Informatics and Mathematical Modelling
Technical University of Denmark

Exercises in the Design and Analysis of Experiments

Henrik Spliid

Spring semester 2006 version

IMM, DTU

Foreword

The exercises in the present booklet are intended for use in the courses given by the author about the design and analysis of experiments. Please respect that the material is copyright protected.

Corresponding to most of the exercises in this collection solutions have been worked out.

The idea is that the student is encouraged to try to do (most of the work in) these exercises and subsequently consult the solutions.

Do not miss the opportunity to learn to apply the different methods by looking into the solutions to early, but wait until you have made a good effort!

In some of the exercises topics that are not covered in the course may appear such as, for example, multi-factor experiments with factors at three levels and their construction (based on Kempthorne's method). I apologize for that, but hope that the reader may still benefit from seeing the concepts and - perhaps - look into the solutions and see some of these interesting methods used.

The numbering of the different exercises is a little irregular, in that it starts with number 10 and then increases with irregular intervals. This expresses that many previous exercises have been deleted as better ones have come up.

During the semester there may also pop up new ones, which then will be placed after the present ones.

A few are still in Danish, if needed they will be translated during the course. later.

Good luck,

Henrik Spliid

IMM, DTU, February 2006

Exercises 1 through 8 are elementary and only meant for repetition:

Exercise 1

From a production line 4 roller bearings were selected randomly and their diameters were measured. The results in cm were as follows:

$$1.0250 \qquad 1.0252 \qquad 1.0249 \qquad 1.0249$$

Compute the sample standard deviation s. Compute the sample standard deviation of the mean $\overline{X}$.

Exercise 2

A civil engineer tested two different types of balks, A and B, of armored concrete. He tested nine balks of which five were of type A and four of type B. Based on the data below he wanted to determine (if possible) whether there is a difference between the two types or not. Which kind of model and assumptions are appropriate for this problem. Which conclusions can he draw based on the data at hand (explain shortly your result).

Strength (coded numbers)

A	B
67	45
80	71
106	87
83	53
89	

Exercise 3

A test panel of 15 individuals participating in testing two brands of beer (A and B) were asked to give their judgments on a scale from 1 to 10. 8 individuals were given brand A and 7 were given B. The distribution was not known to the test panel.

The results were:

Brand A	2	4	2	1	9	9	2	2
Brand B	8	3	5	3	7	7	4	

Test the null hypothesis $\mu_A = \mu_B$ against the alternative $\mu_A \neq \mu_B$, an remember to give the necessary assumptions underlying your test method. Could you suggest a better design for this experiment? Write out precisely how you would then conduct your improved experiment.

Exercise 4

The following data are results from a comparison of two different methods for determining the amount of dissolved oxygen in water (mg per liter). 6 samples were analyzed by both methods as shown in the following table:

Sample no.	1	2	3	4	5	6
Method A (amperometric type)	2.62	2.65	2.79	2.83	2.91	3.57
Method B (visual type)	2.73	2.80	2.87	2.95	2.99	3.67

Estimate the difference between the two methods. Compute a confidence interval for the difference between the methods (degree of confidence 95% for example). Which assumptions did you make in this analysis?

Do you think this experiment is sufficient basis for choosing one of the methods (and not the other). Suggest possible further experiments and/or analyses and discuss which results might be obtained.

Exercise 5

The following data concern two alternative methods for inhaling an asthma spray. Method A is a manually based method and B uses an automatic inhalator. The measurements are breath resistance measured 30 minutes after use of the spray. The data presented to you are shown below, they consist of 10 values for each of the two methods. Obviously, you cannot analyze these data without knowing how they were collected and under which circumstances.

Specify which questions you would ask, before analyzing the data and describe (shortly) how you would conduct the analysis depending on the experimental setup.

A	B	A	B
17.00	11.60	22.80	11.60
21.60	13.65	20.40	17.22
11.20	8.25	14.00	6.20
52.25	41.50	7.50	6.96
12.20	8.40	18.85	9.00
6.05	5.18	4.05	3.00

Exercise 6

Five groups of students in a class measured the distance between two points in the landscape. The results were (meters):

420.6 421.0 421.0 420.7 420.8

Find a 95%-confidence interval for the distance and state your assumptions made in doing this.

Exercise 7

Two kinds of trees, A and B, were planted on 20 pieces of land (plots). A-trees were planted on ten of the plots (randomly selected among the twenty) and B-trees were planted on the reaming ten plots. Six years after being planted the average height of the trees was measured for each plot, and the results were as follows:

A trees	3.2	2.7	3.0	2.7	1.7	3.3	2.7	2.6	2.9	3.3
B trees	2.8	2.7	2.0	3.0	2.1	4.0	1.5	2.2	2.7	2.5

Find a 95% confidence interval for the difference in mean height, and state your assumptions made for doing this.

Exercise 8

Five recipes were used to bake a number of cakes from two types of basic cake mix, A and B. The difference between the two types was that type A mix was added a carbon dioxide source while B was not. The response of the experiment was the volume of the baked cakes. The results were as follows:

Recipe	A	B
1	83	65
2	90	82
3	96	90
4	83	65
5	90	82

The five recipes were somewhat different with respect to amount of water used, mixing time, baking temperature and baking time.The producer of A claims that his product results in a significantly larger volume when compared with B. Do the data support this claim?

Give comments on the experiment and on the analysis of the data that you find relevant. Also discuss the model and the appropriate assumptions on which your analysis is based.

Finally, compute a 95% confidence interval for the mean difference of the volume from the two types of mix and comment on the interval in relation to the above claim.

Regular exercises in the design and analysis of experiments:

Exercise 10

A cornflakes company wishes to test the market for a new product that is intended to be eaten for breakfast.

Primarily two factors are of interest, namely an advertizing campaign and the type of emballage used.

Four alternative advertizing campaigns were considered:

A	TV commercials
B	adds in the newspapers
C	lottery in the individual packages
D	free package (sent by mail to many families)

Four 4 different kinds of emballage were chosen. They differred in the way the product was described on the front of the packages:

I	contains calcium, ferro minerals, phosphorus and B vitamin
II	easy and fast to prepare
III	low cost food
IV	gives you energy to last for the whole day

The investigation was carried out in three cities called 1, 2, 3 og 4. The following results were optained:

Sales figures in multiples of 1000 kr. :

Emballage	City			
	1	2	3	4
I	A 52	B 51	C 55	D 56
II	B 50	C 45	D 49	A 51
III	C 39	D 41	A 37	B 39
IV	D 43	A 41	B 42	C 42

Formulate and analyze a mathematical model for this experiment. Discuss whether the model and the results found seem reasonable.

If any of the sources of variation are (statistically) significant try to see if any of the alternatives considered are better or worse than the other or, in general, if there seems to be a grouping of alternativesd with respect to effect on the sale.

Exercise 17

In an experiment with plates of brass to be used for electrical switches interest has been on the amount of wear on the plates when they are subjected to mechanical stress. Three different alloys were tested. Furthermore the temperature of the plates was varied because the temperature is known to influence the friction.

The following results were obtained:

Amount of material in mg removed under wear test:

	Temperature					
	100^0		75^0		50^0	
Alloy A	25 I	22 I	15 II	18 II	14 III	18 III
Alloy B	24 III	29 III	16 I	19 I	11 II	13 II
Alloy C	27 II	23 II	22 III	20 III	20 I	15 I

The Latin numbers I, II and III refer to three different days when the measurements were performed.

Analyze how the wear is influenced by the two factors. Remember to take into account the distribution of the measurements on the three days.

Exercise 19

The yield of a chemical process was assessed in a pilot experiment. The following factors were considered:

Factor		Factor levels 0	1
A	Amount of active compound	4 mol	5 mol
B	Acidity, pH	6	7
C	Reaction time	2 hours	4 hours
D	Filtering (first pass)	none	after 1/2 hour
E	Filtering (second pass)	none	after 1 hour

The yield relative to the theoretical maximum yield was measured. A priori the uncertainty of the experimental results is assumed to be of the order of magnitude corresponding to a standard deviation of 1.0-1.5%.

Results %	A	B	C	D	E	Construction
67.6	0	0	0	0	1	Reduction by means of
68.5	1	0	0	1	0	defining contrasts:
70.1	0	1	0	0	0	ABE and ACD
72.4	1	1	0	1	1	
78.7	0	0	1	1	1	or (equivalently)
80.1	1	0	1	0	0	E=AB and D=AC
87.7	0	1	1	1	0	(in principle)
88.4	1	1	1	0	1	

Characterize and analyze the experiment under the assumption that D and E do not interact with each other or with the other factors in the experiment.

State other necessary assumptions that you have to make in order to draw conclusions about the effects of the factors.

Exercise 20

In a chemical experiment the purpose was to analyze the influence on the yield from two factors. The factors were amount of additive added and reaction time.

Two mixes were prepared for each combination of the two factors and the reaction took place in these mix'es.

For each mix two samples were taken and the concentration (yield) was measured on each sample.

The results were

Yield from chemical experiment

Reaction time	Amount added	Mix number	Sample number	Results
1	0	1	1,2	y_1, y_2
1	0	2	1,2	y_1, y_2
1	1	1	1,2	y_1, y_2
1	1	2	1,2	y_1, y_2
1	2	1	1,2	y_1, y_2
1	2	2	1,2	y_1, y_2
2	0	1	1,2	y_1, y_2
2	0	2	1,2	y_1, y_2
2	1	1	1,2	y_1, y_2
2	1	2	1,2	y_1, y_2
2	2	1	1,2	y_1, y_2
2	2	2	1,2	y_1, y_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
4	2	2	1,2	y_1, y_2

Question

Formulate a suitable mathematical model for the yield and show how it is analyzed. Complete randomization with respect to the factors is assumed.

Exercise 21

In a certain method of analysis that is considered for standardization it is necessary to prepare a certain solution in the individual laboratories using the method, because the solution degrades rather quickly.

The results from using the method may thus vary between laboratories in that a certain equipment is used, and its calibration may also be a little different at the laboratories.

In order to evaluate whether the method is suitable for standardization a number of randomly selected laboratories participated in a test.

All laboratories analyzed both an A-sample and a B-sample that were sent to the laboratories in the same shipment. This was done twice with a certain (randomly chosen) time interval (test 1 and test 2). Also double determinations were made on both the A- and the B-sample.

The test rounds were carried out at random time points in the period of investigation, such that the individual laboratories could not exchange analysis results etc. In this way the 'test 1' round for laboratory 1 did not take place at the same time as 'test 1' at laboratory 2.

It is of primary interest to assess the deviations between the theoretically correct values and the values found by the method considered.

Deviations from double determinations

	Lab. 1		Lab. 2		...	Lab. 10	
	Test 1	Test 2	Test 1	Test 2		Test 1	Test 2
A-sample	0.71	...	...	...	...	...	0.11
	0.21	...	...	...	...	...	0.31
B-sample	0.41	...	...	...	...	...	0.81
	0.67	...	...	...	...	...	0.97

Question

Formulate a mathematical model for this experiment and indicate (by deriving proper EMS values) how the model can be analyzed.

Exercise 23

In an experiment with laboratory rats the effect of a certain hormone treatment on the content of phosphorus in the liver of the rats has been measured. The effects of sex and the age of the rats were also of interest.

An investigation was carried out where two rats for each combination of sex, age and hormone treatment were measured according to the following table of data:

Content of phosphorus in liver in mg/100 g
all results being reduced with 200 mg/100 g

		No treatment			Hormone treatm.		
4 weeks	male	-22	sum:	-48.0	5	sum:	15.0
old		-26	ssq:	8.0	10	ssq:	12.5
animals	female	-27	sum:	-57.0	-23	sum:	-38.0
		-30	ssq:	4.5	-15	ssq:	32.0
8 weeks	male	-1	sum:	4.0	15	sum:	36.0
old		5	ssq:	18.0	21	ssq:	18.0
animals	female	-21	sum:	-37.0	5	sum:	17.0
		-16	ssq:	12.5	12	ssq:	24.5

When carrying the experiment rats from 2 litters were used in order to obtain results as precise as possible (small within litter variation).

From each litter 8 young animals were used in such a way that 4 were males and 4 were females and half of each were measured after 4 weeks and the other half were measured after 8 weeks as seen from the table.

It was decided to use 2 litters because one litter containing (at least) 8 males and 8 females was difficult to obtain.

The animals were distributed on the 2 litters according to the following table where the litters are denoted “I” and “II”, respectively:

Distribution on litters		No treatment	Hormone treatm.
4 weeks	male	I	II
	female	II	I
8 weeks	male	II	I
	female	I	II

It is anticipated that the effect of the treatment can differ from one litter to litter. The three factor interaction between the three factors age, sex and treatment is assumed to be zero or at least very small.

Question 1
Characterize the experiment and formulate a suitable mathematical model for the content of phosphorus depending on treatment, age and sex (use "A" for age, "B" for treatment and "C" for sex).

Question 2
Analyze the model formulated in question 1 and compute estimates for significant effects and for the experimental variance.

Question 3
Show how the experiment can be thought of as $\frac{1}{2} \cdot 2^4$-factorial experiment where the litter represent 4th factor, and find the alias relations for this design.

