TAMS38
Computer exercises 6

Preparation: *Read about different models and analysis, especially about the choice of the sample size, regression, response surface and logistic regression.*

Bring collection of formulas and tables, and calculators to the computer exercise class.

1 – Analysis using response surface

The following data set provides data for the final attempt in a experiment via the steepest ascentmethod to find the optimal values of pH and temperature for a chemical reaction.

Purity data from uniform-precision rotable Central Composite Design

	x_1	x_2	*Temp.*	*pH*	*Purity Y*
Factorial	*-1*	*-1*	*53*	*5.0*	*90.1*
portion	*+1*	*-1*	*55*	*5.0*	*91.8*
	-1	*+1*	*53*	*5.8*	*90.7*
	+1	*+1*	*55*	*5.8*	*93.6*
Centre	*0*	*0*	*54*	*5.4*	*94.1*
portion	*0*	*0*	*54*	*5.4*	*94.6*
	0	*0*	*54*	*5.4*	*94.2*
	0	*0*	*54*	*5.4*	*93.9*
	0	*0*	*54*	*5.4*	*94.0*
Axial	*-1.414*	*0*	*52.59*	*5.4*	*89.0*
portion	*+1.414*	*0*	*55.41*	*5.4*	*92.3*
	0	*-1.414*	*54*	*4.83*	*90.7*
	0	*+1.414*	*54*	*5.96*	*92.5*

Do we have curvature using analysis according to 2^2-design with five observations in the central point?

Go to **STAT/DOE/Factorial/Create Factorial Design...** *Choose amount of factors as 2. Click on* **Designs** *and choose* **Number of Center Points:** *5.*

Put in the nine y-values in c7 in the correct order by looking on the levels of factors A and B. Name column c7 as Y. Wait with the four extra points.

Go to **STAT/DOE/Factorial/Analyze Factorial Design...** *In* **Response** *set c7.*

a) Perform tests of curvature at the level 0.01. Conclusion?

Rename c5 and c6 as x1 and x2. Fill with extra points in the c5-c7. Then, create new columns by typing in the session window

```
let c8=c5*c6
let c9=c5**2
let c10=c6**2
```

*b) Rename new column as x1*x2, x1**2 and x2**2. You can now fit a quadric surface to your data. Go to* **Stat/Regression/Regression** *and fill* **Response:** *c7,* **Predictors:** *c5 c6 c8-c10.*

c) Is it possible, using this quadric surface, to find an optimum point? Calculate the values of x1 and x2 for the optimum point and recalculate them to the actual temperature and the actual pH.

d) It might be nice to watch how the fitted quadratic surface looks.

Go to **Calc/Make Mesh Data...** *Put x1-values in c11 from -1.5 to 1.5 with 21 points and x2-values similarly in c12 (not for Z).*

As Z-function you should put the estimated regression expression. Write in the session window

```
let c13=....+....*c11+....*c12+....*c11*c12 ...
```

Fill in the correct coefficients from the regression analysis and remember to put all the terms.

Go to **Graph/3D Surface Plots/Wireframe plot** *and fill in Z: c13 Y: c12 X: c11. What you have obtained?*

2 – Logistic regression, mixed model

In 1974 and 1975 two studies have been done National Opinion Research Center, University of Chicago, Illinios where people's attitude towards women is considered. Each would be asked to consider the following statement: Women should take care of running their homes and leave running the country up to men. Observations from 1305 men and 1566 women are summarized in the table below.

Amount	*Men,* $j = 1$		*Women,* $j = 2$	
year of education, k	*positive*	*negative*	*positive*	*negative*
0	*4*	*2*	*4*	*2*
1	*2*	*0*	*1*	*0*
2	*4*	*0*	*0*	*0*
3	*6*	*3*	*6*	*1*
4	*5*	*5*	*10*	*0*
5	*13*	*7*	*14*	*7*
6	*25*	*9*	*17*	*5*
7	*27*	*15*	*26*	*16*
8	*75*	*49*	*91*	*36*
9	*29*	*29*	*30*	*35*
10	*32*	*45*	*55*	*67*
11	*36*	*59*	*50*	*62*
12	*115*	*245*	*190*	*403*
13	*31*	*70*	*17*	*92*
14	*28*	*79*	*18*	*81*
15	*9*	*23*	*7*	*34*
16	*15*	*110*	*13*	*115*
17	*3*	*29*	*3*	*28*
18	*1*	*28*	*0*	*21*
19	*3*	*13*	*1*	*2*
20	*3*	*20*	*2*	*4*

The above data are available to download on the course website.

After a quick look at the data, it seems that the longer the education, the lower proportion of positive to the statement. We will now analyze the data using logistic regression and answer some questions about the length of training and gender are important.
We choose the dummy variable

$$z_j = \begin{cases} 0 & om\ j = 1\ (men) \\ 1 & om\ j = 2\ (women) \end{cases}$$

to separate two groups men and women.

Put columns y_{jk} for the positive, x_{jk} for the negative, dummy variable z_j and amount of the education years k. There is also a column for the total number $n_{jk} = y_{jk} + x_{jk}$

respondents for each sex j and years of education k. Create a column with the percentage of positive answers $\hat{p}_{jk} = y_{jk}/n_{jk}$.

a) Plot the proportion of the positive answers $\hat{p}_{jk}$ against the years of education k. Choose ***Graph/Scatterplot/With Gropus*** *so that we can see the difference between men and women.*

b) Write down model where the constant term and slope can be different for those two groups (logit $p_{jk} = ...$). Use the dummy variables to separate groups.

Do the logit-analysis for the given model with dummy variables. Go to ***Stat/Regression/Binary Log...***

c) What are the parameters of your model that are significant?

Are you satisfied with deviance? $D =. . . . . P =. . . . . .$

Conclusions?