EXAMINATION QUESTIONS

Department: **IKBM**

Examination in: **STAT210** (*Course code*) **Design of experiments and analysis of variance 1** (*Course name*)

Time for exams: Monday 31. Aug 2015 (*Day and date*) 14:00-17:30 (3.5 hours) (*As from – to and duration of examinations (hours)*)

Course responsible: Guro Dørum, 95863762 (*Name*)

Permissible aids:

C3: all types of calculators, all other aids

The exams papers includes: 7 (*Number of pages incl. attachment*)

All 10 sub questions are given equal weight in the evaluation. You may answer in English or Norwegian (or "Scandinavian").

Course responsible: Guro Dørum

External examiner: Torfinn Torp

Exercise 1

An experiment was performed to compare five bread recipes. It was decided to bake the breads in 4 different ovens, where each oven could fit 5 breads. 20 persons were randomly chosen to taste one bread and give points on a scale from 1 to 10, where 10 is best. The data are given in Table 1.

We will use the following model:

$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$ where

$i = 1,2,3,4,5$ correspond to bread recipe,

$j = 1,2,3,4$ correspond to oven,

y_{ij} is points for bread baked with recipe i in oven j

and $\varepsilon_{ij} \sim NID(0,\sigma^2)$,

$\sum_{i=1}^{5} \tau_i = 0, \quad \sum_{j=1}^{4} \beta_j = 0$

a) Explain briefly why it is natural to consider oven as a blocking factor.

The ANOVA table is given in Table 2.

b) Use the output in Table 2 to test if there is an effect of recipe. State the null hypothesis and the alternative hypothesis and give a conclusion. Use 5 % significance level.

c) Do you think oven should be included as block in the model? Give your answer based on the output in Table 2.

Independent of your answer in c), we will use the following model for the remainder of the exercise:

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

$i = 1,2,3,4,5$ correspond to bread recipe,

j=1,2,3,4 correspond to bread number,

y_{ij} is points for bread j baked with recipe i

and $\varepsilon_{ij} \sim NID(0,\sigma^2)$,

$\sum_{i=1}^{5} \tau_i = 0$

d) Find SS_{Recipe} and SS_E, with their corresponding degrees of freedom, for this new model, based on the output from the first model (Table 2). Test if there is effect of recipe now. Use 5 % significance level.

e) Use Tukey's method for pairwise comparisons to determine which recipes that differ. Use a 5 % significance level. Use the output of Table 3.

Exercise 2

An experiment was performed to investigate variation in vitamin C content in mango that grows on 4 different locations in Tanzania. One was also interested in the difference between 2 different types of mango, green and yellow. 3 green and 3 yellow mango fruits were randomly picked at each location. Finally the vitamin C content was measured for each fruit (in milligram per 100 g). We will consider the following model:

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

$i = 1,2,3,4$ correspond to location,

$j = 1,2$ correspond to type ('green', 'yellow'),

$k = 1,2,3$ correspond to fruit (replicate),

y_{ijk} is vitamin C measured for fruit k of type j on location i,

$\varepsilon_{ijk} \sim NID(0,\sigma^2)$,

$$\sum_{i=1}^{4} \tau_i = 0,\ \sum_{j=1}^{2} \beta_j = 0,\ \sum_{i=1}^{4} (\tau\beta)_{ij} = 0,\ \sum_{j=1}^{2} (\tau\beta)_{ij} = 0$$

a) Explain to a non-statistician what interaction means in this experiment. Use maximum two sentences. Use the output from the ANOVA given in Table 4 to test for effect of interaction between location and tree type. Use 5 % significance level.

We will now consider the reduced model:

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

$i = 1,2,3,4$ correspond to location,

$j = 1,2,\ldots,6$ correspond to fruit,

y_{ij} is vitamin C measured for fruit j on location i,

$\varepsilon_{ij} \sim NID(0,\sigma^2)$,

$$\sum_{i=1}^{4} \tau_i = 0,$$

b) Location 1 and 2 are in the highlands of Tanzania, and location 3 and 4 are in the lowlands. We want to test if there is a significant difference between the vitamin C content in mango grown in the highlands and in the lowlands. Formulate a suitable contrast, state the null hypothesis and the alternative hypothesis, and perform the test. Use 5 % significance level. Use the output in Table 5. Note that 'loc1' corresponds to location 1, 'loc2' to location 2 etc.