Question 4
Which conclusions can be drawn from this experiment if it is again assumed that "litter" does not interact with the other 3 factors and that there, furthermore, is no three factor interaction between these 3 factors?

Exercise 24

From 5 localities (from which drinking water is planned to be obtained) samples of the water were taken. From the samples the content of fluoride in the water was determined. For each locality 3 measurements of the content of fluoride was made at random instances during the period of investigation.

The following data were obtained. The values are mg pr liter of water:

Locality	1	2	3	4	5
Measured values	1.1	2.0	0.5	1.6	1.4
(mg/liter)	0.7	2.0	0.7	2.1	0.8
	0.8	2.5	1.0	2.4	0.6
Sum	2.6	6.5	2.2	6.1	2.8
Average	0.87	2.17	0.73	2.03	0.93
Sum of squares total	2.34	14.25	1.74	12.73	2.96
SSQ within localities	0.0867	0.1667	0.1267	0.3267	0.3467

Question 1
Formulate and analyze a model for the variation of fluoride in relation to the localities. Give estimates of the significant parameters of this model and of the residual variance. When testing the model use significance level $\alpha = 5\%$.

Question 2
It is of particular interest to single out localities which have either a particularly high or low content of fluoride. Therefore, do an analysis with the purpose of splitting the whole group of localities into different, but more homogeneous subgroups with respect to content of fluoride.

Question 3
Suppose now that it is known a priori that localities 2 and 4 are ground water suppliers while 1, 3 and 5 are surface water suppliers and that this could give rise to differences in the fluoride content. For this situation set up an appropriate method of analysis and carry it out for the data at hand.

Exercise 25

A company wishes to assess the wear resistance of 4 different types of textile by means of an automatic wear tester.

The dependent variable is the weight loss (measured in units of 0.1 mg) for a piece of textile of standard shape and after being subjected to the wear tester for a specified time.

The independent variable is the type of textile, and there are 4 different types A, B, C and D.

Two noise variables are expected to be so important that is necessary to take them into account.

The first noise variable is the position in the tester in that the tester has four different positions for simultaneous testing of four pieces of textile.

The other noise variable is the round in which the textiles are tested. The round is of importance because (small) changes in the tester, air humidity, temperature etc. may influence the measured value. In order to save time and money only four rounds are carried out.

Based on this model a Latin square was carried out with 1 observation per cell:

The experiment resulted in the following measurements (y_{ijk}):

Round	Position			
	1	2	3	4
1	A(251)	B(241)	D(227)	C(229)
2	D(234)	C(273)	A(274)	B(226)
3	C(235)	D(236)	B(218)	A(268)
4	B(195)	A(270)	C(230)	D(225)

After subtraction of 220 from all data one gets

$$\sum_i \sum_j \sum_k (y_{ijk} - 220) = 312$$
$$\sum_i \sum_j \sum_k (y_{ijk} - 220)^2 = 13528$$

Question 1
Formulate the linear model corresponding to the Latin square design.

Question 2
Estimate all parameters in the linear model.

Question 3
Estimate the residual variance.

Question 4
Test in the usual way the parameters in the linear model - using the appropriate ANOVA procedure.

Question 5
Organize the textiles according to wear resistance and use an appropriate (range based) method to test, if there are significant differences between individual textiles.

Question 6
List the assumptions used in the above analysis.

Question 7
Explain your results.

Question 8
Discuss advantages and draw-backs by using a Latin square design instead of a randomized block design (using rounds as blocks) and not taking the positions into consideration.

Exercise 26

A testing agency concerning household products intends to test a claim made by a company that produces fire inhibiting products for use in cloth for fire brigades. It is claimed the products are resistant against being washed out. Two alternative fire-inhibiting products (B_0 and B_1) are considered.

The agency adopts the following framework:

The dependent variable (the response) in the fire-test is the amount of cloth which actually burns when tested using a standard fire test where, the cloth has a certain shape and size. And the following factors were considered in the particular test:

A Textile:

A_0 Type of textile = 1, A_1 Type of textile = 2.

B Fire inhibiting treatment:

B_0 Compound = 1, B_1 Compound = 2.

C Method of washing:

C_0 Low intensity washing, C_1 Normal intensity washing.

D Fire treatment direction:

D_0 Perpendicular to threads, D_1 Parallel with threads.

Based on this setup a complete 2^4 factorial experiment was carried out and using one observation per combination of the factors.

The investigation resulted in the following data:

		A_0		A_1	
		B_0	B_1	B_0	B_1
C_0	D_0	4.2	4.5	3.1	2.9
	D_1	4.0	5.0	3.0	2.5
C_1	D_0	3.9	4.6	2.8	3.2
	D_1	4.0	5.0	2.5	2.3

Question 1
Formulate the (usual) linear model corresponding to the experiment carried out.

Question 2
Use (for example) Yates' algorithm to compute contrast and sums-of-squares corresponding to the model formulated.

Question 3
Test in the usual way the parameters in the linear model by means of the usual ANOVA (use $\alpha = 10\%$). Ehen testing the model it can be assumed that three- and four-factor interactions are (at least) almost zero and thus representing the errors of the experiment.

Question 4
Estimate the significant effects of the model.

Question 5
Give a short interpretation of your findings.

Suppose now that it is wanted to do a similar experiment in which only 8 of the 16 single tests can be carried out. In order to do so the agency decides to construct a fractional factorial experiment where the four-factor interaction term acts as the defining contrast.

After drawing lots it was decided that the principal fraction (including the measurement "(1)") should be used.

Question 6
Write out the factor combinations that appear in the 2^{4-1}-factorial design.

Question 7
Determine the alias-relations of the design (with sign).

For the sake of illustration we suppose that the new experiment results in the same data as presented in the table above.

Question 8
Write out the (reduced) data table and do the Yates' algorithm for these data.

Question 9
In the usual way test the parameters in the linear model (corresponding to the reduced design) assuming the all interaction terms corresponding to two or more factors are zero.

Question 10
Give a short interpretation of your findings for the data from reduced design.

Exercise 29

An experiment has been carried out with the purpose of assessing the relation between alternative extraction methods and the amount of an aromatic compound obtained.

Material from several plants (the aromatic compound is extracted from the leaves) was used and for each plant two alternative cuttings of the leaves were applied. In order to assess whether there is important variation within the single plants 3 twigs (small branches) were selected from different positions on the plant.

Half of each twig was cut while the other half was not cut. For each twig two measurements for each cut were made.

The results from the experiment are shown in the following table:

yield im mg/(100 g leaves).

P Plant	K Twig	S No cutting		S Fine cut		Sum K	Sum P
1	1	130	132	156	163	581	1724
	2	138	135	141	144	558	
	3	129	133	165	158	585	
2	1	125	130	150	145	550	1696
	2	139	148	149	140	576	
	3	130	119	171	150	570	
3	1	168	169	190	186	713	2174
	2	171	158	188	188	705	
	3	187	178	193	193	756	
4	1	109	108	121	129	467	1427
	2	116	125	130	140	511	
	3	103	101	121	124	449	
Sum		3281		3740		7021	7021

Question 1

Formulate a suitable mathematical model to describe the yield depending on the factors described, and show how this model can be analyzed.

From an analysis of variance computer program the following decomposition of the sum of squares was obtained. The decomposition corresponds to a completely crossed scheme. The factors are named (S) for type of cut, (P) for plants, (K) for twigs and (G) for the residual:

Source of Variation	Sums of Squares	Degrees of Freedom
K	72.541	2
P	23618.160	3
KP	936.958	6
S	4294.082	1
KS	373.791	2
PS	35.416	3
KPS	717.708	6
G	12.000	1
KG	116.375	2
PG	189.166	3
KPG	128.458	6
SG	0.750	1
KSG	2.625	2
PSG	162.416	3
KPSG	106.208	6
Total	30766.640	47

Question 2
Use these figures to carry out the analysis of the model from question 1.

Question 3
Estimate the significant effects and components of variance in the model under discussion.

Use that

$$\begin{aligned} 12.000 + 116.375 + 189.166 + \cdots + 106.208 &= 718.000 \\ 1 + 2 + 3 + \cdots + 6 &= 24 \end{aligned}$$

Exercise 30

In a laboratory an experiment has been conducted with the purpose of finding out, whether it is worthwhile to carry out a pre-treatment of a certain kind of plant material from which an aromatic compound is to be extracted.

A light cooking and/or a treatment with enzymes can be performed. The plant material can also be cut, or it can be used as it is.

In the table below is shown how the experiment was conducted twice. The first replication was carried out on June the 11th and the second was carried out on September the 22nd.

At both instances an extraction device with two chambers was used, and it is anticipated that the extraction results can be (more or less) different from different chambers. The distribution on the two chambers in the two replications is shown in the table with results.

For the first replication the two-factor interaction AB was used to confound with the difference between the two chambers. For the second replication AC was used.

Question 1
Call the factors A (enzyme), B (fineness) and C (cooking). Describe the type of design used in this experiment.

Question 2
Analyze the experimental data. Use significance level $\alpha = 5\%$ in tests.

Question 3
Estimate significant effects and the variance of the experimental uncertainty, σ^2.

Table of measured data and some preliminary computations.

Yield of extraction experiment: mg aromatic compound per 100 g plant.

Results 11/6-80				Results 22/9-80			
		(C)				(C)	
	(A)	no cooking	cooking		(A)	no cooking	cooking
(B) No cut	without enzyme	124	98	(B) No cut	without enzyme	123	110
	with enzyme	171	145		with enzyme	185	154
Fine cut	without enzyme	153	136	Fine cut	without enzyme	165	158
	with enzyme	192	165		with enzyme	201	172

Number of extraction chamber, respectively

1	1	1	2
2	2	2	1
2	2	1	2
1	1	2	1

The following preliminary computations have been performed:

(1)	124	295	640	1184	(1)	123	308	674	1268
a	171	345	544	162	a	185	366	594	156
b	153	243	86	108	b	165	264	98	124
ab	192	301	76	-26	ab	201	330	58	-56
c	98	47	50	-96	c	110	62	58	-80
ac	145	39	58	-10	ac	154	36	66	-40
bc	136	47	-8	8	bc	158	44	-26	8
abc	165	29	-18	-10	abc	172	14	-30	-4
$124^2 + 171^2 + \cdots + 165^2 = 181240$					$123^2 + 185^2 + \cdots + 172^2 = 207344$				
$1184^2/8 = 175232$					$1268^2/8 = 200978$				
$(1184 + 1268)^2/16 = 375769$									

Exercise 31

An American farmer wished to find out whether the application of potash has a positive influence on the fiber strength of the cotton he grows on his fields.

Therefore he carries out an experiment where he applies different amounts of potash.

At the same occasion he chooses to investigate whether it would be beneficial to use another type of cotton plants instead of the presently used type, called A-cotton.

A company which markets certain new types of seeds advertises that it now has 2 new types, B1 and B2-super, which both are claimed to produce fibers with high strength.

The farmer decides to apply 4 different amounts of potash pr acre, namely 0, 10, 20 and 30 kg. These amunts are used for all three types, and he distributes the resulting 12 combinations of types of seeds and amounts of potash on 12 plots randomly placed on a larger field which he assumes is homogeneous whith respect to growing cotton.

after the growing season he harvests his cotton, but unfortunately two of the plots have been attacked by fungi and the results from these plots are unreliable and thus disregarded:

Index of strength for fibers of cotton (high index = high strength)

Potash	Type of seed		
amount	A	B1	B2
0	7.62	8.02	7.93
10	disreg.	8.15	8.12
20	7.76	8.73	8.74
30	8.00	disreg..	8.75

Which conclusions can, based on this experiment, be drawn concerning the application of potash as fertilizer and concerning the possible introduction of a new type of cotton if he wishes to obtain a high index of strength.

The solution to this exercise illustrates a number of alternative approaches illustrating the connection between regression analysis and analysis of variance.

Exercise 33

Vinasse is a by-product which is produced when melasse is used for the production of technical alcohol. Vinasse contains some nutrients and it is of interest to evaluate its potential as food for cows. The experiment was carried out using Red Danish milking cows.

One problem when applying vinasse as food is that the cows do not seem to like it very much. Therefore it is necessary to mix the vinasse with another more tasteful kind of food in order to make the cows eat the vinasse.

5 different mixes with melasse added in different amounts were tested, and for each mix it was noted how much vinasse the cows did eat per day during the test period. The experiment was carried out using 10 twin pairs according to the design below:

Twin pair	Diet number				
number	1	2	3	4	5
1	x				x
2	x		x		
3	x			x	
4	x	x			
5			x		x
6				x	x
7		x			x
8			x	x	
9		x	x		
10		x		x	

For each of the combinations marked with 'x' the daily intake of melasse was recorded.

Question 1

Characterize the experimental design and explain shortly the advantages of the design. Formulate a mathematical model the intake of vinasse in relation to different diets and taking into account the influence from the twin pairs.

