be given the same score in the evaluation of the exam. You may answer in English or Norwegian (or "Scandinavian"). Use significance level 5% (0.05) for tests.

Exercise 1

Different parts of the data in Table 1 will be used for different parts of this exercise. The main purpose of the experiment was to determine if the two materials, A and B, for shoe soles differed based on measurement of wear. Nine children (numbered 1 to 9) participated in the experiment. Each child used one A shoe and one B shoe for a specified time period after which wear was measured. The assignment of shoe to the left (L) or right (R) foot was done by randomization.

a) An engineer did a two sample t-test based on the variables wear and material. The computer output appears in Table 2. Formulate the null hypothesis, the alternative hypothesis and a conclusion.
b) State and comment on the assumptions for the conclusion in a) above. You may refer to Figure 1 for parts of your answer.
c) Another engineer produced the output of Table 3. Is there a significant difference

between material A and B based on this analysis? Perform the test and formulate a conclusion.

d) Explain briefly the difference between a pairwise (or blocked) design and a two sample design using the shoe example of this exercise as an example. Do you prefer the analysis leading to Table 2 or the one leading to Table 3? Give reasons for your decision briefly.

e) A third engineer was interested in the correlation of wear within children. She decided to use the following random effect model:

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

$i = 1, \ldots, 9$ correspond to the a=9 children,

$j = 1, 2$ correspond to the n=2 observations for each child,

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

All variables τ_i and e_{ij} are independent.

Output from the model is given in Table 4. Estimate the correlation $\rho = corr(y_{i1}, y_{i2})$, i.e., the correlation within a child. Comment on the answer. Calculate a 95% confidence interval for ρ.

Exercise 2

We want to investigate how the spot price of electricity for households is changing in the course of a year. Table 5 contains the average spot price for 1 kWh (referred to as price below, given in the Norwegian currency øre) for each quarter in the years 2002-2009. For most of the exercise we will use the model

$$\begin{aligned} y_{ij} &= \mu + \tau_i + \beta_j + e_{ij}, \ i = 1, 2, 3, 4 \text{ and } j = 1, 2, \ldots, 8, \\ &= \text{price for quarter } i \text{ in year } j. \\ \sum_{i=1}^{4} \tau_i &= 0, \sum_{j=1}^{8} \beta_j = 0, \\ e_{ij} &\sim NID(0, \sigma^2). \end{aligned}$$

a) Use Table 6 to test if the price differs depending on quarter. Formulate the hypotheses and a conclusion.

b) Estimate the standard deviation. Calculate a 95% confidence interval for this standard deviation. Use this interval to test whether the standard deviation differs significantly from 9 øre.

c) Use the output of Table 7 to determine between which quarters there are significant differences in price. Calculate a 95% Tukey confidence interval for the expected difference between quarters 4 and 2 (or equivalently find the missing numbers indicated by (1) and (2) in Table 7).

d) Can we argue that the expected price differs significantly between the months January-June (quarter 1 and 2) and July-December (quarter 3 and 4)? Set up an appropriate contrast, state the hypotheses and formulate a conclusion based on Table 8.

e) A student claimed that it was not necessary to have the effect of years in the model because the years should not be compared. He suggested to run a one-way analysis of variance where the quarter was the only fixed effect of the model. Use the information in Table 6 to create the ANOVA table (without p-value) for this model. Can you claim there is a significant effect of quarter from this analysis? Do you agree with the claim that "it was not necessary to have the effect of years in the model because the years should not be compared"? Give reasons for your answer.

Course responsible: Thore Egeland

External examiner: Torfinn Torp

Table 1

Ex1

	child	wear	side	material
1	1	13.2	L	A
2	1	14.0	R	B
3	2	8.2	L	A
4	2	8.8	R	B
5	3	10.9	R	A
6	3	11.2	L	B
7	4	14.3	L	A
8	4	14.2	R	B
9	5	10.7	R	A
10	5	11.8	L	B
11	6	6.6	L	A
12	6	6.4	R	B
13	7	9.5	L	A
14	7	9.8	R	B
15	8	10.8	L	A
16	8	11.3	R	B
17	9	8.8	R	A
18	9	9.3	L	B

Table 2

```
        Two Sample t-test

data:  wear by material
t = -0.3657, df = 16, p-value = 0.7194
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.869596  2.025152
sample estimates:
       mean of x        mean of y  pooled std.dev.
       10.333333        10.755556          2.449008
```

Table 3

```
          Df  Sum Sq  Mean Sq  F value
material   1   0.802   0.8022   9.6106
child      8  95.294  11.9118 142.7038
Residuals  8   0.668   0.0835
---
```

Table 4

```
> Anova(LinearModel.6, type="III")
Analysis of variance (unrestricted model)
Response: wear
          Df  Sum Sq  Mean Sq  F value Pr(>F)
child      8   95.29    11.91    72.93 0.0000
Residuals  9    1.47     0.16        -      -

            Err.term(s)  Err.df  VC(SS)
1 child             (2)       9   5.874
2 Residuals           -       -   0.163
(VC = variance component)
```

Table 5

Ex2

	price	year	quarter
1	32	2002	1
2	28	2002	2
3	30	2002	3
4	46	2002	4
5	57	2003	1
6	45	2003	2
7	50	2003	3
8	51	2003	4
9	32	2004	1
10	28	2004	2
11	30	2004	3
12	32	2004	4
13	27	2005	1
14	33	2005	2
15	33	2005	3
16	30	2005	4
17	44	2006	1
18	40	2006	2
19	60	2006	3
20	56	2006	4
21	32	2007	1
22	25	2007	2
23	11	2007	3
24	45	2007	4
25	43	2008	1
26	21	2008	2
27	49	2008	3
28	58	2008	4
29	46	2009	1
30	38	2009	2
31	38	2009	3
32	40	2009	4

Table 6

```
             Df  Sum Sq Mean Sq F value   Pr(>F)
year          7 2214.87  316.41  4.9703 0.001932
quarter       3  634.12  211.37  3.3203 0.039547
Residuals 21 1336.88   63.66
```

Table 7

```
Fit: lm(formula = price ~ year + quarter, data = Ex2.tmp)

Quantile = 2.7885
95% family-wise confidence level

Linear Hypotheses:
           Estimate     lwr     upr Std. Error t value Pr(>|t|)
2 - 1 == 0   -6.875 -18.000   4.250      3.989  -1.723   0.3371
3 - 1 == 0   -1.500 -12.625   9.625      3.989  -0.376   0.9814
4 - 1 == 0    5.625  -5.500  16.750      3.989   1.410   0.5072
3 - 2 == 0    5.375  -5.750  16.500      3.989   1.347   0.5445
4 - 2 == 0   12.500     (1)     (2) |    3.989   3.133   0.0238 *
4 - 3 == 0    7.125  -4.000  18.250      3.989   1.786   0.3077
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

Table 8

```
                                Estimate Std. Error   t value   Pr(>|t|) DF
quarter c=( 0.5 0.5 -0.5 -0.5 )     -5.5    2.82092 -1.949719 0.06468833 21
                                 lower CI  upper CI
quarter c=( 0.5 0.5 -0.5 -0.5 ) -11.36642 0.3664242
```

Figure 1