

TMA4267 Linear Statistical Models
Part 4: Design of Experiments (DOE)
Recommended exercise 6 - V2017

March 16, 2017

Keywords: orthogonal design, effect coding, two-level designs, contrasts, 2^k full factorials, blocking, fractional factorial designs, resolution, Pareto plot, Daniel plot.

- Problem 1 focus on showing how collecting data with orthogonal (effect coding) columns of the design matrix makes the parameter estimates easy to interpret and calculate.
- A full 2^4 factorial design is analysed in Problem 2, with emphasis on performing the analysis in R using the **FrF2** library.
- Problem 3 also focus on interpretation of an orthogonal design, and ends with a fractional factorial design.
- The topic of blocking and alias structure is the focus of Problem 4 (to be done without the aid of computer software), and
- finally Problem 5 studies the fractional factorial design (by hand). Important words: generator(s), alias structure and resolution.

Problem 1: Exam V2015, Problem 2

Assume a linear regression model $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where \mathbf{Y} is an n -dimensional random vector, X an $n \times p$ design matrix, $\boldsymbol{\beta}$ a p -dimensional parameter (coefficient) vector and $\boldsymbol{\epsilon}$ n -dimensional multivariate normal with mean $\mathbf{0}$ and covariance matrix $\sigma^2 I$, where I is an $n \times n$ identity matrix.

Assume further that the columns of X are orthogonal.

a) Show that the least squares estimator of β_j , the j th entry of $\boldsymbol{\beta}$, depends only on the j th column of X (i.e., the j th covariate vector) and \mathbf{Y} .

Consider a two-level 2^2 factorial unreplicated experiment in which the levels are coded -1 and 1 . The response vector is $(6 \ 4 \ 10 \ 7)^T$, corresponding to levels $(-1 \ 1 \ -1 \ 1)^T$ of the first factor and $(-1 \ -1 \ 1 \ 1)^T$ of the second.

b) Estimate the interaction effect (twice the coefficient) of the two factors.

Problem 2: Factorial experiments

Use statistical software to solve this problem, see end of problem for hints for commands R.

We want to examine if small changes on 4 critical dimensions in a carburettor will affect the horse powers produced with a standard engine with six cylinders. The data from a 2^4 factorial experiment are given below

Dimensions				Response
<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>y</i>
-	-	-	-	14.6
+	-	-	-	24.8
-	+	-	-	12.3
+	+	-	-	20.1
-	-	+	-	13.8
+	-	+	-	22.3
-	+	+	-	12.0
+	+	+	-	20.0
-	-	-	+	16.3
+	-	-	+	23.7
-	+	-	+	13.5
+	+	-	+	19.4
-	-	+	+	11.3
+	-	+	+	23.6
-	+	+	+	11.2
+	+	+	+	21.8

- a) Fit a full 2^4 regression model. Estimate the main effects and the interactions using statistical software. Construct a Pareto-type plot and a Daniel plot (normal plot) for the estimated effects. Also make main effect and interaction (2-way) effects plots.
- b) Write down the regression model that corresponds to this analysis
- c) Why is there no result for s^2 in the software output? Assume that $\sigma^2 = 4$ is known from experience. Which effects are now significantly different from 0? Find a 95% confidence interval for the most important effects. (These last two questions you need to do by hand.)
- d) If you assume that all three-way and four-way interactions are 0, how can you then estimate σ^2 and σ^2_{effect} ? How can you now find the significant effects? Show the theory and do the analysis using statistical software.
- e) Assume now that the experiment is to be done in two blocks. Let the blocks be determined by the four-factor interaction ABCD. Which effects can now be estimated unconfounded with the block effect?

f) How would you perform the experiment in four blocks? (Try different options).

R

You first need to install the library `FrF2` from the packages tab, or writing `install.packages("FrF2")`. To load the library either write `library(FrF2)` or tick off the package at the packages tab in the lower right window of Rstudio.

Construction of design:

```
plan <- FrF2(nruns=16,nfactors=4,randomize=FALSE)
y <- c(14.6,24.8,12.3,20.1,13.8,22.3,12.0,20.0,16.3,23.7,13.5,19.4,11.3,23.6,11.2,21.8)
plan <- add.response(plan,y)
```

If we would perform real experiments we would randomize them, but to get a clearer overview (when adding the response) we skip it here. Now we have an ordinary data set up to be used with `lm`, as we know from the regression part of the course.

For parts d and e, we add the argument `blocks` to the `FrF2` function. For 4 blocks we need to allow for aliasing with two-factor interactions (to the block generator).

```
design1<-FrF2(16,4,blocks="ABCD",randomize=FALSE)
summary(design1)
design2 <-FrF2(16,4,blocks=4,alias.block.2fis=TRUE))
summary(design2)
design3 <-FrF2(16,4,blocks=c("ABC","AD"),alias.block.2fis=TRUE))
summary(design3)
```

Analysis: Now we fit linear models, but the notation to include 4th order terms is $(.)^4$ and to include up to 2nd order terms is $(.)^2$. And, remember that effects are 2*coefficients in the regression.

```
lm4 <- lm(y~(.)^4,data=plan)
summary(lm4)
anova(lm4) # to see the seqSS mentioned in the solutions to d)
effects <- lm4$coeff*2
```

