Norwegian University of Science and Technology
Department of Mathematical Sciences

Tentative solutions to TMA4267 Linear statistical models, 19 May 2017 – English

Problem 1 Random vector

a) Find a constant matrix $\boldsymbol{C}$ such that $\boldsymbol{Y} = \boldsymbol{C}\boldsymbol{X}$.

$$\boldsymbol{C} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 1 & -1 & 0 \end{pmatrix}$$

Find $\mathrm{E}(\boldsymbol{Y})$ and $\mathrm{Cov}(\boldsymbol{Y})$.

$$\mathrm{E}(\boldsymbol{Y}) = \boldsymbol{C}\,\mathrm{E}(\boldsymbol{X}) = \boldsymbol{C}\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\mathrm{Cov}(\boldsymbol{Y}) = \boldsymbol{C}\,\mathrm{Cov}(\boldsymbol{X})\boldsymbol{C}^T = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 1 & -1 & 0 \end{pmatrix}\begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}\begin{pmatrix} \frac{1}{3} & 1 \\ \frac{1}{3} & -1 \\ \frac{1}{3} & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 1 & -1 & 0 \end{pmatrix}\begin{pmatrix} \frac{1}{3} + \frac{2}{3}\rho & 1-\rho \\ \frac{1}{3} + \frac{2}{3}\rho & -(1-\rho) \\ \frac{1}{3} + \frac{2}{3}\rho & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} + \frac{2}{3}\rho & 0 \\ 0 & 2(1-\rho) \end{pmatrix}$$

What is the distribution of $\boldsymbol{Y}$?
$\boldsymbol{Y}$ is a vector of linear combination of a multivariate normal random vector and is therefore multivariate normal, with mean and covariance matrix given above.

Are Y_1 and Y_2 independent?
Yes, since $\mathrm{Cov}(Y_1, Y_2) = 0$ and $\boldsymbol{Y}$ is multivariate normal, then Y_1 and Y_2 are independent.

1b: Focus on Σ

Σ is positive definite when $c^T \Sigma c > 0$ for all columnvectors $c \neq 0$, and this is the case when all eigenvalues of Σ are positive. We have

$\lambda_1 = 1+2\rho$, so $\lambda_1 > 0$ if $1+2\rho > 0$

$\rho > -\frac{1}{2}$

$\lambda_2 = \lambda_3 = 1-\rho$, so $\lambda_2 > 0$ if $1-\rho > 0$

$\rho < 1$

This means $\underline{\rho \in (-\frac{1}{2}, 1)}$ will give Σ positive definite

We want Σ to be positive definite, because then any linear combination of our X's will have variance >0.

$c^T X$ has $\mathrm{Cov}(c^T X) = \underset{1\times 1}{c^T \Sigma c}$ = Variance of $c^T X$

so if $c^T \Sigma c > 0$ then this is satisfied.

Let $P = [e_1 \ e_2 \ e_3]$ be the matrix with the eigenvectors of Σ as columnvectors. We have

$$Z = \begin{bmatrix} e_1^T X \\ e_2^T X \\ e_3^T X \end{bmatrix} = P^T X.$$

We also have the spectral decomposition of Σ as $\Sigma = P \Lambda P^T$, where $\Lambda = \mathrm{diag}(\lambda_i)$.

$$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \lambda_3 \end{bmatrix}$$

We know that if $\mathbf{X} \sim mvN$, then $A\mathbf{X}+b$ is also mvN, so Z is multivariate normal with

$$E(Z) = 0 \text{ since } E(\mathbf{X}) = 0$$

and $$Cov(Z) = P^T \Sigma P = \underbrace{P^T P}_{I} \Lambda \underbrace{P^T P}_{I} = \Lambda$$

So, $$\begin{bmatrix} e_1^T \mathbf{X} \\ e_2^T \mathbf{X} \\ e_3^T \mathbf{X} \end{bmatrix} \sim N_3 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1+2\rho & 0 & 0 \\ 0 & 1-\rho & 0 \\ 0 & 0 & 1-\rho \end{bmatrix} \right)$$

Finally:

$$\overbrace{e_1^T \mathbf{X} + e_2^T \mathbf{X} + e_3^T \mathbf{X}}^{W} \sim N_1\left(0, (1+2\rho)+(1-\rho)+(1-\rho)\right)$$

$$= N_1(0, 3+2\rho-\rho-\rho) = N_1(0,3)$$

$$P(W > 4) = 1 - P(W \leq 4) = 1 - \Phi\left(\frac{4-0}{\sqrt{3}}\right)$$

$$= 1 - \Phi(2.31) = 1 - 0.9896 = \underline{\underline{0.0104}}$$

Note: $Var(W) = 3$ for all choices of ρ.

