Norwegian University of Science and Technology
Department of Mathematical Sciences

TMA4267 Linear statistical models 2020 spring
Compulsory exercises 4

Problem 1

All answers must be justified: if "no" why, if "yes" give an example.

a) Does there exist a random vector having the following matrix as its covariance matrix?

$$\Sigma_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

$$\Sigma_4 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \Sigma_5 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \Sigma_6 = \begin{bmatrix} 1 & 1/3 & 1/2 \\ 1/2 & 1 & 1/3 \\ 1/3 & 1/2 & 1 \end{bmatrix},$$

$$\Sigma_7 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \Sigma_8 = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 2 & 2 \\ 1 & 2 & 1 \end{bmatrix}, \Sigma_9 = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

b) Do there exist two random vectors having the following matrix as their covariance matrix?

$$\Sigma_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

$$\Sigma_4 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \Sigma_5 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \Sigma_6 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

Problem 2

a) Give an example of a bivariate random vector $X = (X_1, X_2)^T$ and a bivariate vector a such that both X_1 and X_2 are normally distributed but $a^T X$ is not normal. Do there exist such vectors X and a under additional condition that X_1 and X_2 are independent?

b) Does the answer to the last question change if "independent" is replaced by "uncorrelated"? Consider the following example and make analysis. Let $X_1 \sim N(0,1)$, a random variable U does not depend on X_1, and $P(U = 1) = P(U = -1) = 1/2$. Finally $X_2 = UX_1$.

c) Thus. Are the following two conditions equivalent? Condition 1: X_1 and X_2 are normal. Condition 2: (X_1, X_2) is normal.

Problem 3

The following data are given:

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	2	3	2	1	3	2	4	4	5	5	3	2	3	5	4
y_i	14	23	13	5	24	13	39	39	59	58	23	12	23	57	39

A multiple linear regression model is fitted, where the expected value of the response y is a second order polynomial in x. More precisely, the assumed model is:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$$

for $i = 1, ..., 15$, where $\epsilon_1, ..., \epsilon_{15}$ are independent and $N(0, \sigma^2)$. R output is given below.

```
Call:
lm(formula = y ~ x1 + x2)

Residuals:
    Min      1Q  Median      3Q     Max
-0.7786 -0.4026 -0.1349  0.4655  1.2074

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.16495    0.95841   3.302  0.00631 **
x1           0.72854    0.58319   1.249  0.23540
x2           2.04910    0.08128  25.210 9.21e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6168 on 12 degrees of freedom
Multiple R-squared:  0.9993,    Adjusted R-squared:  0.9992
F-statistic:  8664 on 2 and 12 DF,  p-value: < 2.2e-16
```

a) Comment briefly on the model fit. Calculate SSE, SSR and SST.

b) How can man find diagonal elements of the matrix $(X^T X)^{-1}$, using the R output, where X is the design matrix? Calculate these diagonal elements (using only the R output!).

Problem 4

A multiple linear regression model is considered. It is assumed that

$$Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i,$$

where ϵ-s are independent and have the same normal distribution with zero expectation and unknown variance σ^2. 100 measurements are made, i.e. $i = 1, 2, ..., 100$. The explanatory variables take the following values: $x_{i1} = 2$ for $1 \le i \le 25$ and 0 otherwise, $x_{i2} = \sqrt{2}$ for $26 \le i \le 75$ and 0 otherwise, $x_{i3} = 2$ for $76 \le i \le 100$ and 0 otherwise.

a) Let $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ be least-square estimators of β_1, β_2, β_3. Prove that in the considered case $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ are independent, and

$$Var(\hat{\beta}_1) = Var(\hat{\beta}_2) = Var(\hat{\beta}_3).$$

Do these properties hold in the general case? If not, give counterexamples.

b) Perform a test for

$$H_0 : \beta_1 + \beta_3 = 2\beta_2$$

vs.

$$H_1 : \beta_1 + \beta_3 \neq 2\beta_2.$$

The significance level is 0.05. The least-squares estimates of β_1, β_2 and β_3 are 0.9812 1.8851 and 3.4406, respectively. The unbiased estimate of the variance σ^2 is 3.27.

c) Three hypotheses

$$H_0 : \beta_1 = 1 \text{ vs. } H_1 : \beta_1 \neq 1,$$

$$H_0 : \beta_1 + \beta_2 = 3 \text{ vs. } H_1 : \beta_1 + \beta_2 \neq 3,$$

and

$$H_0 : \beta_1 + \beta_2 + \beta_3 = 5 \text{ vs. } H_1 : \beta_1 + \beta_2 + \beta_3 \neq 5,$$

are tested simulteneously. Probability of at least one Type I error must not be greater than 0.05. One of the following two methods can be used: the Bonferrony method and the Šidák method. Which one do you choose? Why? Which null hypotheses are rejected if the P-values are given in the table below? Why?

H_0	$\beta_1 = 1$	$\beta_1 + \beta_2 = 3$	$\beta_1 + \beta_2 + \beta_3 = 5$
P-value	0.2317	0.5134	0.0012

Problem 5

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

Factor	Factor level	
	-1	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

A fractional 2^{5-2} experiment was performed, based on a full 2^3 experiment with factors A,B,C and generators $D = AB$ and $E = AC$.

The response Y was defined as yield measured relative to a theoretical maximum. The responses $Y_1, ..., Y_8$ of the 8 experiments are assumed to be independent and normally distributed with the same variance σ^2.

The design and the 8 responses are presented in the table below

Experiment	A	B	C	D	E	Y
1	−	−	−	+	+	69.3
2	+	−	−	−	−	70.8
3	−	+	−	−	+	71.3
4	+	+	−	+	−	73.2
5	−	−	+	+	−	77.5
6	+	−	+	−	+	79.3
7	−	+	+	−	−	88.9
8	+	+	+	+	+	91.2

a) What are the defining relations and what is the resolution of the design in this Problem?

Estimate the effect of the main factor A and effects of interactions BD, CE and ABCDE. Compare these estimated effects. Comment the results.