EXAMINATION: Experimental design (MSA250/TMS031)
Friday, June 9, 2017, 8:30 - 12:30
Lecturers on call: Kerstin Wiklander and Torbjörn Lundh, tel 772 5355 and tel 772 3503.
Tools: A pocket calculator with emtied memory. At the examination, sheets with statistical distributions and tables will be handed out.

Give explanations to the notation you use and motivation to your conclusions.

1. (2 p) Describe what the term "pure error" means and how one compute the pure error.

   Short solution indication:
   S_E the "pure" error, i.e. the sum of the square of errors from genuine replicates. See p. 369 for more details.

2. (2 p)

   (a) Show the formula for the estimated standard deviation of an calculated effect l_i from factor i in a 2^k-design with n replicates (constant for each level combination).

   (b) Explain short what a Latin square design is.

   Short solution:
   (a) $s_{effect} = \frac{2s_p}{\sqrt{N}}$, where s_p^2 is the pooled sample variance and $N = 2^k n$.
   (b) See section 4.4 in the book.

3. (5 p) Given a randomized experiment of two kinds of yellow tulips, A and B where the respond is flower hight in cm.

A	A	B	A	B	B
48	43	52	50	49	52

   Construct a randomization test and give the p-value. Compare this with the result from a t-test at 5% significance level for the two cases:

   (a) if the new tulip strain B is better, i.e. give rise to more rapid growing tulips

   (b) if one of the tulip strain is better than the other.

   For the last task, using the t-distribution, compute the 95% confidence interval of the expected tulip-hight-difference between B and A.

   Short solution:

$$\binom{6}{3} = 20$$

There is one more extreeme configuration and one different that give an equal difference. This gives the one-sided p-value 0.1 for the randomization and 0.08 for the t-test. For the second task, we will instead have 0.2 and 0.16. The confidence interval is $(-2.4, 10.4)$

4. (5 p) In an investigation on the speed of the ball in table tennis, a 2^3 factorial desing was conducted. The factors were type of rubber on the paddle (racket) (factor R), type of glue (factor G) and type of ball (factor B). A professional player was selected for the experiment where the speed (km/h) of the ball was measured under equal conditions (besides the factor settings). Everything was randomized and the mean from $n = 4$ measurements are given in the table below. The calculated standard deviation for an effect was $s_{effect} = 0.25$. The factor settings and the data:

Factor	Levels coded by $-$	coded by $+$
R	Donic	Palio
G	Butterfly	Stiga
B	Nittaku	Yasaka

R	G	B	Mean speed
$-$	$-$	$-$	63.3
$+$	$-$	$-$	64.9
$-$	$+$	$-$	63.8
$+$	$+$	$-$	64.7
$-$	$-$	$+$	64.1
$+$	$-$	$+$	66.5
$-$	$+$	$+$	64.6
$+$	$+$	$+$	66.8

Estimate and test all effects.

Short solution:
$l_R = 1.775$ $l_G = 0.275$ $l_B = 1.325$ $l_{RG} = -0.225$ $l_{RB} = 0.525$ $l_{GB} = 0.125$ and $l_{RGB} = 0.125$.
Divide all by $s_{effect} = 0.25$, which give the test statistics (for example the value of 7.1 for factor R). The absolute values that are $\geq t_{24,0.025} = 2.064$ give a significant result for that effect on significande level 5%. Those are: R, B and the interaction RB.

5. (6 p) Measurements of the height of a plant species was taken in three different areas. The sample sizes were 11 in all three cases. The aim of the study was to find some differences between the areas. Part of the result is given in the Anova-table below.

Source	SS	df	MS	F
Area		2		
Residual	2.94			
Total	4.15	32		

   (a) Fill in the missing parts in the Anova table. State also the assumptions for using Anova as a test method and the two hypotheses.

   (b) Is there evidence of any difference between the areas? Motivate your conclusion.

   (c) Give the model matrix X but with the sample size two only.

Short solution:

(a)

Source	*SS*	*df*	*MS*	*F*
Area	1.21	2	0.605	6.17
Residual	2,94	30	0.098	
Total	4,15	32		

All variables independent and normally distributed with constant theoretical variance. $H_0 : \mu_1 = \mu_2 = \mu_3$ and $H_1 : \mu_i \neq \mu_j$ for some $i \neq j$.

(b) Since $F = 6.17 > F_{2,30,0.05} = 3.32$, the null hypotesis of equal expeced values can be rejected on significance level 5%. (The same conclusion also for significance level 1% since $F_{2,30,0.01} = 5.39$.)

(c) The design matrix with the model formulation $y_{ij} = \mu_i + \epsilon_{ij}$ or expressed by $y_{ij} = \mu + \tau_i + \epsilon_{ij}$ (with the sum of the τ's equal to zero) :

$$X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \text{ or } X = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

6. (2 p) All experiments could not be done during the same day in a two-level factorial desing with three factors. It was assumed that there could be different conditions the two days that it would take and it was decided to use a block design. Show how to perform this by adding information in the factorial design below and explain the consequences in such a design.

A	*B*	*C*
−	−	−
+	−	−
−	+	−
+	+	−
−	−	+
+	−	+
−	+	+
+	+	+

Short solution:

Use a block design and identify the block effect by the ABC-effect. They will then be alias (both in design and consequenses). See more on pages 211-212.

7. (4 p) The least squares estimators of the parameters in a linear model $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$ is expressed as $\mathbf{b} = (X^T X)^{-1} X^T \mathbf{y}$ using matrix notation.

   (a) Show this formula.

(b) To use the formula you need to express X according to your model and design set-up. Give an example of a design matrix X in a regression model with two independent explanatory variables ("regressors") and their interaction.

Short solution:
See pages 416 and 417 in *Statistics for Experimenters*, 2nd edition.

8. (4 p) Desribe a general second-order experimental design for three factors as

   (a) a formula $\hat{y} = \ldots.$

   (b) a sequence of graphical descriptions with dots and edges.

Short solution:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3.$$

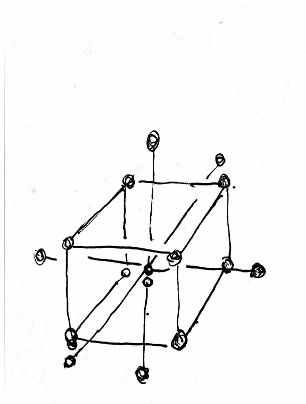

Figure 1: 8. (b)