Changed 12 February 2015

TMA4267 – Linear Statistical Models
Exercise 3 – V2015

The exercises have been collected from exercise sets 3 and 4 of 2014 spring.

Problem 1: Treatment of tennis elbow

This is parts of an exam problem, so only calculations by hand are intended.

The term *tennis elbow* is used to describe a state of inflammation in the elbow, causing pain. This injury is common in people who play racquet sports, however, any activity that involves repetitive twisting of the wrist (like using a screwdriver) can lead to this condition. The condition may also be due to constant computer keyboard and mouse use.

In a randomized clinical study the aim was to compare three different methods for treatment of tennis elbow, A: physiotherapy intervention, B: corticosteroid injections and C: wait-and-see (the patients in the wait-and-see group did not get any treatment but was told to use the elbow as little as possible).

We will look at the short-term effect of treatment by studying measurements at 6 weeks. All patients participating in the study only had one affected arm. There were several outcomes measured for the study, and we will study an outcome measure called *pain-free grip force*. This was measured by a digital grip dynamometer and normalized to the grip force of the unaffected arm. A pain-free grip force of 100 would mean that the affected and the unaffected arm performed equally good. Summary statistics for each of the treatment groups are presented in the table below. Here the standard deviation is defined with the number of observations in the group in question minus 1 in the denominator (the unbiased estimator).

Treatment	Sample size	Average	Standard deviation
A (physiotherapy)	63	70.2	25.4
B (injection)	65	83.6	22.9
C (wait-and-see)	60	51.8	23.0
Total	188	69.0	

We would like to investigate if the expected pain-free grip force varies between the treatment groups. Write down the null- and alternative hypothesis and perform one hypothesis test using the summary statistics in the table above.
What are the assumptions you need to make to use this test?
What is the conclusion from the test?

Problem 2: Gene activity

This is parts of an exam problem, so only calculations by hand are intended.

We will look at the results from a biological study where the aim was to compare the gene activity of three genes (B_1, B_2, and B_3) in three different cell populations (A_1, A_2 and A_3). The gene activity was measured using the polymerase chain reaction method. Measurements are continuous and will be analysed on the logarithmic scale (base 2). In total 45 measurements were made, whig equal number of observations for each combination of gene and cell population.

A two-way analysis of variance model (ANOVA) with interaction was fitted to the data, and gave the following results.

Source	Df	Sum Sq	Mean Sq	F value	p-value
A	*	3.240	*	5.935	*
B	*	253.161	126.581	463.768	<2e-16
AB	*	*	30.645	112.277	<2e-16
Error	*	9.826	0.273		
Total	*	388.807	8.837		

Eight of the entries in the result table are each replaced with an asterisk. Explain the column headings and calculate numerical values for these missing entries. Is the interaction term AB significant? How would you now proceed for further analyses?

Problem 3: Simple matrix calculations

Solve the problems by hand AND by use of `R` (when possible).

The matrix A is given by

$$A = \begin{bmatrix} 9 & -2 \\ -2 & 6 \end{bmatrix}$$

a) Construct A as a matrix in R.
Command: `matrix`.

b) Is A symmetric?
Command: `t`.

c) Show that A is positive definite. (Not in `R`. Use the direct definition of positive definiteness.)

d) Find the eigenvalues and the eigenvectors of the matrix. A. Are the eigenvectors found by `R` normalized?
In `R` matrix multiplication is performed by the command `%*%`.
Command: `eigen`.

e) Write the spectral decomposition of the matrix A. In `R` matrix multiplication is performed by the command `%*%`.

f) Find A^{-1}.
Command: `solve`.

g) Find the eigenvalues and the eigenvectors of the matrix A^{-1}. Is there a relationship between the eigenvalues and the eigenvectors of A and A^{-1}?
Command: `solve, eigen`

h) Why can A be a covariance matrix?

i) Find the correlation matrix corresponding to A.
Command: `diag, sqrt`
Check your computations with `cov2cor`.

j) If

$$\mathrm{E}(\mathbf{X}) = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \tag{1}$$

$$\mathrm{Cov}(\mathbf{X}) = A, \tag{2}$$

and

$$B = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, \tag{3}$$

$$\mathbf{d} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \tag{4}$$

find in `R` the expectation and covariance for:

$$\mathbf{s} = B\ \mathbf{X}, \tag{5}$$

$$t = \mathbf{d}'\ \mathbf{X}, \tag{6}$$

$$\mathbf{v} = \begin{bmatrix} \mathbf{X} \\ 3\mathbf{X} \end{bmatrix}. \tag{7}$$

Problem 4: Covariance

Let $\boldsymbol{V}$ be a random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\mathrm{E}((\boldsymbol{V}-\boldsymbol{\mu})(\boldsymbol{V}-\boldsymbol{\mu})^T) = \boldsymbol{\Sigma}$. Show that $\mathrm{E}(\boldsymbol{V}\boldsymbol{V}^T) = \boldsymbol{\Sigma} + \boldsymbol{\mu}_V\boldsymbol{\mu}^T$.

Problem 5: Symmetric projection matrices

A matrix $\boldsymbol{A}$ is a projection matrix if it is idempotent, $\boldsymbol{A}^2 = \boldsymbol{A}$. An idempotent matrix $\boldsymbol{A}$ is an orthogonal projection matrix if, in the decomposition of a vector $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{y} + (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{y})$, $\boldsymbol{A}\boldsymbol{y}$ and $\boldsymbol{y} - \boldsymbol{A}\boldsymbol{y}$ are always orthogonal, that is, $(\boldsymbol{A}\boldsymbol{y})^T(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{y}) = \mathbf{0}$.

a) Prove that the eigenvalues of a projection matrix are 0 and 1.

b) Prove that the projection given by $\boldsymbol{A}$ is orthogonal if and only if $\boldsymbol{A}$ is symmetric.

c) Assume that it is known that the rank of a symmetric matrix (actually: a diagonalizable quadratic matrix) equals the number of nonero eigenvalues of the matrix. Use the result in a) together with this result to show (understand) the following:
If a $(n \times n)$ symmetric projection matrix $\boldsymbol{A}$ has rank r then r eigenvalues are 1 and $n - r$ are 0.

d) What is the relationship between the trace and rank of a symmetric projection matrix?

e) The hat-matrix related to a multiple linear regression is defined from the $(n \times p)$ design matrix $\boldsymbol{X}$. In our course the first column of the design matrix is a vector of 1s, to accomondate an intercept term in the regression parameter vector, and that the design matrix has full rank. We will also need a socalled centering matrix. We define the following:

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$$

$$\boldsymbol{J} = \mathbf{1}\mathbf{1}^T = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

Where $\boldsymbol{J}$ is a $(n \times n)$ matrix.
Show that each of the following matrices are symmetric and idempotent: $\boldsymbol{H}$, $(\boldsymbol{I} - \boldsymbol{H})$, $\frac{1}{n}\boldsymbol{J}$, $(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J})$, $(\boldsymbol{H} - \frac{1}{n}\boldsymbol{J})$. Also find the rank (or trace) of the matrices.