Question 2

The results of the experiment are shown in the following table

Twin pair number	Diet number 1	2	3	4	5	Sum
1	0.2				4.2	4.4
2	0.6		3.1			3.7
3	0.9			5.2		6.1
4	0.8	1.3				2.1
5			4.2		5.1	9.3
6				2.9	3.8	6.7
7		3.1			6.1	9.2
8			2.0	1.9		3.9
9		1.0	2.2			3.2
10		2.1		4.6		6.7
Sum	2.5	7.5	11.5	14.6	19.2	55.3

Analyze the model formulated in question 1. Focus on whether the intake of vinasse depends on the diet used.

$$
\begin{aligned}
0.2^2 + 4.2^2 + 0.6^2 + \ldots + 4.6^2 &= 210.17 \\
2.5^2 + 7.5^2 + \ldots + 19.2^2 &= 776.66 \\
4.4^2 + 3.7^2 + \ldots + 6.7^2 &= 361.03 \\
55.3^2 &= 3058.09
\end{aligned}
$$

Question 3

The different diets correspond to adding melasse in the amounts 0.5:1, 1:1, 1.5:1, 2:1 and 2.5:1, respectively, where the amount of melasse is mentioned first. Call the ratio between melasse and vinasse z (=1/2 $\times$ number of diet) and test a hypothesis that the intake can be considered to be linearly related to z such as:

$$Y = \mu + a \cdot z + E$$

Question 4

Estimate the possible linear relation between the vinasse intake and the z-ratio. Also, estimate the experimental variance σ^2.

Exercise 34

A laboratory was given the task to analyze the dispersion stability of a certain type of paint in relation to a number of factors. The following factors, all at three levels, were to be considered.

A:	Type of solvent 0: 50:50 xylol-butanol 1: xylol 2: 50:50 xylol-ethanol	C:	Disperser type 0: tetramethyl ammonium hydroxide 1: ammonium hydroxide 2: morpholine
B:	Amount of factor A (solvent) 0: 40% 1: 50% 2: 60%	D:	Mixing method 0: ultra sound 1: Waring-blender 2: Manton Gauling Colloid mill

The measurement of the dispersion stability is done by producing a certain amount of paint which then is stored under controlled conditions for a fixed time period after which the degree of dispersion in the paint is assessed microscopically.

Question 1

Initially it is suggested to carry out an investigation in which all factors and their combinations are taken into account, that is $3 \times 3 \times 3 \times 3$ measurements. Unfortunately, however, the laboratory cannot store so many samples at the same time. Therefore the experiment has to be split up. Assuming that the laboratory can store, for example, about 30 samples you are asked to work out a reasonable design for the experiment. You do not need to write out the detailed plan, but it suffices to show the principle used and how a few treatment combinations are handled.

Question 2
After seeing the design and calculating the total cost of the experiment it is agreed that it probably is not necessary to carry out all possible factor combinations. It is therefore decided to implement a fractional design using 27 measurements, that is a $\frac{1}{3} \cdot 3^4$-factorial design.

Work out such a design (give examples on measurements to be performed and measurements not to be performed and examples of alias relations for some interesting effects).

Question 3
In order to improve the design worked out in question 2 the laboratory suggests to block the experiment such that it is split into 3 blocks each having 9 measurements. The reason

for this is that from one mix (batch) of raw material it is not possible to formulate all 27 samples required for the design, but 9 samples can be made in a practical and homogeneous way. (You may also answer this question by showing the principle for constructing the design and by finding of few measurements that have to be conducted and their allocation to a block.)

As an aid in solving the exercise the so-called "standard" sequence of effects in a 3^4 factorial is given:

I	D	BCD
A	AD	BCD^2
B	AD^2	BC^{2}D
AB	BD	BC^{2}D^2
AB^2	BD^2	ABCD
C	ABD	ABCD^2
AC	ABD^2	ABC^{2}D
AC^2	AB^{2}D	ABC^{2}D^2
BC	AB^{2}D^2	AB^{2}CD
BC^2	CD	AB^{2}CD^2
ABC	CD^2	AB^{2}C^{2}D
ABC^2	ACD	AB^{2}C^{2}D^2
AB^{2}C	ACD^2	
AB^{2}C^2	AC^{2}D	
	AC^{2}D^2	

Exercise 38

In an experiment the purpose is to evaluate the influence from using certain sugar types in the production of mycelial inoculum which is a material used in the production of penicillin.

The material is grown in a medium which contains the necessary ingredients and the process takes place under controlled circumstances (temperature, humidity, clean air, etc.).

In the experiment a 4% addition of the different types of sugar and the following data were found:

Wet weight of inoculum grown using different types of sugar

Sugar type	Technical lactose	Technical glucose	Sucrose	Melasse	Dextrin
Yield	2.0	1.8	2.3	2.4	3.4
in five	2.1	1.4	2.8	2.3	3.9
experiments	1.6	1.7	2.7	2.9	3.8
	2.1	1.9	2.4	3.0	3.7
	2.0	2.0	1.9	3.1	2.9
Total	9.8	8.8	12.1	13.7	17.7

$2.0 + 1.8 + ... + 2.9 = 62.1$ and $2.0^2 + 1.8^2 + ... + 2.9^2 = 166.25$

Question 1 Do an analysis which clarifies if there a differences between yields of inoculum when using different types of sugar.

Question 2 Can one or more of the sugar types be identified as giving a yield which is significantly different from yields from other sugar types.

Question 3 It is concluded that dextrin is suitable to add, and in order to assess optimal amount to add another experiment is conducted. The data were:

Amount of dextrin	2%	4%	6%	8%
Yield	1.9	3.1	3.4	2.8
in four	2.6	3.4	3.6	2.9
experiments	2.3	3.0	3.0	3.1
	2.7	3.3	2.9	2.6
Total	9.5	12.8	12.9	11.4

Using these data, assess how the yield depends on the amount of dextrin used by estimating the mean yield as a polynomial such as:

$$E(Y) = a + bx + cx^2 + dx^3$$

where z is the amount of dextrin (in %). It is given that $1.9 + 3.1 + ... + 2.6 = 46.6$, and $1.9^2 + 3.1^2 + ... + 2.6^2 = 138.56$

Exercise 39

A Martindale Wear Tester is a machine used to measure mechanical wear resistance. The machine is constructed with four plates on which an abrasive surface is mounted. The material to be tested is mounted in the machine and is moved mechanically over the abrasive surface. The weight loss after a certain number of movements is taken as an indicator of the wear resistance.

In the present case a laboratory is asked to compare 5 different types of plastic which are to be used for coating certain electrical parts. The parts are used in building machinery where the mechanical wear can be substantial.

An experiment was conducted and the following results were obtained.

Type of plastic	A	B	C	D	E	sum
Round 1	79	139	90	74		382
2	93	144	96		125	458
3	58	136		76	136	406
4	106		75	95	101	377
5		125	73	114	95	407
Sum	336	544	334	359	457	2030

$$79^2 + 139^2 + \ldots + 95^2 = 218558$$

The data are measured amounts of material worn off in the tester.

Question 1

What is the name of the design used. Explain the idea behind the design and formulate a suitable mathematical model for the measured values in relation to the types of plastic and the rounds.

Question 2

Perform an analysis (of variance) of the influence of the different types of plastic on the wear resistance and try to figure out whether the plastic types can be grouped.

Question 3

The contractor of the investigation asks whether the laboratory has taken into account that the four mounting plates can be different. In that connection it turns out that the investigation was carried out as follows:

Positions a–d	A	B	C	D	E
Round 1	d	c	b	a	
2	d	c	a		b
3	d	a		b	c
4	b		c	a	d
5		a	c	b	d

3a) What are the problems using this design (explain shortly)

3b) Suggest an alternative way of conducting the experiment in case is has to be repeated (explain shortly). Analyze the above results under the assumption that the suggested alternative was in fact used.

Exercise 41, new

This is a cvorrected version of exercise 41. The table with sums of squares has been corrected. That is all. The solution will also change a little, but not substantially.

In a precision study the intention was to assess which factors influence the uniformity with which certain ceramic tiles are produced.

The tiles are to be used for inner linings for ovens. In the present study the heat transfer ability of the tiles is focused on. The reason for this is, that experience has shown that the heat transfer varies pretty much between tiles with the result that large temperature differences can occur behind the tiles which, in turn, may cause tiles to fall off.

Among the possible causes of variation is the clay from which the tiles are produced, and it can vary from one shipment to the next. One shipment will typically suffice for one week of work in the plant.

Another cause can be the pre-treatment and addition of certain additives and colors, because insufficient mixing and lack of homogeneity may occur. This part of the manufacturing takes place in batches which typically correspond to 12 hours of production.

Finally at the final part of the manufacturing, covering cutting and burning of the individual tiles, there may occur differences in size and degree of burning which can lead to differences in the heat transfer ability. This part of the manufacturing takes place in so-called 'productions' formed from the abovementioned batches. Typically one batch leads to about 20-25 'productions'.

In the analysis of the heat transfer ability raw material from 3 different weeks was used. From each week 3 batches were selected and from each batch samples from 4 'productions' were measured. From each 'production' 4 tiles were measured, as shown in table 1 below.

Question 1
Formulate a suitable mathematical model to describe the variation of the K-values (heat transfer coefficient) where the described possible sources of variation are taken into account. Also explain how the model can be analyzed (using EMS-values).

Question 2
By means of the sums of squares shown in table 2 (which corresponds to a completely cross classified splitting of the total variation) construct the proper analysis of variance table corresponding to the model formulated and do the analysis for the K-values.

Question 3
Give estimates corresponding to the significant sources of variation including the variance corresponding to the residual variation (the variation between tiles from the same 'production').

Produc-tion	Week 16			Week 22			Week 31		
	Batch			Batch			Batch		
	1	2	3	1	2	3	1	2	3
1	13.7	23.0	25.6	39.3	28.2	29.6	28.3	32.3	28.0
	17.9	23.3	24.4	30.5	27.6	27.4	29.5	29.9	29.9
	17.4	24.3	26.0	28.6	27.5	27.5	29.3	33.2	28.5
	17.9	25.9	25.1	27.6	29.7	27.7	28.4	31.7	30.0
2	26.8	25.7	25.6	29.9	26.0	30.0	30.7	27.9	26.0
	25.5	23.2	24.3	30.4	24.9	29.6	29.7	30.4	27.1
	26.1	24.9	24.5	28.6	26.0	28.2	30.4	29.2	27.6
	25.7	23.6	24.9	29.1	25.9	29.4	30.4	27.3	27.6
3	24.9	26.1	26.6	27.9	30.5	26.1	28.5	28.4	25.4
	25.7	26.3	29.2	27.5	29.9	26.1	26.5	26.4	25.0
	25.8	26.4	27.1	27.6	30.9	26.2	28.2	27.1	26.4
	26.1	26.5	29.0	26.8	31.4	25.8	26.9	26.8	26.9
4	22.6	27.8	29.1	31.4	29.7	28.0	31.8	25.8	25.8
	23.7	27.1	27.1	32.0	31.0	28.7	31.1	27.5	26.3
	22.6	26.8	27.2	30.8	30.2	26.8	30.0	26.2	24.0
	23.0	29.6	29.7	32.8	30.3	27.3	30.2	27.2	27.2

Table 1. K-values×10 for ceramic tiles.

Sources of variation	SSQ	Degrees of f.d.	
W	411.3	2	(W = Week)
B	7.1	2	(B = batch)
BW	204.0	4	
P	19.1	3	(P = Production)
PW	230.4	6	
PB	87.5	6	
PBW	209.0	12	
G	4.2	3	(G = Repetitions)
GW	15.0	6	
GB	9.9	6	
GBW	21.7	12	
GP	11.4	9	
GPW	47.1	18	
GPB	14.0	18	
GPBW	56.7	36	
I alt	1348.4	143	

Table 2. Sums of squares computations for ceramic tiles.

Exercise 42

As part of a so-called “response surface”-experiment 5 control variables (factors) were considered. In order to be able to estimate the best direction for a search for optimal set points for the factors a screening experiment was conducted.

Initially it is assumed that the 5 factors influence the response additively (to a reasonable approximation). Therefore a reduced factorial experiment is suggested. Specifically a $\frac{1}{4} \cdot 2^5$ factorial was conducted.

The five factors in question were :

A:	temperature	(70°C, 80°C)
B:	concentration of active ingredient	(40%, 60%)
C:	reaction time	(30 min, 40 min)
D:	concentration of catalyst additive	(0.10%, 0.20%)
E:	pH in solution	(5.5, 6.5)

in that the setpoints until now were (75°C, 50%, 35 min, 0.15%, 6.0), where the yield of the process, based on experience, only is about 50% of the theoretical maximum yield.

Question 1
Following the suggestions of the statistical department of the laboratory, the following experiment was conducted, in that the usual standard notation is used:

		A_0				A_1			
		B_0		B_1		B_0		B_1	
		C_0	C_1	C_0	C_1	C_0	C_1	C_0	C_1
D_0	E_0	(1)							abc
	E_1				bce	ae			
D_1	E_0			bd			acd		
	E_1		cde					abde	

The construction of the design is based on the defining relations $I_1 = BCD$ and $I_2 = ACE$.

Show that this design is adequate (under the given assumptions) for determining the dependence of the yield of the 5 factors, and show how the design can be constructed.