For Normal plots showing the effects this called `DanielPlot`, plot of main effects is called `MEPlot` and plots of interactions are called `IAPlot`. The barplot below can be seen as a Pareto-plot.

```
DanielPlot(lm4, half=FALSE, alpha=0.05)
MEPlot(lm4)
IAPlot(lm4)
barplot(sort(abs(effects[-1])), decreasing=FALSE, las=1, horiz=TRUE)
```

Run	A	B	C	Response
1	-1	-1	-1	550
2	1	-1	-1	669
3	-1	1	-1	633
4	1	1	-1	642
5	-1	-1	1	1037
6	1	-1	1	749
7	-1	1	1	1075
8	1	1	1	729

Table 1: The etch experiment with one replicate.

Problem 3: Process development

This is an exam question from TMA4255, August 2012.

We will look at a designed experiment to develop an etching process. There are three design factors (A, B and C). The three design factors were run at two levels each.

- A: The gap between the electrodes. Low: 0.80 cm. High: 1.20 cm.
- B: The flow of gas. Low: 125. High: 200.
- C: The power applied to the cathode. Low: 275 W. High: 325 W.

The response variable is the etch rate for the process ($\text{\AA}/\text{m}$). A high value for the etch rate is desired. A 2^3 factorial design was run, and the results are presented in Table 1.

From these results we have the following design of experiment (DOE) effect estimates.

A	B	C	AB	AC	BC	ABC
-126.5	*	274.0	-42.0	-190.5	-9.5	13.0

a) Fill in the missing effect estimate for factor B. Construct a main effects plot for factor B, and explain with words how the estimated main effect of factor B is interpreted.

In addition to this first replicate of the experiment a second replicate was made, such that each factor combination was run twice, that is $n = 16$. The result from fitting a multiple linear regression model to these data, with intercept, A, B, C, AB, AC, BC and ABC as covariates, is given in the printout below. We call this the *full* regression model. Note that the estimates (“Estimate”) are the estimated regression coefficients and not the effect estimates.

	Estimate	Std. Error	<i>t</i> value	<i>p</i> -value
Intercept	776.062	11.865	65.406	3.32e-12
A	-50.812	11.865	-4.282	0.002679
B	3.688	11.865	*	0.763911
C	153.062	11.865	12.900	1.23e-06
AB	-12.437	11.865	-1.048	0.325168
AC	-76.812	11.865	-6.474	0.000193
BC	-1.062	11.865	-0.090	0.930849
ABC	2.813	11.865	0.237	0.818586
Residual standard error: 47.46 on 8 degrees of freedom				

b) What does the “Std. Error” column give, and why is the “Std. Error” the same for all covariates?

Now turn to factor B, in the second row of the printout table. What is the relationship between an estimated coefficient for B and the estimated effect for B? Calculate the missing number for the *t*-statistic. What are the hypotheses underlying the *p*-value and what is the conclusion of that test?

What are the significant covariates in the model?

c) We now assume that we use a regression model with intercept and covariates A, C and AC, and get the following printout for this new regression model. Let us call this the *reduced* regression model.

	Estimate	Std. Error	<i>t</i> value	<i>p</i> -value
Intercept	776.06	10.42	74.458	<2e-16
A	-50.81	10.42	-4.875	0.000382
C	153.06	10.42	14.685	4.95e-09
AC	-76.81	10.42	-7.370	8.62e-06
Residual standard error: 41.69 on 12 degrees of freedom				

Why are the estimated coefficients for A, C and AC in the reduced model the same as in the full model? Why have the estimated standard deviations changed from the full to the reduced model?

What would you suggest are the optimal choices of level (low or high) for each of these two factors when the aim is to use the fitted regression model to arrive at the combination with the highest estimated etch rate?

Calculate a 95% prediction interval for the etch rate based on your chosen levels for A and C.

d) We now assume that in a pilot study with three factors only runs 1, 4, 6 and 7 of Table 1 were performed.

What type of experiment is this?

What is the generator and the defining relation for the experiment?

Write down the alias structure of the experiment.

What is the resolution of the experiment?

Problem 4: Blocking

This should be solved by hand (no software).

A 2^5 experiment must be divided into 8 blocks with four observations in each block. We assume that factor A has the greatest impact on the response and we want therefore to construct a design where no main effect and none of the two-factor interactions involving factor A: AB, AC, AD and AE are confounded with the block-effects. Find out how you can block the experiment.

Problem 5: Fractional factorial

This should be solved by hand (no software).

- a) Suppose you want to run a 2^{4-1} fractional factorial experiment and have chosen $D=ABC$ as the generator for the design. What is the resolution of the design?
- b) In a 2^{8-4} experiment the generators are $E=ABC$, $F=ABD$, $G=ACD$ and $H=BCD$. What is the resolution of this design?
- c) Suppose you run the 2^{8-4} experiment in two blocks using $B1=AB$ as blocking factor. How will this affect the estimation of the effects?
- d) If you examine the results in a) and b). How many factors do you think it is possible to investigate in 32 runs and still have a resolution IV design? Can you give an argument for that it is possible to investigate at least that many factors in a 32 run resolution IV design?