I did write $\rho = 0.5$ so that it was possible to show e_1, e_2, e_3, and not confuse the student on the missing value of ρ. Hopefully noone got confused by that...

Problem 2: Modelling systolic blood pressure

a) model A: 2 ?'s

<u>BMI: Std. Error missing</u>

Std. Error is $\hat{SD}(\hat{\beta}_{BMI})$, where

$Cov(\hat{\beta}) = (\underline{X}^T\underline{X})^{-1}\sigma^2$ and

design matrix $n \times p$

$Var(\epsilon_i)$

$Var(\hat{\beta}_{BMI})$ is the corresponding diagonal element of $Cov(\hat{\beta})$, call this $c_{jj} \cdot \sigma^2$.

$$\hat{SD}(\hat{\beta}_{BMI}) = \sqrt{c_{jj}} \cdot \hat{\sigma}$$

where $\hat{\sigma}^2 = \frac{SSE}{n-p}$ and SSE is the sum-of-sq-errors

$$SSE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$
$$= Y^T(I-H)Y$$

vector of responses.

$\underline{X}(X^T\underline{X})^{-1}X^T$

Numerical value: $T_j = \frac{\hat{\beta}_j}{\hat{SD}(\hat{\beta}_j)} \Leftrightarrow \hat{SD}(\hat{\beta}_j) = \frac{\hat{\beta}_j}{T_j} = \frac{1.010550}{10.129}$

j for BMI

$$\underline{\hat{SD}(\hat{\beta}_j) = 0.0998 \approx 0.1}$$

This is the estimated standard deviation of the estimated regression coefficient.

$\Pr(>|t|)$ missing for SEX

We want to test

$$H_0: \beta_{SEX} = 0 \quad \text{vs} \quad H_1: \beta_{SEX} \neq 0$$

and use as test statistic (let j denote SEX)

$$T_j = \frac{\hat{\beta}_j - 0}{\hat{SD}(\hat{\beta}_j)}$$

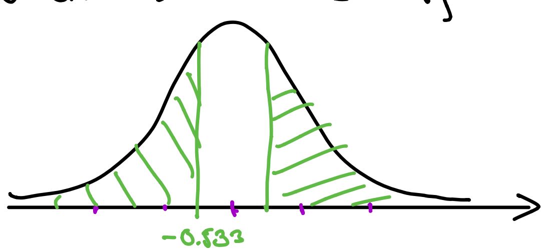
where $\hat{\beta}_j$ is the jth element of $\hat{\beta} = (X^T X)^{-1} X^T Y$ and $\hat{SD}(\hat{\beta}_j)$ as given on previous page.

The p-value of the test is "$\Pr(>|t|)$" and calculated as

$$2 \cdot P(T_j > |t_j|)$$

where t_j is the observed value of the test statistic.

When H_0 is true $T_j \sim t_{n-p}$. We have observed



$t_j = -0.533$ and
$n-p = 2593$

The t-distribution with 2593 degrees of freedom is very close to the $N(0,1)$ distribution.

$$\text{p-value} = 2 \cdot P(\underbrace{T_{2593}}_{\approx N(0,1)} > 0.533) \overset{\text{Tabeller } \& \text{ formel } N(0,1)}{\approx} 2 \cdot (1 - 0.7019) = \underline{\underline{0.5962}}$$

$P(Z \leq 0.533) = 0.7019$

This p-value is larger than any sensible choice of significance level, so we do not reject H_0 and believe SEX does not influence SYSBP in our model.

$H_0: \beta_{SEX} = 0$ vs $H_1: \beta_{SEX} \neq 0$

Model fit of model A:

- $R^2 = 25\%$ which is low, but for a medical problem we migth not be able to get much higher.
- The regression is significant, that is,

  $H_0: \beta_1 = \beta_2 = \dots = \beta_6 = 0$ vs H_1: at least one $\neq 0$

  is rejected, p-value $< 2.2 \cdot 10^{-16}$.

