STAT 730: Homework 4

1. **Profile analysis**. Consider the hypothesis $H_0 : \mathbf{CBM} = \mathbf{D}$ in the multivariate linear model with one-way MANOVA structure. Specifically, assume we have $k = 3$ groups and $p = 4$ measurements so $\mathbf{B} = \begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{13} & \mu_{14} \\ \mu_{21} & \mu_{22} & \mu_{23} & \mu_{24} \\ \mu_{31} & \mu_{32} & \mu_{33} & \mu_{34} \end{bmatrix}$. Note that $\mathbf{X} =$ block-diag$(\mathbf{1}_{n_1}, \dots \mathbf{1}_{n_k})$. Two common hypotheses are (i) that the group profiles are parallel, and (ii) the group profiles coincide (no difference between groups). Formulate hypothesis (i) in terms of the $kp = 12$ groups means and find the corresponding $\mathbf{C}$ & $\mathbf{M}$; $\mathbf{D} = \mathbf{0}_{2 \times 3}$. Formulate hypothesis (ii) in terms of the $kp = 12$ groups means and find the corresponding $\mathbf{C}$; $\mathbf{D} = \mathbf{0}_{2 \times 4}$. Now perform these two hypotheses on the iris data.

2. Testing $H_0 : \mathbf{R}\boldsymbol{\mu} = \mathbf{r}$ under $\mathbf{x}_1, \dots, \mathbf{x}_n \overset{iid}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

   (a) Argue that this test is the same as $H_0 : \boldsymbol{\mu}_\mathbf{y} = E(\mathbf{y}_i) = \mathbf{r}$ for the transformed data $\mathbf{y}_i = \mathbf{R}\mathbf{x}_i$. Note that then $\mathbf{Y} = \mathbf{X}\mathbf{R}'$.

   (b) For the head length/breadth data where $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})'$ are the head length/breadth of son 1 followed by the head length/breadth of son 2, test $H_0 : \mu_1 = \mu_3, \mu_2 = \mu_4$ using the approach in part (a) and Tsagris' R function. This tests that the mean breadths do not change from son 1 to son 2 and the mean lengths do not change from son 1 to son 2.

   (c) Test for bivariate normality of the $\mathbf{y}_i$ using `mardiaTest` from the MVN package, comment on the χ_2^2 Q-Q plot.

The following problems use the same data set. The file insulin.dat contains longitudinal data from a study on $n = 36$ rabbits; 12 rabbits were randomly assigned to each of 3 groups: group 1 rabbits received the standard insulin mixture, group 2 rabbits received a mixture containing 1% less protamine than the standard, and group 3 rabbits received a mixture containing 5% less protamine. Rabbits were injected with the assigned mixture at time 0, and blood sugar measurements were taken on each rabbit at the time of injection (time 0) and 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 hours post-injection. Each data record in the file insulin.dat represents a single observation; the columns of the data set are (1) rabbit number, (2) hours (time), (3) response (blood sugar level), and (4) insulin group (1, 2, or 3).

3. Obtain a profile plot for these data, indicating group number. Comment on any trends that you see.

4. Let $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3$ denote the population covariance matrices for the three groups. Obtain and report estimates $\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3$. Are they roughly similar? Test $H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3$ using Tsagris' Box's M-test function.

5. Perform a one-way MANOVA on these data across the three treatment groups; here $k = 3$ and $p = 7$. Create the MANOVA table from software output and report conclusions of $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3$. Test for normality of the $\hat{\mathbf{u}}_{ij} \in \mathbb{R}^7$ using `mardiaTest` from the MVN package, comment on the χ_7^2 Q-Q plot.

6. Contrasts via first-differences. Consider longitudinal outcomes, although this idea is very general. Take $\boldsymbol{\mu} = (\mu_1, \ldots \mu_p)' \in \mathbb{R}^p$. A first difference matrix $\mathbf{D}_p \in \mathbb{R}^{(p-1)\times p}$ gives how our mean function $\mu(t)$ changes from $t-1$ to t, i.e. $\Delta\mu(t) = \mu(t) - \mu(t-1)$, at the values $t = 2, \ldots, p$:

$$\mathbf{D}_p\boldsymbol{\mu} = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \begin{bmatrix} \mu_2 - \mu_1 \\ \mu_3 - \mu_2 \\ \vdots \\ \mu_p - \mu_{p-1} \end{bmatrix}.$$

This is the "discrete 1st derivative." If it is $\mathbf{0}$ then the means are constant $\mu_1 = \cdots = \mu_p$ and the response does not change across time.

