EXAM

Department: IKBM

Examination in: STAT210 Design of experiments and analysis of variance

Time for exams: Monday 5. September 2016 14:00-17:30 (3.5 hours)

Course responsible: Trygve Almøy 95 14 13 44

Permissible aids: **C3: all types of calculators, all other aids**

The exams papers includes: 5

Number of pages incl. attachment

Question 1 counts for 40 %, Question 2 counts for 20 %, and Question 3 counts for 40 %.
The sub-questions have equal weights.
You may answer in English or Norwegian (or "Scandinavian").
If you are testing, you are supposed to state the null hypothesis, the alternative hypothesis and (if not given in the text) the level of significance.

Question 1

We have five sorts of potatoes (treatments) A, B, C, D, and E. Of these sorts A, C, and D are Norwegian, while B and E are foreign sorts. The experiment was done as a randomized complete block design. Three blocks were used. The response variable was kilogram dry matter (Norwegian: Tørrstoff) per experimental unit. The experimental plan and the observed results are given in Table 1.

Sort (as treatment)	*Block*			*Mean*
	j = 1	*j = 2*	*j = 3*	
i = 1 (A)	6.40	6.43	6.40	$\bar{Y}_{1\cdot}$ = **6.410**
i = 2 (B)	5.66	5.40	4.62	$\bar{Y}_{2\cdot}$ = **5.227**
i = 3 (C)	6.84	6.52	6.19	$\bar{Y}_{3\cdot}$ = **6.517**
i = 4 (D)	6.67	6.72	6.15	$\bar{Y}_{4\cdot}$ = **6.513**
i = 5 (E)	5.12	5.58	5.00	$\bar{Y}_{5\cdot}$ = **5.233**
Mean	$\bar{Y}_{\cdot 1}$ = **6.138**	$\bar{Y}_{\cdot 2}$ = **6.130**	$\bar{Y}_{\cdot 3}$ = **5.672**	$\bar{Y}_{\cdot\cdot}$ = **5.980**

Table 1: Data from Question 1

We used the following model.

Model 1: $Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$, i = 1, 2, …, 5; j = 1, 2, 3.

Here τ_i is the effect of Sort i, and β_j is the effect of Location (Block) j.
All ε_{ij}'s are assumed to be independent and normally distributed with expectation 0 and variance σ^2. We assume that both the treatment and the block are fixed effects.

Restrictions: $\tau_1 + \tau_2 + \tau_3 + \tau_4 + \tau_5 = 0$, and $\beta_1 + \beta_2 + \beta_3 = 0$.

In Table 2 you can find the ANOVA table, however *Degrees of freedom* (Df) are removed.

```
Response: Yield

          Sum Sq Df  F value    Pr(>F)
Sort      5.647  ?   23.887  0.0001671
Location  0.712  ?
Residuals 0.473  ?
```

Table 2. Parts of ANOVA table for block design Question 1.

a) Give the degrees of freedom for Sort, Location and Residuals.

b) Find the estimates of μ, τ_1 and β_1.

c) Interpret and give an estimate of σ^2.

d) Discuss possible advantages obtained by blocking.
Do you agree that location (the block) should be included in the model? Argue for your answer.

e) Test if there is effect of sort, use level of significance 5 %.

f) Show that a consequence of the model is:
The expected difference between two sorts in one location is the same for all locations.

A student had never heard about block design, so he fitted (incorrectly) Model 2 for the same data.

Model 2: $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$. (Similar assumptions as in Model 1).

Use Model 2 for the sub questions **g**), **h**) and **i**).

g) Test if there is significant difference between the sorts. (Hint: Use the output in Table 2)

Let $\Gamma = \frac{\tau_1 + \tau_3 + \tau_4}{3} - \frac{\tau_2 + \tau_5}{2}$

h) Show that Γ is a contrast, which represents the *expected* difference between the Norwegian and the foreign sorts.

i) Estimate the contrast and test if the Norwegian sorts gives more dry matter than the two foreign sorts.

Question 2
One investor wanted to produce blue mussel (Norwegian: Blåskjell) along the coast of Norway. The response called yield (measured as edible food in percent of the total mass) could possibly vary over the locations. This was investigated by a random selection of 3 locations along the Norwegian coast, and 5 mussels sampled in each location.
A *random effect model* was analyzed with the results in Table 3.

Location	mean
A	20.4
B	20.2
C	21.8

	Mean Sq	Sum Sq	Df	F value	Pr(>F)
Location	3.80	7.60	2	1.07	0.3750
Residuals	3.57	42.80	12		

Table 3: For use in Question 2.

a) Give the model and the assumptions.
Test if Location has significant effect.
Estimate the variance components (usually called σ_τ^2 and σ^2).

b) Construct a 95 % confidence interval for the average yield along the Norwegian coast (usually named μ).

Question 3
It is well known that pH in eggs are increasing during storage. However, the pH level could also be dependent on washing regime. From a master thesis, we borrow some data on 32 randomly selected eggs.