- Model requirements: looking at the residual plot in the left panel of Figre 2, $\hat{y}$ on x-axis, $\hat{\varepsilon}$ on y-axis

  ~linearly of covs ok?

  maybe problem with homosc. assumption $\varepsilon \sim N(0, \sigma^2 I)$

  it looks roughly random. It might be a slight downwards trend and larger variance for small values of $\hat{y}$, but that is not clear. The qq-plot (right panel) does not look like a straight line — the tails are deviating a lot — which may imply that the assumption of normality of errors is violated

  also Anderson-Darling in Fig 2. reject H_0: normal.

2b) model A : SYSBP vs covs
B : $-\frac{1}{\sqrt{SYSBP}}$ vs covs.

Residual plots: left panel of Figure 2 and 4 are not very different, maybe less of a downward trend in model B-plot than model A. The qq-plot for model B is very good - we can not reject normality of residuals.
Anderson Darling test in Fig 3
⇒ I would prefer model B.

A full model migth include variables that do not influence the response, and thereby fit noise instead of signal, thus overfitting migth be a problem. This will in particular give a bad performance for prediction because overfitting increases the variance of $\hat{\beta}_j$'s.

In best subset selection we examine all possible regression models, that is, with 6 covariates (like we have) we have $2^6 = 64$ possible ways of including or not the 6 different covariates.
We start by looking at all models of equal size, and select the best based on SSE or R^2. In the printout the best model for each of the sizes 1-6 is given. E.g. best model with one covariate include only BPMEDS, best with 2 include AGE + BPMEDS, etc.

After the best model for each size is found we need to choose between models of different sizes, and for that we can't use SSE because SSE will never decrease when new covariates are added to the model. Instead we use a penalized version by adding penalty term for including many covariates. The BIC criterion is based on adding a penalty to $-2 \cdot$ loglikelihood of the fitted model.

$$BIC = n \cdot \ln(\hat{\sigma}^2) + \ln(n)\,(|M| + 1)$$

for σ^2

penalty term

size of model = number of covariates

We choose the model with the lowest BIC.

In our case this is the model with 4 covariates, but this model is not very different from the model with 3 covariates.

Choose model with lowest BIC $\Rightarrow$ AGE + BMI + TOTCHOL + BPMEDS

2c: Testing linear hypotheses

$H_0: \beta_{sex} = \beta_{cursmoke} = \beta_{totchol} = 0$

$H_1:$ at least one $\neq 0$

$C\beta = d$:

$$3\times 7 \quad C = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\beta = \begin{array}{cc} \begin{bmatrix} \beta_0 \\ \beta_{SEX} \\ \beta_{AGE} \\ \beta_{CURSM} \\ \beta_{BMI} \\ \beta_{TOTCH} \\ \beta_{BPMED} \end{bmatrix} & \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix} \end{array}$$

$$d = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$C\beta = d \iff \begin{bmatrix} \beta_{SEX} \\ \beta_{CURS} \\ \beta_{TOTC} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Test statistic:

$$F_{obs} = \frac{\frac{1}{r}\Delta SSE}{\frac{SSE}{n-p}}$$

$r = 3$ number of Hyp

$SSE =$ in model B

$SSE_{H_0} =$ in model C

$\Delta SSE = SSE_{H_0} - SSE$

$n-p =$ in model B (= 2593 here)

Since $\hat{\sigma}^2 = \frac{1}{n-p} SSE$, we may use $\hat{\sigma}$ from model B and C to get SSE and SSE_{H_0}

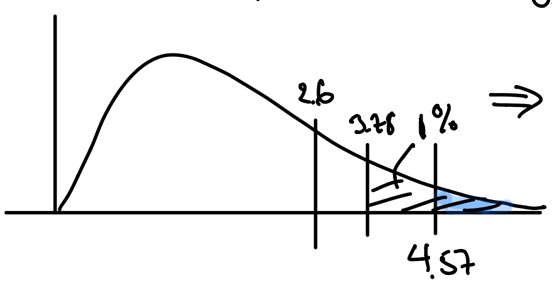
When H_0 is true $F_{obs} \sim F_{r,\, n-p}$

$$SSE = 0.005819^2 \cdot 2593 = 0.08780$$

$$SSE_{H_0} = 0.005831^2 \cdot 2596 = 0.08826$$

$$F_{obs} = \frac{\frac{SSE_{H_0} - SSE}{3}}{\underbrace{\frac{SSE}{2593}}_{\hat{\sigma}^2}} = \frac{\frac{0.08826 - 0.08780}{3}}{0.005819^2} = 4.57$$