Question 2

The experiment given above resulted in the following response data:

$$\begin{array}{lcl} (1) & = & 36\% \\ ae & = & 45\% \\ bd & = & 45\% \\ abde & = & 56\% \\ cde & = & 39\% \\ acd & = & 30\% \\ bce & = & 76\% \\ abc & = & 61\% \end{array}$$

Based on these data you are asked to estimate the direction of the steepest ascent (no testing is wanted, only estimation).

Question 3

Suppose a new experiment with the 5 variables (factors) having the values (65°C, 70%, 40 min, 0.25%, 5.0) is to be conducted.

What is the estimate of the expected yield for this experiment, based on the above results?

Exercise 46

The purpose of a certain laboratory experiment is to assess the influence on the yield of a chemical process of a number of control variables. In the process a pharmaceutical product is produced.

The influence of the following factors is considered:

A: acidity during reaction	(pH = 5, pH = 7)
B: concentration of catalyst	(5%, 10%)
C: temperature of reaction	(40°C, 60°C)
D: filler ($CaCO_3$)	(5:1, 10:1)

The process is currently running with a yield about 60% of the maximum attainable yield. The purpose of the experiment is to find out whether the yield changes if one or more of the above factors are changed within the limits given.

Question 1
Discussion has been about how a suitable experimental design could be. The full 2^4 factorial cannot be carried out under completely stable experimental conditions because the individual measurements can require as much as two hours of work. Thus, the experiment, if a complete factorial is chosen, must be distributed over 4 days. The AB interaction is of particular interest, and therefore this term should not be confounded with days. Under this requirement a design using 4 days is wanted.

Construct a suitable plan and show the distribution of the 16 measurements on the 4 days.

Question 2

As an alternative to the above complete 2^4 factorial design is is suggested to take into account, that the interactions between more than two factors generally can be considered small, and that the factor D, especially, does not at all interact with other factors, such that a fractional design could be relevant.

Therefore, construct a $\frac{1}{2} \cdot 2^4$ design in such a way that, under the sketched assumptions, main effects for all four factors and two-way interaction terms for the factors A, B and C can be estimated (no blocking considered in this question).

Find the alias relations and the individual measurements of the design you suggest.

Question 3

Since there is still problems about controlling the experimental conditions, a suitable distribution over two days of the experiment is wanted, that is using 2 blocks each containing 4 observations.

The BC interaction term can be used to confound with the day-to-day variation.

Show how the 8 treatments found in question 2 can be distributed over the 2 days in a suitable way.

Exercise 47

In a psycho-acoustic experiment with farm pigs the purpose was to find out whether noise affects the usage of the food given to the pigs, the growth rate and the quality of the meat. In the experiment a number of alternative artificial schemes of noise were constructed from ventilation noise and applied at different levels of noise. A natural noise scheme recorded in a real pig farm was used as control.

In the experiment teams each consisting of 3 pigs (all female) were used. The experiment ran from the pigs weighed 23 kg until they weighed 89 kg. 3 pigs were used in each team in order to make sure sure that data were obtained for all teams. The measured values were the average values obtained for 2 of the pigs in the team selected randomly among the 3 participating if all 3 pigs in the team were sound and alive at the end of the experiment. If one of the pigs was sick, dead or otherwise not well the average of the 2 other pigs was used.

In all teams at least 2 sound and well pigs could be used.

The individual teams were formed from litters of pigs. 10 litters were used and from each litter 3 teams consisting of 3 pigs were formed.

In the following design one 'x' indicates one team.

	Litter number									
Noise scheme	1	2	3	4	5	6	7	8	9	10
Constant 75dB(A) A	x	x	x	x	x	x				
Varying 75dB(A) B	x	x	x				x	x	x	
Constant 45dB(A) C	x			x	x		x	x		x
Varying 45dB(A) D		x		x		x	x		x	x
Natural noise E			x		x	x		x	x	x

Question 1

What is this design called, and which mathematical model is usually used for it. What can be estimated using this design.

Question 2

The response considered here is the growth rate of the pigs measured in g/day during a certain period of time in the experiment.

The results obtained are shown in the following table where all data have been reduced with 600 g/day

	Litter number										
Noise scheme	1	2	3	4	5	6	7	8	9	10	Sum
Constant 75dB(A) A	5	5	1	10	-15	7					13
Varying 75dB(A) B	5	5	-2				4	24	22		58
Constant 45dB(A) C	30			35	-15		-5	29		2	76
Varying 45dB(A) D		13		38		19	3		6	13	92
Natural noise E			23		15	15		37	23	30	143
Sum	40	23	22	83	-15	41	2	90	51	45	382

Based on these data analyze the influence from the noise scheme on the growth rate.

3a) Estimate the growth rate in g/day under the 5 different noise schemes. Also, estimate the measurement uncertainty (variance).

3b) Perform an analysis which reflects whether the differences between the 4 artificial noise schemes can be related to the differences between noise levels and/or the differences between constant and varying noise schemes.

Exercise 48

An investigation was carried out concerning the application of aluminum plates for coverage of outdoor equipment such as cabinets, for example.

Plates of aluminum were placed (for a certain period of time) in an environment similar to what could expected in daily use.

As test sites for the experiment 3 different production plants using the plates in question were chosen.

At each of the test sites a number of persons were selected and they were asked to evaluate the appearance of the plates without knowing which aluminum alloys were actually used for the individual plates. The following table shows the data obtained:

Results from evaluators:

Plant	Evaluator	Alloy (L)					Sums
(S)	(B)	1	2	3	4	5	
	A	18	19	13	15	14	79
I	B	21	21	15	16	16	89
	C	18	18	10	13	14	73
	D	19	15	12	14	12	72
	Sum	76	73	50	50	56	313
	E	16	16	13	12	13	70
II	F	16	16	11	10	9	62
	G	15	14	10	10	10	59
	H	14	14	10	8	8	54
	Sum	61	60	44	40	40	245
	I	13	14	8	8	7	50
III	J	13	16	11	10	11	61
	K	17	20	5	8	4	54
	L	15	15	11	9	11	61
	Sum	58	65	35	35	33	226
Sums		195	198	129	133	129	784

Question 1
Formulate a suitable mathematical model to describe the variation of the evaluations depending on the plants, the alloys and the evaluators (persons). Also explain how the model can be analyzed (using EMS-values).

Question 2
By means of the sums of squares shown in table 2 (which corresponds to a complete cross classified splitting of the total variation) construct the proper analysis of variance table corresponding to the model formulated and do the analysis for the evaluations.

effect	SSQ	f
S (Plant)	209.23	2
B (Evaluator)	29.73	3
SB	51.57	6
L (Alloy)	439.06	4
SL	29.94	8
BL	34.28	12
SBL	63.92	24
Total	857.73	59

Question 3

Estimate significant effects.

Question 4

Among the 5 alloys the alloys 1 and 2 contain an anti-corrosive agent, while 3, 4 and 5 do not not contain such an additive. Do an evaluation whether this difference can be a main reason for the differences seen in the evaluations of the appearances of the plates.

Exercise 49

In an experiment concerning production of a certain type of penicillin two factors were considered. The measured value is the yield of the production process.

De two factors are

A:	Catalyst	0%, 0.02%, 0.04%
B:	Glucose	0%, 0.25%, 0.50%

The impact of the two factors are to be evaluated separately and other compounds in the growth medium are kept as constant as possible.

Two experiments were conducted:

Experiment 1

Yield

B	A 0.00	0.02	0.04	sum
0.00	54	58	64	176
0.25	65	80	39	184
0.50	94	47	38	179
sum	213	185	141	539

Experiment 2

Yield

B	A 0.00	0.02	0.04	sum
0.00	32	64	100	196
0.25	87	93	39	219
0.50	120	55	35	210
sum	239	212	174	625

Since the experiment is time consuming and difficult to control a blocking was used. This blocking is shown in the table below, and it is seen that a block size of 3 was used. The Latin numbers I, II,...,VI indicate the 6 blocks.

Experiment 1

	A		
B	I	II	III
	III	I	II
	II	III	I

Experiment 2

	A		
B	V	VI	IV
	VI	IV	V
	IV	V	VI

Question 1
Characterize the design and explain the purpose and the idea of the blocking principle used. What is the design called?

Question 2
From a computer program that can compute sums-of-squares according to "Kempthorne's method", the following results were obtained for the two experiments separately and commonly:

Experiment	effect	totals for levels 0	1	2	ssq	f
1	A	213	185	141	878.22	2
	B	176	184	179	10.89	2
	AB	140	161	238	1774.89	2
	AB^2	172	191	176	66.89	2
2	A	239	212	174	710.89	2
	B	196	219	210	89.56	2
	AB	126	186	313	6077.56	2
	AB^2	160	223	242	1228.22	2
Commonly	A	452	397	315	1584.33	2
1+2	B	372	403	389	80.33	2
	AB	266	347	551	7189.00	2
	AB^2	332	414	418	785.33	2

The total sum of squares is $ssq_{\text{total}} = 11248.00$

Analyze the importance of the two factors for the yield by means of these computations and the data.

Question 3
Give estimates for the yield depending on the factors A and B and, if relevant, of the interaction between these factors.

Which combination (A,B) would you recommend to apply if the highest possible yield is wanted?

Exercise 50

An experiment has been conducted in order to study the influence from at certain pigment to be used in paint. Four different types of paint were used for studying the effect of the pigment. The measured variable is a reflective index.

The procedure was that 4 types of paint were prepared and subsequently the paint was applied to a test surface using 3 alternative methods of applying the paint.

Of interest is of course the influence on the reflective index from both the types of paint and the methods of applying the paint.

From experience it is known that some variation can occur when producing the different types of paint. Therefore new batches of the types of paint were produced on each of the days. Thus, one batch of paint type A was produced on day 1 and a new batch was produced on day 2, etc.

On the single day enough paint of each type to be used for all 3 application methods was produced. Thus there is a restriction on the randomization. It could be illustrated as in the following table giving the measurement sequence:

Randomization					
Day	Application method	Type of paint A	B	C	D
1	1	3	11	5	7
	2	1	12	4	9
	3	2	10	6	8
2	1	20	13	17	24
	2	21	14	16	22
	3	19	15	18	23
3	1	28	31	36	27
	2	29	32	34	26
	3	30	33	35	25

Data obtained

Day	Application method	A	B	C	D	
		Type of paint				
1	1	64	66	74	66	270
	2	68	69	73	70	280
	3	70	73	78	72	293
		202	208	225	208	843
2	1	65	65	73	64	267
	2	69	70	74	68	281
	3	71	72	79	71	293
		205	207	226	203	841
3	1	66	66	72	67	271
	2	69	69	75	68	281
	3	70	74	80	72	296
		205	209	227	207	848
Totals		612	624	678	618	2532

Question 1
Comment on the randomization and the consequences with respect to how to carry out the analysis of variance. Formulate a reasonable mathematical model for the reflective index. Explain how the model is analyzed, and explain in particular how to interprete the variation between types of paint within one day and between application methods within one type of paint on one day.

Question 2
From an analysis of variance computer program the following complete decomposition of the sum of squares was obtained. Use these sums-of-squares to test the influence on the reflective index from, especially, the type of paint and the application method.

D = day, M = method for application, T = type of paint

TERM	SSQ	df
D	2.167	2
M	228.667	2
DM	1.667	4
T	308.000	3
DT	5.833	6
MT	12.667	6
DMT	11.000	12
Total	570.000	35

Question 3
Estimate significant effects in the model under consideration and also the residual variance.

Exercise 51

In an experiment about extraction of a dyestuff from a certain kind of plants different methods of pre-treatment was considered. The purpose of the pre-treatment is to degrade the cell walls in the plants in order to facilitate the extraction of the dyestuff.

In the experiment two kinds of enzymes that were thought to be suited for this purpose were tried. Some plants were not treated with enzymes, but instead they were cooked for 20 minutes.

Since the plant material can vary with time (because of different deliveries at different days) the analyses were carried out on 4 different days with some time interval between them. Thus there may be some variation between days but within one day the plant material can be considered relatively homogeneous.

The following results for the experiments given as mg dye extracted per experiment were found:

	Treatment			
Day no.	Cooking	Enzyme α	Enzyme β	Sums
1	21	32	33	86
2	21	29	37	87
3	51	72	66	189
4	30	39	54	123
Sums	123	172	190	485

The following computations are given

$$
\begin{aligned}
21^2 + 32^2 + \cdots + 54^2 &= 22683 \\
123^2 + 172^2 + 190^2 &= 80813 \\
86^2 + 87^2 + 189^2 + 123^2 &= 65815 \\
485^2 &= 235225
\end{aligned}
$$

Question 1
Formulate a mathematical model for this experiment. What is a design as the one considered called? Which assumptions are usually made concerning the effects of the variables 'day' and 'treatment' and their possible interaction? Carry out an analysis of how the alternative treatments influence the yield of dyestuff.

Question 2
Estimate all effects in the model formulated in question 1 including an estimate of the day-to-day variance and of the residual variance (corresponding to the pure experimental error).

Question 3
In order to choose the future method of production it is of particular interest to assess whether the variation of the data related to treatments can be attributed to the enzyme treatment as such as opposed to cooking, and if there is possibly a (significant) difference between the two enzyme treatments. Carry out an analysis which answers these two questions.

