# Tutorial - Week 10

*Please read the related material and attempt these questions before attending your allocated tutorial. Solutions are released on Friday 4pm.*

In this tutorial, we are interested in multivariate regression with $p$-dimensional response. Using the characterization of Anderson given on page 4 of the lecture notes, we are given a random sample of $n$, $p$-dimensional vectors, $X_1, X_2, \ldots, X_n$, such that

$$X_i \sim N_p(\beta Z_i, \Sigma)$$

for each $1 \leq i \leq n$ and where $Cov(X_i, X_j) = 0$ for each $i, j$. Also, $\beta$ is $p \times q$, each $Z_i$ is of dimension $q \times 1$, $\Sigma$ is of dimension $p \times p$. Moreover, $\Sigma$ is a positive definite matrix.
We partition

$$\beta = [\beta_1 \ \beta_2]$$

for $\beta_1$ a $p \times q_1$ matrix and $\beta_2$ a $p \times q_2$ matrix where $q = q_1 + q_2$.
Of interest is the hypothesis test

$$H_0 : \beta_1 = 0 \text{ vs } H_1 : \beta_1 \neq 0$$

## Question 1

Write a routine to simulate data, ie the $X_i$, from this model, given $n, q_1, q_2$ and $p$ such that $n > p + (q_1 + q_2)$ as well as $\beta$ and $\Sigma$. Simulate each of the $Z_i$s from $N_q(\mu, \Sigma_Z)$ where again $\mu$ and $\Sigma_Z$ are assumed given and the latter is positive definite.

## Question 2

Write a routine to calculate Wilk's LRT test statistic

$$\Lambda_n = \frac{|\widehat{\Sigma}_\Omega|}{|\widehat{\Sigma}_\omega|}$$

where $\widehat{\Sigma}_\Omega$ and $\widehat{\Sigma}_\omega$ are as described on page 7 of the Lecture Notes.

## Question 3

Use your solutions to Questions 1 and 2 to examine the accuracy in the classical large $n$, fixed $p$ regime of the approximation

$$-n \log(\Lambda_n) \sim \xi^2_{pq_1}$$

under the null hypothesis.

## Question 4

Use your solutions to Questions 1 and 2 to examine the accuracy in the classical large $n$, fixed $p$ regime of the approximation

$$-k \log(\Lambda_n) \sim \xi^2_{pq_1}$$

under the null hypothesis and where $k = n - q - (p - q_1 + 1)/2$.

## Question 5

Let $y_{n_1} := p/q_1 \to y_1 \in (0,1)$, $y_{n_2} := p/(n-q) \to y_2 \in (0,1)$ and $f(x) = \log(1 + xy_{n_2}/y_{n_1})$. Now take $p \to \infty$, $q_1 \to \infty$ and $n - q \to \infty$. Recall that by the Theorem on page 14 of the lecture notes and under the null hypothesis

$$T_n := v(f)^{-1/2}[-\log(\Lambda_n) - pF_{y_1,y_2}(f) - m(f)] \to N(0,1)$$

where

$$\begin{aligned}
m(f) &:= \frac{1}{2}\log\left[\frac{(c^2-d^2)h^2}{(ch-y_2d)^2}\right] \\
v(f) &:= 2\log\left[\frac{c^2}{c^2-d^2}\right] \\
F_{y_1,y_2} &:= \frac{y_{n_2}-1}{y_{n_2}}\log(c_n) + \frac{y_{n_1}-1}{y_{n_2}}\log(c_n - d_nh_n) + \frac{y_{n_1}+y_{n_2}}{y_{n_1}y_{n_2}}\log\left[\frac{c_nh_n - d_ny_{n_2}}{h_n}\right]
\end{aligned}$$

and where

$$\begin{aligned}
h_n &:= \sqrt{y_{n_1} + y_{n_2} - y_{n_1}y_{n_2}} \\
a_n, b_n &:= \frac{(1 \mp h_n)^2}{(1-y_{n_2})^2} \\
c_n, d_n &:= \frac{1}{2}\left[\sqrt{1 + \frac{y_{n_2}}{y_{n_1}}b_n} \pm \sqrt{1 + \frac{y_{n_2}}{y_{n_1}}a_n}\right].
\end{aligned}$$

Examine this result by a simulation study.