$F_{(3,\,2593)}$ with area 0.05 to the right: 2.60
0.01: 3.78



$\Rightarrow$ F_{obs} > critical value
4.57   2.60

($\underset{\substack{\shortparallel \\ \text{p-value}}}{0.003}$ according to R)

$\Rightarrow$ reject H_0

$\Downarrow$

we prefer model B

Problem 3 Design of experiments

What type of experiment is this?

We see that we have a full factorial design in the factors A, B, C, but there is a fourth factor D added. This is a half fraction of a 2^4 design, also called a 2^{4-1}-design.

	A	B	C	D	ABC
1	-1	-1	-1	1	-1
2	1	-1	-1	-1	1
3	-1	1	-1	-1	1
4	1	1	-1	1	-1
5	-1	-1	1	-1	1
6	1	-1	1	1	-1
7	-1	1	1	1	-1
8	1	1	1	-1	1

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

The generator for the design is $D = -ABC$ (which is seen from the table above after the ABC column is added). The defining relation is then $I = -ABCD$.

What is the resolution of the experiment?

The resolution of the design equals the number of letters in the defining relation and is given using Roman numerals. Thus, the resolution is IV.

Write down the alias structure of the experiment.

Main effects and 3-factor interactions: $A = -BCD$, $B = -ACD$, $C = -ABD$, $D = -ABC$

2-factor interactions with eachother: $AB = -CD$, $AC = -BD$, $AD = -BC$

Why perform the experiments in random order?

To minimize the potential influence of external factors not part of the experimental plan.

Problem 4: Underfitting

$$Y = \underbrace{X_1}_{n\times k}\beta_1 + \underbrace{X_2}_{n\times(p-k)}\beta_2 + \varepsilon \qquad \varepsilon \sim N(0,\sigma^2 I)$$

$$Y = X_1\alpha_1 + \varepsilon^* \quad \text{underfitted model}$$

$$\hat{\alpha}_1 = (X_1^T X_1)^{-1} X_1^T Y$$

$$H_1 = X_1 (X_1^T X_1)^{-1} X_1^T$$

$$SSE_1 = Y^T (I - H_1) Y$$

i) Show that H_1 is idempotent and find the trace of H_1

$$H_1 H_1 = X_1 (X_1^T X_1)^{-1} \underbrace{\underbrace{X_1^T X_1}\,\underbrace{(X_1^T X_1)^{-1}}}_{I} X_1^T = X_1 (X_1^T X_1)^{-1} X_1^T = H_1$$

$$tr(H_1) = tr(X_1 (X_1^T X_1)^{-1} X_1^T) = tr(\underbrace{X_1^T}_{k\times n} \underbrace{X_1}_{n\times k} \underbrace{(X_1^T X_1)^{-1}}_{k\times k})$$

$tr(AB) = tr(BA)$

$$= tr(\underset{k\times k}{I}) = \underline{\underline{k}}$$

ii) Show that $SSE_1 = Y^T(I-H_1)Y$

$$SSE_1 = (Y - X_1\hat{\alpha}_1)^T(Y - X_1\hat{\alpha}_1)$$

$$Y - X_1\hat{\alpha}_1 = Y - \underbrace{X_1(X_1^TX_1)^{-1}X_1^T}_{H_1}Y = Y - H_1Y = (I-H_1)Y$$

Since H_1 is idempotent, $I-H_1$ is also idempotent.

$$SSE_1 = Y^T\underbrace{(I-H_1)^T(I-H_1)}_{I-H_1}Y = \underline{\underline{Y^T(I-H_1)Y}}$$

iii) Find $E\left(\frac{SSE_1}{n-k}\right)$ $\quad n, k, \sigma^2, \beta_2, X_1, X_2$

Hint: trace formula: $E(X^TAX) = tr(A\Sigma) + \mu^TA\mu$
$E(X) = \mu$, $Cov(X) = \Sigma$.