Question 4
Since there seems to be large differences between results obtained on different days it could be of interest to try to do some selection of batches of plant material, that could be expected to give a high yield. One day typically corresponds to one batch. Therefore do an analysis which answers the question whether one or more of the batches (days) seem to be different from (perhaps better than) the other batches.

Exercise 52

Some small electronic Al components to be used for controlling work shop machinery are considered. Interest is on a certain compound in which the components are embedded and qualities such as electrical conductivity, resistance to humidity and mechanical wear resistance of the compound are measured.

The compound is a plastic compound, and its qualities may depend on a number of factors. In the present experiment the following factors are considered:

A:	addition of hardener	(10%, 15%)
B:	hardening temperature	(60°, 80°)
C:	pretreatment (heat) of raw product	(no, yes)
D:	concentration of solvent in raw product	(4%, 8%)
E:	humidity in hardener	(dry , humid)

For the factors A and B it is assumed that, besides main effects, there can be an interaction of importance, while it for all the factors C, D and E can be assumed that they (predominantly) have additive effects.

Question 1
Initially it is clear that it should not be necessary to conduct measurements for all possible factor combinations, and the aim is therefore to work out an experimental design based on only 1/4 of the full factorial design. Work out such a design by introducing the factors D and E into the complete factorial design defined by the factors A, B and C.

Show the alias relations of the design suggested. Comment.

Work out the particular design in which the measurement based on A = 10%, B = 60°, C = no, D = 4% and E = dry is included.

Question 2
The described experiment can hardly be carried out under constant conditions, since the fabrication of the individual test items is cumbersome (takes time). Therefore is is decided to do the experiment using two days with 4 measurements pr day.

Suggest a suitable design which distributes the 8 measurements over two days.

Question 3
After conducting the first experiment it is decided to carry out an extra experiment where focus is on the 4 most important factors A, B, C and D.

A $\frac{1}{2} \cdot 2^4$ factorial experiment where the factor D is introduced by the relation D=ABC is conducted.

Show that the following design is an adequate design and show the alias relations for the design.

The following measurements were made:

(1) ab ac ad bc bd cd abcd

The following results for an index of wear resistance were obtained:

$$\begin{array}{lcl} (1) & = & 60 \\ \text{ad} & = & 70 \\ \text{bd} & = & 55 \\ \text{ab} & = & 61 \\ \text{cd} & = & 84 \\ \text{ac} & = & 68 \\ \text{bc} & = & 57 \\ \text{abcd} & = & 83 \end{array}$$

This experiment was carried out without blocking.

Question 4
In an earlier experiment an estimate of the experimental variance was found $\hat{\sigma}^2 = 2.2^2 = 4.84$, and it is based on 4 degrees of freedom.

Perform an analysis of the results obtained under the assumption that the factor D does not interact with the other factors, but that there can be two-factor interactions between the factors A, B and C.

Exercise 54

The purpose of this exercise is to get acquainted with or repeat a number of fundamental concepts and principles that are important in constructing experimental designs in the laboratory, for example.

Generally you should try to formulate your reply in a way as if you were a statistical consultant giving advice to the laboratory management. And you are free to elaborate on your own on the problems you may identify or find interesting to pursue.

One important task is to prevent or reduce the influence from unwanted, but inevitable sources of variation that may invalidate the precision of the experimental data and subsequently of the estimation and testing.

We imagine that we want do an experiment where a pharmaceutical substance is to be tested for side effects on the treated individuals.

The experiment is carried out on animals (rats fx), and the quantity measured is the amount of a certain residual chemical substance resulting from the metabolism of the product in question in the blood serum of the animals.

The objective is to give the pharmaceutical substance to the animals in alternative doses, $d_1 = 20$ mg/animal, $d_2 = 50$ mg/animal and $d_3 = 80$ mg/animal, and 24 hours later the amount of the residual substance in the serum is measured. The measured value is denoted Y.

Question 1
What do we usually call the variable $d_i,\ i = 1, 2, 3$ in such an experiment. What do we call the variable Y (there are a number of possibilities in both cases)?

Question 2
Which proporties of the animals could (should!) be considered, when they are selected for participation in the experiment.

Set up an imaginary specification table for the animals which could be used in the future when the laboratory purchases animals from its animal supplier. The specification should primarily be set up with focus on the proporties of the animals that you think could influence the intended measurements. Obvious examples are (of course) gender and age, but what else?

Question 3
Suppose it is required that the animals must be of the same 'family'. Discuss how 'family' could be defined, and how the degree of being of 'the same family' or not may influence the experimental results for animals that are treated in the same way.

Question 4
Suppose it is only required that the animals have the same weight. Discuss how this

weight requirement may influence the experimental results in contrast to the situation where the weight of the animals used is allowed vary in a natural way.

Question 5
Mathematically the outcome of an experiment made on one animal can (in the usual way) be written

$$Y_{ij} = \mu_i + E_{ij} = \mu + \alpha_i + E_{ij}$$

where μ_i denotes the theoretical mean and E_{ij} is the random deviation in measurement no. j using the i'th treatment. The index j can be the number of an animal. α_i is the socalled effect of treatment i, $\sum \alpha_i = 0$.

Y_{ij} is the amount of the residual substance in the blood serum – as described above.

What kind of figure is μ_i ?
What kind of figure is E_{ij} ?

Explain how the above model could be modifyed if more than one measurement is taken per animal (does it change our conception of μ or E_{ij}).

Discuss the randomization of the model suggested if only one measurement is taken for each animal, and if, alternatively, say, two measurements are taken per animal.

Question 6
You are also asked to specify how the proporties taken up in the specification list, question 2, and in questions 3 and more detailed may influence μ_i and/or E_{ij} in the model (give a few examples).

Question 7
Again we consider one of the treatments, d_i. For this treatment r animals are selected with results as follows (one measurement per animal):

d_i
Y_{i1}
$\vdots$
Y_{ir}

We still assume that $Y_{ij} = \mu_i + E_{ij}$. The treatment mean μ_i is to be determined with a certain precision. In the actual experiment there is a certain mix of measurement uncertainties, biological variation, natural differences between animals etc. All this amounts to a standard deviation of E_{ij} of about 0.2 mg in the determination of Y_{ij} for one blood sample.

How many animals (r) are needed if μ_i is to be determined with a precision such that the variance:

$$V(\hat{\mu}_i) \leq (0.07 \text{ mg})^2.$$

How many animals are needed if the width of a 95% confidence interval for μ_i shall be at most 0.20 mg ?

Question 8

Suppose that we in question 7 can improve the precision of the measurements by reducing the biological variation such that the standard deviation of E_{ij} becomes only 0.15 mg. How many animals do we then need in order to comply with the requirements of question 7?

Question 9

Before starting the actual experimentation alternative experimental designs are discussed.

It is firstly assumed that the animals (rats) that are bought only satisfies the requirement that they are of about the same age and in the same feeding condition (equally fed), and that about half are females, and so on.

It is suggested not to take these proporties into consideration, because the purpose is to determine μ_i in general - as it is argued (!).

The animals are thus allocated to the treatments by a random selection procedure. The following table illustrates this by showing the numbering of the individual animals and the treatments. The measurements are carried out in random order. Animal no. 1 is allocated to d_2 etc.

Treatments	d_1	d_2	d_3
No. of	11	1	2
animal	10	3	5
	7	20	13
	22	14	24
	8	18	21
	4	23	19
	16	6	15
	12	9	17

Animals 1-12 are all females and the others are 12 males randomly numbered within genders.

Discuss problems you see by applying this design.

What do we call the “process”, by which the animals with respect to gender are allocated to the three treatments ?

Which other proporties of the animals are being handled in the same way as gender. Name at least three such proporties.

Question 10

Suppose the gender of the animal influences the metabolism in such a way that it for male animals can be assumed that the amount of the residual compound in the blood is about 0.1 mg lower than the overall μ_i for any treatment i and for females it is (correspondingly) about 0.1 mg higher than μ_i .

How will this (assumed effect) influence the three estimates $\hat{\mu}_1$, $\hat{\mu}_2$ og $\hat{\mu}_3$ in the concrete design shown above.

Question 11
Construct an alternative design which eliminates the above problem about the gender of the animals and formulate the proper mathematical model for this alternative design. Call the random error in the model $Z_{ij..}$ (properly indexed corresponding to your suggested design).

Question 12
It is assumed that the random error, E_{ij} in the first design has variance

$$V(E_{ij}) = (0.2 \text{ mg})^2,$$

where, e.g., the variation related to the gender of the animals is included. In an earlier experiment of similar nature where only male rats were used a variance corresponding to the random error was estimated to be around $(0.15 \text{ mg})^2$.

How large do you expect the random error will be in in the experiment and model you formulated in question 11.

If the random error in question 11, $Z_{ij..}$, has the variance $V(Z_{ij..})$, the relation between this variance and the variance in the first experiment, $V(E_{ij})$, can be written

$$V(E_{ij}) = V(Z_{ij..}) + V(D_{ij}),$$

where D_{ij} is the difference between animal (i, j) and the ideal measurement result as a result of its gender.

Assess the order of magnitude of $V(D_{ij})$.

Suppose we want the same precision in the experiment set up in question 11 as was asked for in question 9 for the first experiment. How many animals would then be needed per treatment.

Question 13
In a screening experiment where the animals were separated as follows, and only 12 animals were used.

The results for these 12 animals were:

	d_1 20 mg	d_2 50 mg	d_3 80 mg
Females	0.10	0.32	0.66
	0.23	0.51	0.49
Males	0.11	0.26	0.54
	0.16	0.40	0.48

Data: Content of residual compound (mg).

Formulate the natural model for this experiment, analyse the data and estimate the parameters and the residual variance of the model.

Estimate the mean response in one of the treatments including a 95% confidence interval. Finally construct a 95% confidence interval for the residual variance and compare it with the previously (question 12) made assumptions about the uncertainty of the experiment.

Question 14
We now return to the problem discussed in question 5. The experiment was as follows

d_1	d_2	d_3
$Y_{1,1}$	$Y_{2,1}$	$Y_{3,1}$
$\vdots$	$\vdots$	$\vdots$
$Y_{1,8}$	$Y_{2,8}$	$Y_{3,8}$

and we assumed that $Y_{ij} = \mu_i + E_{ij} = \mu + \alpha_i + E_{ij}$.

Suppose, after the collection and analysis of the data, the experimenter gets the idea that the amount of residual compound in the blood may depend on or be correlated with the weight (size) of the liver of the animal in question. It could then be of interest to try to compensate for the liver weight in the analysis.

We could imagine that the weight of the liver of an animal can fx be measured rather precisely by a scanning technique or, equivalently, that the function of the liver could be measured by some other method.

The liver weight or function measure is called x_{ij} for (ij)'th animal. It is assumed that x_{ij} is not influenced by the treatment - which obviously will be true, if it is measured prior to the experiment.

How could one take x_{ij} into account in analysing the experiment. Formulate a reasonable model and indicate how it can be analysed by means of a suitable computer procedure. A variable like x_{ij} is called a covariate.

Question 15
We return to question 5, and now suppose the experimenter is not completely satisfied with

the influence of the biological variation, that seems to be present in the first experiments. In that case he could use animals from a limited number of litters and distribute them evenly over the single experiments as shown in the following design:

		Treatment		
		d_1	d_2	d_3
Litter 1	Female	y	y	y
	Male	y	y	y
Litter 2	Female	y	y	y
	Male	y	y	y
Litter 3	Female	y	y	y
	Male	y	y	y

What is the standard term for a collection of animals as defined by one litter. Formulate the proper model for this experiment and discuss the randomization of the design. Sketch the ANOVA table.

Question 16
Since the animal weight could possibly also be related to the response it could be a possibility to divide the animals i groups according to size (fx small (α), medium (β) and large (γ)). One possible design could then be:

		Treatment		
		d_1	d_2	d_3
Litter 1	Female	y (α)	y (β)	y (γ)
	Male	y (α)	y (β)	y (γ)
Litter 2	Female	y (γ)	y (α)	y (β)
	Male	y (γ)	y (α)	y (β)
Litter 3	Female	y (β)	y (γ)	y (α)
	Male	y (β)	y (γ)	y (α)

Formulate the proper model for this design, discuss the randomization and sketch the ANOVA table for the design/model.

Question 17
Alternatively one could apply the actual weight of the single animal by measuring it before the experiment is started. Using i for the treatment no., j for the litter the animal is coming from and k for the gender of the animal, the weight is called x_{ijk} as shown in the following table:

		Treatment		
		d_1	d_2	d_3
Litter 1	Female	y (x_{111})	y (x_{211})	y (x_{311})
	Male	y (x_{112})	y (x_{212})	y (x_{312})
Litter 2	Female	y (x_{121})	y (x_{221})	y (x_{321})
	Male	y (x_{122})	y (x_{222})	y (x_{322})
Litter 3	Female	y (x_{131})	y (x_{231})	y (x_{331})
	Male	y (x_{132})	y (x_{232})	y (x_{332})

Formulate the proper model for this design, discuss the randomization and sketch the ANOVA table for the design/model.

Question 18

Finally write a short conclusion that may be used by the laboratory to decide which design could be applied and which measurements could be beneficial to do supplementing the primary response.