The washing water had two different temperatures (Warm and Cold, named W and C), and to levels of the time the eggs spend in the washing water (Short and Long, named S and L).
Tables and Figures are at the end of this question

a) Construct an interaction effect plot (plot of means). Explain how the plot can support the statement that interaction should be included. (The means are presented in Table 4).

b) State the model where interaction between the two factors with respect to pH is included. Also, add necessary restrictions and assumptions. Test if the interaction effect is significant.

c) Find the fitted value and the residual for the first observation in Table 7.
What does it mean if the residual has a positive value?
If interaction was excluded from the model, what would be the fitted value and the residual for the first observation?

d) Test if long washing time gives significantly higher pH compared to short washing time if cold water is used for both washing times.

e) The residual plots in Figure 1 is from a model without interaction. Explain how this might reveal that interaction should be included.

	L	S
C	9.04	9.02
W	9.22	8.98

Table 4: The sample means for each combinations Question 3.

Response: Ph

	Sum Sq	Df	F value	Pr(>F)
Temperature	0.035113	1	3.1325	0.08763
Time	0.137813	1	12.2949	0.00155
Temperature:Time	0.094612	1	8.4408	0.00709
Residuals	0.313850	28		

Table 5: ANOVA table for Question 3.

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.065	0.0187	484.385	< 2e-16
Temperature(C)	-0.033	0.0187	-1.770	0.08763
Time(L)	0.066	0.0187	3.506	0.00155
Temperature(C):Time(L)	-0.054	0.0187	-2.905	0.00709

Table 6: Information for Question 3.

Ph	Time	Temperature
9.28	S	C
8.82	S	C
8.85	S	C
8.85	S	C
9.05	L	C
9.05	L	C
9.11	L	C
9.05	L	C
9.02	S	W
9.01	S	W
8.93	S	W
8.91	S	W
9.03	L	W
9.30	L	W
9.30	L	W
9.22	L	W
9.03	S	C
9.17	S	C
9.06	S	C
9.11	S	C
8.92	L	C
9.04	L	C
9.12	L	C
9.01	L	C
8.89	S	W
9.06	S	W
9.00	S	W
9.01	S	W
9.14	L	W
9.30	L	W
9.20	L	W
9.26	L	W

Table 7: Data for Question 3

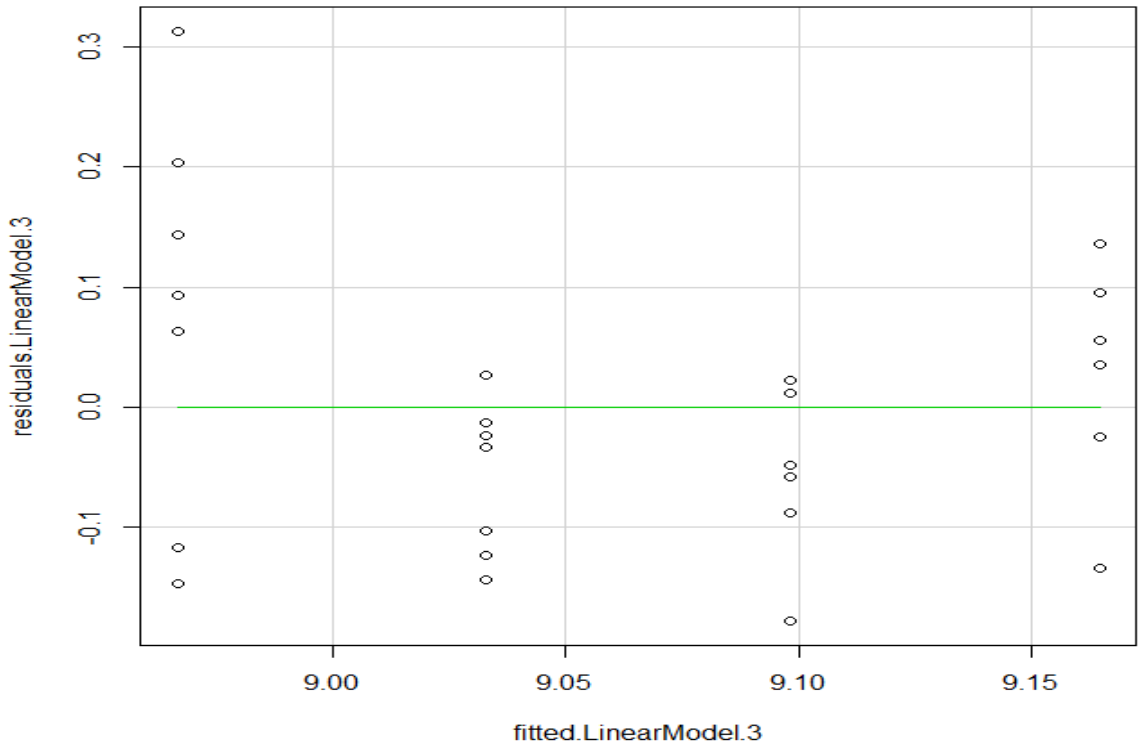

Figure 1: Residuals against fitted values when interaction was excluded.