Another experiment was performed to estimate the mean vitamin C content in mango for all of Tanzania. 10 new locations were randomly chosen, from each location 5 mango fruits were randomly picked, and the vitamin C content was measured. The model used was:

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

$i = 1,2,...,10$ correspond to location,

$j = 1,2,...,5$ correspond to fruit,

y_{ij} is the vitamin C content measured in fruit j on location i,

and $\tau_i \sim NID(0,\sigma_\tau^2)$ and $\varepsilon_{ij} \sim NID(0,\sigma^2)$.

All τ_i and ε_{ij} are assumed to be independent.

Table 6 gives some summary data and output from the ANOVA.

c) Use the output from Table 6 to estimate the variance components, and estimate the correlation between two trees in the same location.

d) Construct a 95 % confidence interval for the expected vitamin C content in mangos in Tanzania. Use this confidence interval to test if the expected vitamin C content differs significantly from 26 milligram per 100 g. Use the output from Table 6.

e) Construct a 95 % confidence interval for the variance within location.
A student estimates this variance as SS_E/N, where N is the total number of observations in the experiment. Explain briefly why this student is making a mistake.

Table 1

	Recipe1	Recipe2	Recipe3	Recipe4	Recipe5
oven1	6	7	5	4	8
oven2	5	7	4	5	7
oven3	4	8	7	6	6
oven4	5	6	6	4	7

Table 2

```
Response: Points
           Sum Sq Df F value  Pr(>F)
Recipe     18.80  4  4.5484 0.01819
Oven        1.35  3  0.4355 0.73160
Residuals  12.40 12
```

Table 3

```
Multiple Comparisons of Means: Tukey Contrasts


Fit: aov(formula = mod)

Quantile = 3.0869
95% family-wise confidence level

Linear Hypotheses:
                       Estimate
Recipe2 - Recipe1 == 0  2.00000
Recipe3 - Recipe1 == 0  0.50000
Recipe4 - Recipe1 == 0 -0.25000
Recipe5 - Recipe1 == 0  2.00000
Recipe3 - Recipe2 == 0 -1.50000
Recipe4 - Recipe2 == 0 -2.25000
Recipe5 - Recipe2 == 0  0.00000
Recipe4 - Recipe3 == 0 -0.75000
Recipe5 - Recipe3 == 0  1.50000
Recipe5 - Recipe4 == 0  2.25000
```

Table 4

```
Response: VitaminC
               Sum Sq Df F value   Pr(>F)
Location      24.5503  3  5.8169 0.006929
Type           1.7920  1  1.2738 0.275703
Location:Type  3.4235  3  0.8111 0.506153
Residuals     22.5096 16
```

Table 5

```
        mean         sd data:n
loc1 27.8890 0.4773418      6
loc2 29.6500 1.5478249      6
loc3 27.5525 1.6500476      6
loc4 26.9245 0.4457841      6

Response: VitaminC
          Sum Sq Df F value   Pr(>F)
Location  24.550  3  5.9033 0.004685
Residuals 27.725 20
```

Table 6

```
   VitaminC
Min.   :19.82
1st Qu.:24.83
Median :27.39
Mean   :27.48
3rd Qu.:30.44
Max.   :34.33
```

```
Response: VitaminC
          Mean Sq Sum Sq Df F value Pr(>F)
Location    26.46 238.18  9    2.97 0.0084
Residuals    8.91 356.58 40       -      -

            Err.term(s) Err.df VC(SS)
1 Location          (2)     40   3.51
2 Residuals           -      -   8.91
(VC = variance component)

          Expected mean squares
Location              (2) + 5 (1)
Residuals             (2)
```