Exercise 55

Consider a situation where 10 samples of penicillin plus a standard are to be compared by means of the so-called cylinder plate method. With this method 6 samples can be compared on one Agar plate by using 6 'wells' containing penicillin.

Before applying the penicillin to the gel by inserting the wells the gel has been inoculated with micro organisms.

The penicillin leaks out from the wells and after a certain time the radius of the circular zone around the wells where the growth of the micro organisms is reduced is measured. This measure is taken as an indicator of the concentration (or strength) of the active penicillin in the sample applied to the particular well.

The 10 samples and the standard are called, respectively,

$$A, B, C, D, E, F, G, H, I, J \quad \text{and} \quad S$$

In the experiment 6 replicates of a design using 2 plates as follows was used:

Plate 1:	S	A	B	C	D	E				
Plate 2:	S						F	G	I	H	J
Plate 3:	S	A	B	C	D	E				
Plate 4:	S						F	G	I	H	J
Plate 5:	S	A	B	C	D	E				
Plate 6:	S						F	G	I	H	J
Plate 7:	S	A	B	C	D	E				
Plate 8:	S						F	G	I	H	J
Plate 9:	S	A	B	C	D	E				
Plate 10:	S						F	G	I	H	J
Plate 11:	S	A	B	C	D	E				
Plate 12:	S						F	G	I	H	J

Thus a total of 12 plates are analyzed.

Question 1

Formulate a suitable model for this experiment.

Question 2

Compute the precision with which a sample, for example A, can be compared to the standard (as a measure of precision the variance of the estimated difference between the sample and the standard can be used).

Question 3

Compute the precision with which a sample, for example A, can be compared to another sample on the same plates, for example B. Compute the precision for the comparison between samples on different plates (A versus F for example).

Question 4

Suggest an alternative design using only 11 plates which enables us to do the above comparisons with the same precision for all pairs of samples and/or the standard (use an incomplete balanced block design with block size 6 and number of treatments equal to 11).

Question 5

Sketch the analysis of the suggested design.

Question 6

In certain situations it can be advantageous to take into account yet another source of variation such as, for example, the physical placement on the plate (if it is not symmetrical, for example). How can that be done.

Question 7

Sketch the analysis of the design suggested in question 6.

Exercise 58

We consider an experiment where the yield of a certain fermentation process is to be assessed. The process is a so called "surface" culture where the growth medium is placed in a glass container, is sterilized, and a suspension of spores of the particular fungus (in this case P. Chrysogenum) is added. The spores float to the surface where the mycelium in which the penicillin is produced is formed.

The growth medium has a complicated composition and consequently the determination of of optimal growth conditions is also complicated. A large number of factors may be of importance.

In the following 6 factors are considered. They are all thought to be of importance for the growth rate, that is the amount of penicillin produced:

A: Corn-syrup	2%-4%
B: Lactose	2%-4%
C: Precursor (starter)	0%-0.006%
D: $NaNO_3$	0%-0.3%
E: Glucose	0%-0.5%
F: Mixture of nutrients	0%-0.4%

The measurements are to conducted under controlled temperature conditions by application of an air conditioned (climatised) box in which only a limited number of samples can be placed at the same time.

Question 1
Suppose all factors are to be evaluated on p levels r times. How many individual measurements do you then have in a complete factorial design?

Write the results for $p = 2, r = 2$ and for $p = 3, r = 2$.

Question 2
Suppose now that all 6 factors are purely additive (at least almost). We then want an "optimal" design of the type p^{k-q}, that is a design where you with as few measurements as possible still can estimate the effects of all factors. Determine such a design for the case $p = 2$ and $r = 1$.

Write out the measurements that can be suggested and give expressions for the estimates for the main effects A and F and for their sums of squares.

Question 3
We again assume additive effects for the 6 factors A–F, but now all effects are to be evaluated on $p = 3$ levels. What is the the minimum number of observations required if the level (μ) and all main effects are to be estimated (remember that for all factors 3

parameters are to be determined (fx A_0, A_1 and A_2), but they sum to zero (fx $A_0+A_1+A_2 = 0$)

Sketch the construction of an experimental design which utilizes the same principles as described in connection with question 2, but now for $p = 3, r = 1$, and again for the factors A–F. How many factors could theoretically be handled in the design if additivity between factor effects can still be assumed?

Question 4

The design constructed in question 3 is to be run in a climatised box. If the box can contain at most 12 individual measurements, how could these measurements then be distributed between repeated usages of the box (or, if several boxes are at hand, between boxes) (you only need to sketch the solution to this question).

Question 5

The design from question 2 is to be extended in such a way that all measurements are carried out twice. One possibility is to repeat all measurements at the same time (doubly repeated measuring) or, alternatively, carry out the whole plan twice using single measurements each time. Explain the difference that these two possibilities make in relation to the proper mathematical model for the experiment, the randomization and the blocking. The condition that there can at most be 12 individual measurements in the climatised box at the same time is retained.

Exercise 59

The stability of a certain drug has been evaluated. The factors in question were the set points of a number and process variables applied in the production of the drug. The stability was expressed through the amount of a primary degradation product measured 12 months after production and packaging.

Seven factors were considered, and the the individual experiments were carried out on two batches of raw material (called '-1' and '+1'). It would be reasonable to treat the batches as two blocks.

Name	Meaning	Levels	
A :	Temperature of storage	18^{o}C	24^{o}C
B :	Concentration of preservative	0.0%	0.5%
C :	pH-justification of product	6.90	7.20
D :	Heat treatment of ampuls	44^{o}C	60^{o}C
E :	Additive for stabilization	0%	2%
F :	Storage condition	calm	shake a little
G :	Light	dark	day/night light

The measured response was relative content of bi-product (degradation product) given as a percentage. Design and data are displayed in the following table:

Factorial design and confounding							Batch=		
A	B	C	D	E=BCD	F=ACD	G=ABD	ABCD	Response	Code
-1	-1	-1	-1	-1	-1	-1	+1	3.78	(1)
1	-1	-1	-1	-1	1	1	-1	2.62	a fg
-1	1	-1	-1	1	-1	1	-1	3.65	b eg
1	1	-1	-1	1	1	-1	+1	3.70	ab ef
-1	-1	1	-1	1	1	-1	-1	3.62	c ef
1	-1	1	-1	1	-1	1	+1	3.35	ac eg
-1	1	1	-1	-1	1	1	+1	4.28	bc fg
1	1	1	-1	-1	-1	-1	-1	3.24	abc
-1	-1	-1	1	1	1	1	-1	2.83	d efg
1	-1	-1	1	1	-1	-1	+1	3.62	ad e
-1	1	-1	1	-1	1	-1	+1	4.79	bd f
1	1	-1	1	-1	-1	1	-1	2.69	abd g
-1	-1	1	1	-1	-1	1	+1	3.45	cd g
1	-1	1	1	-1	1	-1	-1	3.06	acd f
-1	1	1	1	1	-1	-1	-1	4.09	bcd e
1	1	1	1	1	1	1	+1	3.18	abcd efg

It is assumed that the following conceptual model is applicable:

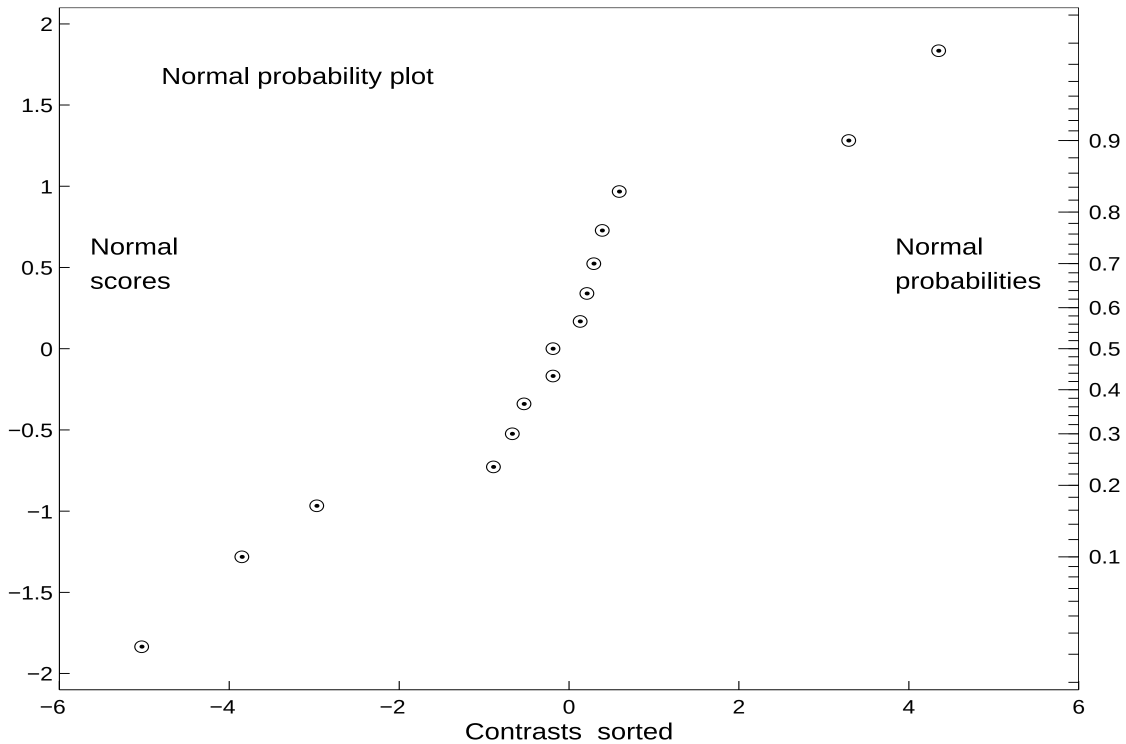
$$Y = \mu + A + B + AB + C + AC + BC + D + AD + BD + CD + E + F + G + Batch + \epsilon$$

where, for example, the influence from the factor A has two levels A_0 and A_1, ($A_0 = -A_1$), depending on whether the temperature is 18^{o}C or 24^{o}C, and correspondingly for the other factors. The contribution ϵ represents the experimental uncertainty with variance σ_ϵ^2.

The model relies on the assumption that besides the main factor effects only two- factor interactions between the factors A, B, C and D are thought to be present (different from zero).

The following computations and corresponding normal plot have been carried out:

code	Response	Yates algorithm			Contrast	SSQ
(1)	3.78	6.40	13.75	28.24	55.95	–
a fg	2.62	7.35	14.49	27.71	-5.03	1.581
b eg	3.65	6.97	13.93	-2.42	3.29	0.677
ab ef	3.70	7.52	13.78	-2.61	-2.97	0.551
c ef	3.62	6.45	-1.11	1.50	0.59	0.022
ac eg	3.35	7.48	-1.31	1.79	-0.19	0.002
bc fg	4.28	6.51	-1.31	0.44	-0.67	0.028
abc	3.24	7.27	-1.30	-3.41	0.39	0.010
d efg	2.83	-1.16	0.95	0.74	-0.53	0.018
ad e	3.62	0.05	0.55	-0.15	-0.19	0.002
bd f	4.79	-0.27	1.03	-0.20	0.29	0.005
abd g	2.69	-1.04	0.76	0.01	-3.85	0.926
cd g	3.45	0.79	1.21	-0.40	-0.89	0.050
acd f	3.06	-2.10	-0.77	-0.27	0.21	0.003
bcd e	4.09	-0.39	-2.89	-1.98	0.13	0.001
abcd efg	3.18	-0.91	-0.52	2.37	4.35	1.183



Question 1: Do an analysis with the purpose of identifying which factors and/or interactions seem to influence the stability of the product significantly, and formulate the reduced model which you think is most reasonable in view of the data obtained (in the same notation as used above).

Question 2: Estimate μ and the effects in the identified model, that is, for example, for the factor A the effect is to be understood as $A = A_1 - A_0$, i.e. the change in mean response if the factor A is altered from '18^{o}C' to '24^{o}C'.

Question 3: Estimate the uncertainty variance, σ_ϵ^2 (based on the data and your choice of final model).

Question 4: Finally, determine the optimal factor combination (when the objective is to have the smallest possible amount of bi-product) and estimate the mean response for this factor combination.

CIVILINGENIØREKSAMEN *Page* 61 *until page* 78.

Written test: May 24, 1994 *Course no.* : 04242

Name of course: **Statistical Design of Experiments**

Allowed to use: All usual textbooks, exercises and solutions, calculator, statistical tabels, etc., but only such previous examination multiple choice tests that have been handed out during the course.

This test was answered by

______________________ ______________________ ________

(name) (signature) (table no.)

Introduction

Put only one cross for each question. If more than one cross is applied the question is considered unanswered. If a cross is put by mistake it can be corrected by "blackening" it out, putting a cross in the intended box and adding: Corrected by ..... If there is doubt as to what is meant by a correction, the question is considered unanswered.

5 points is given for a correctly answered question and -1 point is given for an incorrect answer. An unanswered question or a cross in the box "don't know" gives 0 points.

The number of points required to obtain a certain mark or to pass the exam is determined during censoring of the answered tests.

The cross'es are put in the original questionaire which must be delivered back after the examination.