$$E(SSE_1) = E(Y^T(I-H_1)Y)$$

Know: $E(Y) = X_1\beta_1 + X_2\beta_2$
$Cov(Y) = \sigma^2 I$

$$E(SSE_1) = tr((I-H_1)\sigma^2 I) + (X_1\beta_1 + X_2\beta_2)^T\underbrace{(I-H_1)(X_1\beta_1 + X_2\beta_2)}$$

$$(I-H_1)X_1\beta_1 + (I-H_1)X_2\beta_2$$

$$\underbrace{X_1\beta_1 - \underbrace{H_1X_1}_{X_1}\beta_1}_{0}$$

$$= \sigma^2 \operatorname{tr}(\underbrace{I}_{n\times n} - H_1) + \underbrace{(X_1\beta_1 + X_2\beta_2)^T (I-H_1)}\ \underbrace{(I-H_1)(X_1\beta_1 + X_2\beta_2)}$$

$$\beta_2^T X_2^T (I-H_1)(I-H_1) X_2 \beta_2$$

$$= \sigma^2 (n-k) + \beta_2^T X_2^T (I-H_1) X_2 \beta_2$$

$$E\left(\frac{SSE_1}{n-k}\right) = \sigma^2 + \frac{\beta_2^T X_2^T (I-H_1) X_2 \beta_2}{n-k}$$

Problem 5: Independence of linear combinations

$X \sim N_p(\mu, \Sigma)$, $M_X(t) = \exp(\mu t + \frac{1}{2} t^T \Sigma t)$

$\underset{q \times p}{A}$ and $\underset{r \times p}{B}$

$$\underset{(q+r)}{Y} = \begin{bmatrix} AX \\ BX \end{bmatrix}.$$

i) Show that $Y \sim N_{q+r}$ using mgf.

$$M_Y(t) = E(\exp(\underbrace{t^T}_{1\times(r+q)} Y)) = E\left(\exp\left(t^T \begin{bmatrix} AX \\ BX \end{bmatrix}\right)\right)$$

$$= E\left(\exp\left(\underbrace{t^T \begin{bmatrix} A \\ B \end{bmatrix}}_{t^{*T}} X\right)\right) = E\left(\exp\left(\underbrace{t^{*T}}_{1 \times p} X\right)\right)$$

this is the mgf of X evaluated at $t^* = [A^T\ B^T]t$

$$t^{*T} = t^T \begin{bmatrix} A \\ B \end{bmatrix}$$

$$t^* = (t^{*T})^T = \left(t^T \begin{bmatrix} A \\ B \end{bmatrix}\right)$$

$$= \underbrace{[\underset{p\times q}{A^T}\ \ \underset{p \times r}{B^T}]}_{p \times (r+q)} \underset{(r+q)\times 1}{t}$$

$$M_Y(t) = \exp\left(t^{*T}\mu + \tfrac{1}{2} t^{*T} \Sigma t^{*} \right)$$

$$= \exp\left(t^T \begin{bmatrix} A \\ B \end{bmatrix} \mu + \tfrac{1}{2} t^T \begin{bmatrix} A \\ B \end{bmatrix} \Sigma \begin{bmatrix} A^T & B^T \end{bmatrix} t \right)$$

This we recognise as

$$Y \sim N_{req}\left(\begin{bmatrix} A \\ B \end{bmatrix} \mu \, , \, \begin{bmatrix} A \\ B \end{bmatrix} \Sigma \begin{bmatrix} A^T & B^T \end{bmatrix} \right)$$

ii) Condition for when $A\underline{X}$ and $B\underline{X}$ are independent.

$A\underline{X}$ and $B\underline{X}$ are independent $\iff$ all components $A\underline{X}$ are independent of all components of $B\underline{X}$

$\iff$ all covariances between these components are 0

$$\mathrm{Cov}(Y) = \begin{bmatrix} A \\ B \end{bmatrix} \Sigma \begin{bmatrix} A^T & B^T \end{bmatrix}$$

$$= \begin{bmatrix} A\Sigma A^T & A\Sigma B^T \\ B\Sigma A^T & B\Sigma B^T \end{bmatrix}$$

$A\Sigma B^T = 0$ and $B\Sigma A^T = (A\Sigma B^T)^T$

That is, $A\Sigma B^T = 0$ is a condition for when $A\underline{X}$ and $B\underline{X}$ are independent

(The condition is necessary and sufficient)