If the means all lie on a straight line, then the difference of these 1st differences, i.e. the "discrete 2nd derivative" is equal to zero:

$$\mathbf{D}_{p-1}\mathbf{D}_p\boldsymbol{\mu} = \begin{bmatrix} \mu_3 - 2\mu_2 + \mu_1 \\ \mu_4 - 2\mu_3 + \mu_2 \\ \vdots \\ \mu_p - 2\mu_{p-1} + \mu_{p-2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

What is this matrix $\mathbf{D}_{p-1}\mathbf{D}_p$ for the rabbit data where $p = 7$?

If the means lie on a parabola, i.e. are quadratic (note constant functions and lines are also quadratic), then the difference of these previous differences, the "discrete 3rd derivative" is equal to zero:

$$\mathbf{D}_{p-2}\mathbf{D}_{p-1}\mathbf{D}_p\boldsymbol{\mu} = \begin{bmatrix} \mu_4 - 3\mu_3 + 3\mu_2 - \mu_1 \\ \mu_5 - 3\mu_4 + 3\mu_3 - \mu_2 \\ \vdots \\ \mu_p - 3\mu_{p-1} + 3\mu_{p-2} - \mu_{p-3} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

What is this matrix $\mathbf{D}_{p-2}\mathbf{D}_{p-1}\mathbf{D}_p$ for the rabbit data where $p = 7$? Test that the trajectories are (i) at most linear, and (ii) at most quadratic within each group using a `lm` MANOVA fit without an intercept. Are any of the treatment group means significantly nonlinear in time at the $\alpha = 0.1$ level?

Hint: You will need to use the transpose of the difference matrices as $\mathbf{M}$ in H_0 : $\mathbf{CBM} = \mathbf{0}$ for both sets of tests. The $\mathbf{C}$ matrix will be 1×3 simply picking out the group.

7. Fit the general linear model to these data assuming quadratic trajectories for each group *that have the same intercept*; in R you might use the `gls` function from `nlme`. Theoretically, there should be no mean differences among rabbits at baseline because the treatment hasn't started working yet. Try unstructured, AR(1), and compound symmetry $\mathbf{\Sigma}$, with and without different variances at each time t_j. Find the model with the smallest AIC. Within this model, test that the linear and quadratic terms are the same across groups. The model is

$$y_{ij} = \beta_0 + (\beta_1 + \beta_2 I\{g_i = 2\} + \beta_3 I\{g_i = 3\})t_j + (\beta_4 + \beta_5 I\{g_i = 2\} + \beta_6 I\{g_i = 3\})t_j^2 + u_{ij},$$

where $\mathbf{u}_i \overset{iid}{\sim} N_7(\mathbf{0}, \mathbf{\Sigma})$.

8. Fit a mixed model where *each rabbit* has its own trajectory, i.e.

$$y_{ij} = (\beta_0 + \gamma_{i0}) + (\beta_1 + \beta_2 I\{g_i = 2\} + \beta_3 I\{g_i = 3\} + \gamma_{i1})t_j + (\beta_4 + \beta_5 I\{g_i = 2\} + \beta_6 I\{g_i = 3\} + \gamma_{i2})t_j^2 + u_{ij},$$

where $\mathbf{u}_i \overset{iid}{\sim} N_7(\mathbf{0}, \sigma^2 \mathbf{I}_7)$ and $\boldsymbol{\gamma}_i = (\gamma_{i0}, \gamma_{i1}, \gamma_{i2})' \overset{iid}{\sim} N_3(\mathbf{0}, \mathbf{\Omega})$. In R you might use the `lme` function from `nlme`. Compare AIC from this model to that obtained in the last problem.