The original questionaire is to delivered back in all circumstances, and even if the examinee chooses not to give any answers or leaves the examination before it is finished.

Remember to fill out the present page of this questionaire with **name** and **signature**.

About the test in general: When *EMS*–values are to be applied, it is required that the method presented in the lectures and in the solutions to exercises is used (*EMS* = Expected Mean Square = expected values for S^2–values) (the method of the text book is changed to: "In each row, if any of the subscripts on the row component match the subscript in the column, write 0 if the row corresponds to a fixed effect and 1 if the row corresponds to a random effect."). When writing a general analysis of variance model a notation using parantheses indicates a hierarcical factorial structure, and distinction between deterministic and random factors is done by applying small letters for deterministic effects and capital letters for random effects. Furthermore, as an example, ϕ_t denotes $\sum_{i=1}^{a} t_i^2/(a-1)$ for a deterministic effect t_i at a levels, while σ_B^2 denotes a component of variance corresponding to a random effect B.

Problem no. I

An experiment has been conducted and the effect of the concentration of a growth stimulating compund has been tested. The experiment was performed such that a starter culture was prepared from a certain kind of micro organisms growing on some small plates. When the culture is growing satisfactorily the plates are divided in two parts and the stimulating compound is added to each half in predetermined concentrations.

In an initial experiment the following results were obtained. The data are measurements of the growth rate of the micro organisms :

	Log_{10}(Conc. mg/l)			
Plate no.	0.0	0.7	1.4	Sum
1	4.7	6.9	–	11.6
2	–	8.4	9.7	18.1
3	6.3	–	8.4	14.7
Sum	11.0	15.3	18.1	44.4

Spørgsmål 1 / I.1 : The design used can be called:

1 ☐ A Youden square

2 ☐ A symmetrically balanced incomplete block design

3 ☐ A Greaco-Latin square

4 ☐ An incomplete factorial with two factors

5 ☐ A randomised factorial with two factors

6 ☐ Don't know

Question 2 / I.2 : For the design in question it holds true that the primary objective of putting two single experiments on each plate, as shown, and opposed to just using 6 plate each with one experiment is:

1 ☐ To be able to estimate the variation between plates better

2 ☐ It is cheaper to use 3 plates than to use 6 plates

3 ☐ It is easier to randomise over 3 than over 6 plates

4 ☐ The experiment will often be much more precise with two experiments per plate

5 ☐ One wants to confound the effect from plates with the effect from the concentration

6 ☐ Don't know

From a computer program the following results were obtained:

Source of variation	Sum of squares	Degrees of freedom
Plates	10.57	2
Log_{10}(Conc)	5.14	2
Residual	0.33	1
Total	16.04	5

$$Q_{0.0} = 11.0-(11.6+14.7)/2 = -2.15$$
$$Q_{0.7} = 15.3-(11.6+18.1)/2 = +0.45$$
$$Q_{1.4} = 18.1-(18.1+14.7)/2 = +1.70$$

Question 3 / I.3 : We now want to find out whether the organisms grow linearly with Log_{10}(Conc), and in order to do this the variation between the three levels of concentration is split up in a linear and a 2nd order contribution. This gives:

1 ☐ $SSQ_{linear} = 4.94$ og $SSQ_{2nd\ order} = 0.20$, both with 1 degree of freedom

2 ☐ $SSQ_{linear} = 3.93$ og $SSQ_{2nd\ order} = 1.21$, both with 1 degree of freedom

3 ☐ $SSQ_{linear} = 2.47$ og $SSQ_{2nd\ order} = 2.67$, both with 1 degree of freedom

4 ☐ $SSQ_{linear} = 4.19$ og $SSQ_{2nd\ order} = 0.95$, both with 1 degree of freedom

5 ☐ $SSQ_{linear} = 4.84$ og $SSQ_{2nd\ order} = 0.10$, both with 1 degree of freedom

6 ☐ Don't know

When one half of a plate is to be prepared a certain amount of time is needed. During that period there may occur, despite all precautions, some growth on the other half. It is therefore considered in which sequence the two single experiments on the different plates could be performed.

Question 4 / I.4 : Which suggestion, among the ones presented below would you recommend (in that "1" denotes the first and "2" denotes the last) The data are otherwise organised as shown in the beginning of the problem.

1	2	
	2	1
1		2

1 ☐

1	2	
	2	1
2		1

2 ☐

1	2	
	1	2
2		1

3 ☐

2	1	
	1	2
2		1

4 ☐

2	1	
	1	2
1		2

5 ☐

Don't know

6 ☐

Problem no. II

We want to assess whether a number of factors influence the yield of an extraction process. The raw material which is used consists of leaves from a certain plant. The plants studied were grown at 3 randomly selected localities where they grow naturally. From each locality 12 samples were taken, as shown in the data table below.

In the laboratory 3 potential extraction compounds were tested, namely ether, ethanol and acetone, and extraction was performed at 2 different temperatures.

All other sources of variation, including the sequence in which the single meauserements were carried out, are completely randomised. The following results were obtained (yield in mg/kg):

Locality		I		II		III		sum
Ether	15^0C	62	47	67	60	46	49	331
	20^0C	76	84	80	89	69	65	463
Ethanol	15^0C	47	39	57	65	57	48	313
	20^0C	46	55	58	61	51	42	313
Acetone	15^0C	81	92	88	84	76	62	483
	20^0C	89	99	112	105	83	88	576
Sum		817		926		736		2479

The threee factors are denoted by L_i, t_j and o_k for locality, temperature and extraction compound, respectively, and distinction between fixed and random factors is made by using small or capital letters. The following mathematical model is considered to be adequate for describing the results of the experiment in that the pure experimental error is called E:

$$Y_{ijk\nu} = \mu + L_i + t_j + LT_{ij} + o_k + LO_{ik} + to_{jk} + LTO_{ijk} + E_{\nu(ijk)}$$

A computer program for doing analysis of variance has produced the following printout:

Source of variation	Sum of square	Degrees of freedom	Mean square	Expected mean squares
Locality(L)	1515.06	2	757.53	$12\sigma_L^2 + 6\sigma_{LT}^2 + 4\sigma_{LO}^2 + 2\sigma_{LTO}^2 + \sigma_E^2$
Temperature (T)	1406.25	1	1406.25	$18\phi_t + 6\sigma_{LT}^2 + 2\sigma_{LTO}^2 + \sigma_E^2$
L×T	28.50	2	14.25	$6\sigma_{LT}^2 + 2\sigma_{LTO}^2 + \sigma_E^2$
Extraction compound (O)	7942.72	2	3971.36	$12\phi_o + 4\sigma_{LO}^2 + 2\sigma_{LTO}^2 + \sigma_E^2$
L×O	284.28	4	71.07	$4\sigma_{LO}^2 + 2\sigma_{LTO}^2 + \sigma_E^2$
T×O	766.50	2	383.25	$6\phi_{to} + 2\sigma_{LTO}^2 + \sigma_E^2$
L×T×O	199.50	4	49.88	$2\sigma_{LTO}^2 + \sigma_E^2$
Residual	665.49	18	36.97	σ_E^2
Total	12808.30	35		

Question 5 / II.1 : The experimental design shown could be called

1 ☐ A mixed fixed and random effect factorial design

2 ☐ An incomplete balanced block design

3 ☐ An unbalanced block design

4 ☐ A split plot design

5 ☐ A nested factorial design

6 ☐ Don't know

Question 6 / II.2 : THe influence from the temperature, that is the effect t_j, can be estimated directly. In that "+" og "–" correspond to 20^0C and 15^0C, we find:

1 ☐ ± (1406.25/18–14.25/6)

2 ☐ ± (1406.25–14.25)16

3 ☐ ± 6.25

4 ☐ ± 1406.25/18

5 ☐ ± 1352/18 = ± 75.11

6 ☐ Don't know

Question 7 / II.3 : Testing of the interaction between temperature and extraction compound can be done directly by computing one of the following F–test quantities. Which one?

1 ☐ F(6,2) = 766.50/199.50

2 ☐ F(2,4) = 383.25/49.88

3 ☐ F(2,18) = 383.25/36.97

4 ☐ F(2,4) = (766.50/2)/(284.25/4)

5 ☐ $F(f_1,f_2)$ = (383.25–36.97)/(49.88–36.97), f_1 and f_2 are computed in a special way (approximate F–test)

6 ☐ Don't know

Question 8 / II.4 : One can (based on the given model) estimate a component of variance corresponding to the interaction between locality and extraction compound. The result is

1 ☐ σ^2_{LO} = (284.28–199.50)/(4–2)

2 ☐ σ^2_{LO} = (284.28–199.50)/(4+2)

3 ☐ σ^2_{LO} = (71.07–49.88)/(4–2)

4 ☐ σ^2_{LO} = 71.07/4

5 ☐ σ^2_{LO} = (71.07–49.88)/4

6 ☐ Don't know

The LTO–effect is tested firstly at a 5% level of significance in order, subsequently, to test the two–factor interactions LT_{ij}, LO_{ik} and to_{jk}.

Question 9 / II.5 : After the initial test of the LTO–effect an improved estimate for the pure experimental error variance can be computed. Which of the following proposals should be preferred?

1 ☐ σ^2_E = (49.88+36.97)/(2)

2 ☐ σ^2_E = (199.50+665.49)/(4+18)

3 ☐ σ^2_E = 36.97+(49.88–36.97)/(18–4)

4 ☐ σ^2_E = 49.88/4

5 ☐ $\sigma^2_E = 48.05 = 6.93^2$

6 ☐ Don't know

After testing LTO and afterwards testing the terms LT_{ij}, LO_{ik} and to_{jk} it is concluded that the following model describes the data well:

$$Y_{ijk\nu} = \mu + L_i + t_j + o_k + to_{jk} + E_{\nu(ijk)}$$

After selecting this model the sum of squares for the pure experimental error (the residual sum of squares) is computed. We find 1177.77 with 28 degrees af freedom.

Question 10 / II.6 : Further we are interested in assessing how different localities influence the yield. Which estimate do you think is the relevant one.

1 ☐ $(\hat{L}_1, \hat{L}_2, \hat{L}_3) = (\frac{817}{12} - \frac{2479}{36} \;\; , \;\; \frac{926}{12} - \frac{2479}{36} \;\; , \;\; \frac{736}{12} - \frac{2479}{36})$

2 ☐ $(\hat{\mu} + \hat{L}_1, \hat{\mu} + \hat{L}_2, \hat{\mu} + \hat{L}_3) = (\frac{817}{12}, \frac{926}{12}, \frac{736}{12})$

3 ☐ $\hat{\sigma}^2_L/12 = 1515.06$

4 ☐ $\hat{\sigma}^2_L = 12 \cdot 1515.05 + 42.06$

5 ☐ $\hat{\sigma}^2_L = (1515.06/2 - 42.06)/12$

6 ☐ Don't know

Problem no. III

In a laboratory experiment the aim is to assess the influence of alternative methods of heat treatment on the hardness of a certain surface treatment. In the experiments planned both the influence from temperature as well as the treatment time is to be studied.

The experiment is carried out using a certain thermostatised bath in which a number of test items can be placed at the same time.

In the experiment 6 items are placed in the bath at the same time and the items are again removed from the bath when the prescribed treatment time has elapsed. The treatment times for the inidividual items are found by random allocation (randomisation).

This procedure is repeated at 3 different temperatures, and this is repeated on two different days.

The following design shows the randomisation that was used, in that a total of 36 items were measured:

Experimental design 1

	60^0C		70^0C		80^0C		
5 min	12	7	3	1	13	14	
10 min	8	9	2	6	17	16	Day 1
15 min	11	10	4	5	15	18	

	60^0C		70^0C		80^0C		
5 min	21	23	35	34	25	26	
10 min	20	19	32	33	30	27	Day 2
15 min	22	24	36	31	28	29	

The items 1–6 were treated firstly at the temperature 70^0C, followed by the items 7–12 at 60^0C, etc.

Question 11 / III.1 : The variable "day" can be characterized in one of the following ways:

1 ☐ A deterministic (fixed), but nested factor

2 ☐ A split plot effect

3 ☐ A block effect

4 ☐ An incomplete block effect

5 ☐ A mixed deterministic (fixed) and random experimental error.

6 ☐ Don't know

As can be seen, a temperature level on one day at the same time constitutes a block and represents the temperature level as such.

Question 12 / III.2 : The design shown is usually called

1 ☐ A mixed fixed and random effect factorial design

2 ☐ An incomplete, balanced block design

3 ☐ A complete, balanced block design

4 ☐ A split plot design

5 ☐ An incomplete factorial design

6 ☐ Don't know

Question 13 / III.3 : Which of the following possible objectives do you think is served by repeating the design over two days as shown?

