TENTAMEN: Experimental design (TMS031/MSA250)
Torsdagen den 12 april, 2012, kl 8:30-12:30
Lärare och jour: Aila Särkkä, tel. 772 35 42
Hjälpmedel: A valfri miniräknare (med tömt minne) och tabellbok (minicalculator (with emptied memory) and a book of tables).

1) a) What is the main idea in the least squares estimation? What assumptions do you have to make?

   b) Let us have a 2^2 factorial design and fit a first order (plane) model to the data. After having fitted the model, you look at the curvature or the interaction term. Why? What additional information would you need (in addition to the original experimental points)? What assumptions do you have to make in order to use the information from the curvature/interaction term? (8p)

2) A chemist performed the following 2^3 factorial design in order to study a chemical reaction, and obtained the following results:

Run	Temperature ($^\circ C$)	Concentration (%)	Stirring rate (rpm)	Yield (%)
1	50	6	60	54
2	60	6	60	57
3	50	10	60	69
4	60	10	60	70
5	50	6	100	55
6	60	6	100	54
7	50	10	100	80
8	60	10	100	81

The standard deviation of each observation is approximately 2.

   a) Analyze these data (estimate the main and interaction effects, explain the assumptions you make).

   b) What are the main conclusions you can draw from the data?

   c) If the chemist would like to perform two further runs and her object is to obtain high yield values, what settings of temperature, concentration and stirring rate would you recommend? (10p)

3) Have you ever thought, when waiting to get someone's parking space, that the driver you are waiting for is taking longer than necessary? In order to examine this question, we hung out in parking lots and recorded the time that it took for a car to leave a parking place. The data were divided into two groups: those where there was someone in another car waiting for the space and those where there was no one waiting. The data (waiting time in seconds) are as follows:

No one waiting (X_1): 36.30 42.07 39.97 39.33 33.76 33.91 39.65 84.92 40.70 39.65 39.48 35.38 75.07 36.46 38.73 33.88 34.39 60.52 53.63 50.62

Someone waiting (X_2): 49.48 43.30 85.97 46.92 49.18 79.30 47.35 46.52 59.68 42.89 49.29 68.69 41.61 46.81 43.75 46.55 42.33 71.48 78.95 42.06

a) Suggest a randomization test to test whether the two means (waiting time when no one is waiting and waiting time when someone is waiting) differ. Give the hypotheses and explain how you do the randomization. How many combinations are there in total?

b) Assume that we have made the randomization 4999 times, and computed the test statistic $\bar{X}_1 - \bar{X}_2$ for each randomization. For the data above the test statistic has the value -9.68 and among the randomized test statistics there are 17 that are smaller than or equal to -9.68, and 13 are larger or equal to 9.68. Compute the p value.

c) Perform a two sample T test to test the same hypotheses as in b). (The sample sizes are $n_1 = n_2 = 20$, mean values $\bar{x}_1 = 44.42$ and $\bar{x}_2 = 54.11$, and the sample standard deviations $s_1 = 14.10$ and $s_2 = 14.39$.)

d) What are the conclusions in b) and c)? Discuss the similarities/differences of the results and the assumptions needed, and give your recommendation on which test to use and why. (12p)

4) We are investigating the effect of diet on the coagulation time of blood of rabbits, which were housed in cages placed on four shelves (four different heights). The main interest is to find out whether the diet has an effect on the coagulation time. If we also wanted to acknowledge the potential effect of height of the cage on the coagulation time, we should organize the experiment using a randomized complete block design. The coagulation times for the 16 rabbits are as follows:

	Diet A	Diet B	Diet C	Diet D
Shelf 1	62	63	68	56
Shelf 2	60	67	66	62
Shelf 3	63	71	71	60
Shelf 4	59	64	67	61

a) Give the model that includes the blocking variable. Mention also the assumptions you make.

b) How would you do the randomization to allocate a diet and a shelf for each rabbit? Note that each shelf has four positions.

c) Fill in the ?'s in the incomplete ANOVA table:

Source	df	SS	MS	F ratio
Diet	?	191.5	?	?
Shelf	?	38.0	?	?
Residuals	?	45.5	?	
Total	?	275.0		

d) What are your conclusions based on the ANOVA table and F test.

e) What design would you have if you ignored the possible effect of the height of the shelf? What design would you use if you took into account both the height of the cage (shelf) and the position on the shelf? (12p)

5) We have made the following experiment:

Run number	**a** **A**	**b** **B**	**c** **C**	**ab** **D**	**ac** **E**	**bc** **F**	**abc** **G**	Response y
1	-	-	-	+	+	+	-	68.4
2	+	-	-	-	-	+	+	77.7
3	-	+	-	-	+	-	+	66.4
4	+	+	-	+	-	-	-	81.0
5	-	-	+	+	-	-	+	78.6
6	+	-	+	-	+	-	-	41.2
7	-	+	+	-	-	+	-	68.7
8	+	+	+	+	+	+	+	38.7

a) Which design is this? Give the resolution and the defining relation.

b) What assumptions have you made in order to get the confounding pattern below

$$l_A = -10.9 \rightarrow A + BD + CE + FG$$
$$l_B = -2.8 \rightarrow B + AD + CF + EG$$
$$l_C = -16.6 \rightarrow C + AE + BF + DG$$
$$l_D = 3.2 \rightarrow D + AB + CG + EF$$
$$l_E = -22.8 \rightarrow E + AC + BG + DF$$
$$l_G = 0.5 \rightarrow G + AF + BE + CD$$

c) What can you say based on the results above (in b)? Which effects are active?

d) What would be a natural next step if you were allowed additional eight runs? (8p)

Good luck!