1 ☐ To be able to construct a good test of the temperature effect

2 ☐ To eliminate possible interactions between treatment time and temperature

3 ☐ To be able to test and estimate the effect from treatment time

4 ☐ To make sure that the influence from 'days' is eliminated by testing and estimating the temperature effect

5 ☐ To be able to estimate the variation between days

6 ☐ Don't know

From a general computer program the following printout corresponding to the measured values has been obtained:

Source of variation	SSQ	f	S^2
Days (D)	52.23	1	52.23
Temperature (T)	318.12	2	159.06
D × T	82.12	2	41.06
Treatment time (B)	126.56	2	63.28
D × B	19.76	2	9.88
T × B	25.04	4	6.26
D × T × B	32.32	4	8.08
Rest	45.32	18	2.52
Total	701.47	35	

Question 14 / III.4 : Initially one wants to test the influence from the temperature (T). At F–test is constructed and one finds:

1 ☐ F_T = F(2,1) = 159.06/52.23

2 ☐ F_T = F(2,2) = 159.06/41.06

3 ☐ F_T = F(2,4) = 159.06/8.08

4 ☐ F_T = F(2,18) = 159.06/2.52

5 ☐ An approximate F–test should be used:
F_T = F(2,f) = 159.06/(8.08–2.52), where the degrees of freedom "f" are computed in a special way.

6 ☐ Don't know

An engineer thinks that the above design is a little too impractical for the laboratory to carry out since there in reality can be placed 12 items in the bath at the same time. Therefore he suggests the following alternative plan where the items 1–12 (randomly selected) are treated at 80^0C, followed by the items 13–24 at 60^0C etc.:

Experimental design 2

	60^0C				70^0C				80^0C				
5 min	13	20	16	15	28	34	33	29	12	8	4	1	carried out in one day
10 min	17	21	24	23	31	32	30	25	2	6	7	3	
15 min	22	19	18	14	26	27	36	35	5	9	11	10	

Question 15 / III.5 : For this design one of the following statemants apply. Which one?

1 ☐ The design is a complete, balanced and randomised block design with 2 factors (temperature and time) and 2 blocks

2 ☐ The 2 factors time and temperature are a random and a deterministic (fixed) factor, respectively

3 ☐ The design is not an adequate design because of the randomisation suggested

4 ☐ The design is especially adequate in order to test the temperature effect

5 ☐ The design no 2 is more accurate than the design no. 1, since the block size is doubled (12 now against 6 before)

6 ☐ Don't know

Problem no. IV

An ivestigation has been carried out in order to evaluate the stability of a pharmaceutical substance.

The data lay-out is shown in the table below. The measured values are amounts of a degredation substance in the samples.

		15^0C		20^0C		25^0C	
		30 days	60 days	30 days	60 days	30 days	60 days
Batch 1	Sample 1	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2
	Sample 2	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2
	Sample 3	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2
Batch 2	Sample 1	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2
	Sample 2	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2
	Sample 3	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2
Batch 3	Sample 1	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2
	Sample 2	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2
	Sample 3	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2	y_1 y_2

The aim has been to evaluate the dependence between the stability and the storage temperature (t_i) and the number of days passed since production (d_j). In the preparation of the production the raw material is arrives in batches (B_k) and raw material from 3 batches was used in the experiment.

From each of these batches 18 samples were taken such that there were used exactly 3 samples per combination of t_i, d_j and B_k.

Finally two measurements were made (y_1 and y_2) per sample.

In order to analyse the results of the investigation the following mathematical model is formulated :

$$\begin{aligned} Y_{ijkl\nu} &= \mu + t_i + d_j + td_{ij} + B_k + TB_{ik} + DB_{jk} \\ &\quad + TDB_{ijk} + P(TDB)_{l(ijk)} + E_{\nu(ijkl)} \end{aligned}$$

where $i = (1,2,3)$, $j = (1,2)$, $k = (1,2,3)$, $l = (1,2,3)$ and $\nu = (1,2)$.

From a computer program performing analysis of variance the following printout, based on the organisation of the data shown in the table:

Source of variation	SSQ	f	S^2
T (Temp)	152.22	2	76.11
D (Døgn)	118.12	1	118.12
T × D	12.12	2	6.06
B (Batches)	26.56	2	13.28
T × B	19.76	4	4.94
D × B	12.04	2	6.02
T × D × B	13.32	4	3.33
P (Samples)	8.56	2	4.28
T × P	19.04	4	4.76
D × P	0.08	2	0.04
T × D × P	5.12	4	1.28
B × P	41.43	4	10.36
T × B × P	29.45	8	3.68
D × B × P	15.10	4	3.78
T × D × B × P	11.11	8	1.39
Rest	25.32	54	0.469
Total	509.35	107	

Question 16 / IV.1 : The sum of squares for the $P(TDB)$–effect in the model is found as:

1 ☐ SSQ = 15.10 with 4 degrees of freedom

2 ☐ SSQ = 129.89 with 36 degrees of freedom

3 ☐ SSQ = 8.56 with 2 degrees of freedom

4 ☐ SSQ = (5.12+29.45+15.10+11.11) with (4+8+4+8) degrees of freedom

5 ☐ SSQ = 121.33 with 36 degrees of freedom

6 ☐ Don't know

Question 17 / IV.2 : In order to test the effects of the model EMS–values (Expected Mean Square) are computed. For the main effect t_i is found

1 ☐ $\mathrm{E}\{\mathrm{S}_t^2\} = 36\phi_t + 12\sigma^2_{TB} + 6\sigma^2_{TDB} + 2\sigma^2_{P(TDB)} + \sigma^2_E$

2 ☐ $\mathrm{E}\{\mathrm{S}_t^2\} = 36\phi_t + 18\phi_{td} + 12\sigma^2_{TB} + 6\sigma^2_{TDB} + 2\sigma^2_{P(TDB)} + \sigma^2_E$

3 ☐ $\mathrm{E}\{\mathrm{S}_t^2\} = 36\phi_t + 2\sigma^2_{P(TDB)} + \sigma^2_E$

4 ☐ $\mathrm{E}\{\mathrm{S}_t^2\} = 36\phi_t + 6\sigma^2_{TDB} + 2\sigma^2_{P(TDB)} + \sigma^2_E$

5 ☐ $\mathrm{E}\{\mathrm{S}_t^2\} = 18\phi_t + 6\sigma^2_{TB} + 3\sigma^2_{TDB} + \sigma^2_{P(TDB)} + \sigma^2_E$

6 ☐ Don't know

Problem no. V

An experiment is to be carried out. The purpose is to evaluate how the friction (loss of energi) in a certain type of transmission (a bearing in combination with a gearing) depends on the following factors, which at the same time are organised in accordance with their expected importance:

A : Type of grease (grafite/oil)
B : Radial load on bearing (low/high)
C : Horizontal load of bearing (low/high)
D : Speed of rotation in transmission (low/high)
E : Temperaturr during operation (60^0C,90^0C)
F : Moment transmitted in transmission (low/high)

As can be seen the factors are to be assessed on only two levels.

It is assumed that the factors in question essentially act additively in relation to the measured friction and in the first experiment and using this assumption it is decided to take as few single measurements as possible .

Question 18 / V.1 : Which of the following designs would you recommend under the described circumstances?

1 ☐ A $2^{-1} \times 2^6$ faktorial in 4 blocks constructed from I_1=ABC and I_2=DEF.
2 ☐ A fractional $2^{-1} \times 2^6$ factorial design.
3 ☐ A fractional $2^{-3} \times 2^6$ factorial design.
4 ☐ A partially confounded $2^{-2} \times 2^6$ factorial design.
5 ☐ A $1/4 \times 2^6$ factorial design.
6 ☐ Don't know

Question 19 / V.2 : It is decided to carry out an experiment based on the complete factorial for the factors A, B and C and on introducing the factors D, E and F into this factorial, resulting in a design for 6 factors in 8 measurements. Which of the following generator equations do you think is best suited as basis for the design under construction?

1 ☐ D=AC, E=ABD and F=BCD

2 ☐ I_1=ABC, I_2=DEF and I_3=ABCDEF

3 ☐ D=ABC, E=ABCE and F=CDE

4 ☐ D=ABE, E=ABD, F=DE

5 ☐ None of the suggestions are adequate

6 ☐ Don't know

Question 20 / V.3 : A design based on the generator equations D=AB, E=AC and F=BC is considered. In this design the factor F will be confounded as follows:

1 ☐ F=BC=ACEF=ABDF

2 ☐ F=ABDF=ACEF=BC=BCDEF=ABE=ACD=DE

3 ☐ F=DF=ABF=EF=ACF=BC

4 ☐ F=ABDF=$ABDF^2$=ACEF=$ACEF^2$=ABDF=$ABDF^2$=BCF

5 ☐ None of the suggestions are adequate

6 ☐ Don't know

Question 21 / V.4 : The above generator equations (D=AB, E=AC and F=BC) are used to construct an experiment. Which of the following possibilities correspond to these equations?

1 ☐ $\boxed{\text{(1) ade bdf abef cef acdf bcde abc}}$

2 ☐ $\boxed{\text{(1) adf bde abef cef acde bcdf abc}}$

3 ☐ $\boxed{\text{(1) ad be abdf cf acde bdcf abce}}$

4 ☐ $\boxed{\text{(1) adef be ab ce acde bcef abcd}}$

5 ☐ $\boxed{\text{(1) abdf bded abef cdef acde bcdf abc}}$

6 ☐ Don't know

Question 22 / V.5 : The same generator equations (D=AB, E=AC and F=BC) are used to construct the following design:

$\boxed{\text{de a abdf bef acef cdf bc abcde}}$,

as one of the 8 possibilities that can be chosen. For reasons of randomisation the experiment has to be carried out in 2 blocks each containing 4 single measurements by confounding blocks with the interaction term ABC.

Which one of the following designs corresponds to this choice?

1 ☐ $\boxed{\text{a de abdf acef}}$ and $\boxed{\text{bc bef cdf abcde}}$

2 ☐ $\boxed{\text{de abdf cdf bc}}$ and $\boxed{\text{a bef acef abcde}}$

3 ☐ $\boxed{\text{de abdf acef bef}}$ and $\boxed{\text{a bc cdf abcde}}$

4 ☐ $\boxed{\text{de abdf acef bc}}$ and $\boxed{\text{a bef cdf abcde}}$

5 ☐ $\boxed{\text{a abdf acef de}}$ and $\boxed{\text{de bef cdf abcde}}$

6 ☐ Don't know

Problem no. VI

An experiment in which 4 factors are considered is to be planned. The factors are A, B, C and D, and they are all to be assessed on three levels, but in such a way that only 1/3 of the complete factorial is to be carried out. For defining the underlying complete factorial A, B and C are used.

Question 23 / VI.1 : Which of the following suggested generator equations for the design to be constructed do you think is well suited?

1 ☐ I = ABC

2 ☐ D = ABD

3 ☐ D = ABC = $ABCD^2$

4 ☐ D = ABC

5 ☐ $I_1 = AB^2C^2D^2$ and $I_2=BC^2D$

6 ☐ Don't know

As one possiblity the experimentator can choose to construct a design by introducing the factor D in such a way that the index equation $l = i + j + 2k$ (modulo 3) is used, in that i, j, k and l are the indices of the factors A, B, C and D, respectively.

A $1/3 \times 3^4$ factorial design consists of 3^3 single experiments, and the principal fraction has the following appearance -in principle:

$$\boxed{(1)\ x\ x^2\ y\ xy\ x^2y\ y^2\ xy^2\ x^2y^2\ z\ xz\ x^2z\ yz\ xyz\ ...\ x^2y^2z^2}$$

Question 24 / VI.2 : In that standard notation is used for the single experiments the whole design can be constructed by using:

1 ☐ x = ad , y = bd and z = cd^2

2 ☐ x = ad , y = bd and z = cd

3 ☐ x = bd , y = ad and z = d

4 ☐ x = abcd , y = bd and z = xy = ab^2cd^2

5 ☐ All suggestions are wrong

6 ☐ Don't know

Corresponding to the experiment described in question VI.2 the factor D is introduced into the design by the equation $D=ABC^2$, and one could further want to carry out the 27 single measurements on raw material from 3 batches. From each batch enough material is taken for conductiing 9 single experiments.

In order to distribute the single experiments on the 3 batches it is further chosen to confound the batches with the AB^2C effect.

Question 25 / VI.3 : The batch, where the single experiment "(1)" appears can - in principle - be written as

$$\boxed{(1)\ x\ x^2\ y\ xy\ x^2y\ y^2\ xy^2\ x^2y^2}$$

As x and y we can use:

1 ☐ x = ad and y = bd

2 ☐ x = ac^2d^2 and y = bc

3 ☐ x = bd and y = acd

4 ☐ x = abc^2d and y = bd^2

5 ☐ All suggestions are wrong

6 ☐ Don't know

Question 26 / VI.4 : In the above design where D = ABC^2 and batches = AB^2C, the confounding of the main effect A with other effects can be described by one of the following alias relations: Which one?

1 ☐ A = BC^2D^2

2 ☐ A = AB^2CD = BC^2D^2 = AB^2C = batches

3 ☐ A = AB^2CD

4 ☐ A = AD = AB^2C^2

5 ☐ A = AB^2CD = BC^2D^2

6 ☐ Don't know

Question 27 / VI.5 : A source of variation such as the "batches' is most often called

1 ☐ A nested factorial effect (in relation to A, B, C and D)

2 ☐ A random main effect

3 ☐ A block effekt

4 ☐ A confounded factorial effect

5 ☐ A main experimental error

6 ☐ Don't know

End of test. Have a good summer holiday.