Additional Multivariate Exercise 12

Data and univariate analyses: The Egyptian skulls data were described in the Manly textbook (Example 1.2). One-way ANOVAs for each of the 4 variables readily reproduces the F-values and P-values listed in the first paragraphs of Example 4.3.

MANOVA tests: Manual calculation of the four common MANOVA tests is demonstrated in Manly's Example 4.3. The results for Wilk's lambda, Lawley-Hotelling trace and Pillai's trace statistics agree completely with those obtained in Minitab and Stata, both in the values of the test statistics and in the associated approximations by F-distributions. Also the calculated value of Roy's root statistic is confirmed by the software, but Minitab does not give an F-approximation (or P-value) for this statistic, and Stata uses a somewhat different value for df_2 (145 instead of 140) in its conversion to an F-distribution. All four statistics are very strongly significant ($P < 0.001$), which is unsurprising when several of the univariate tests were strongly significant as well.

Estimates and tests for variance: Both Minitab and Stata will output the error SSCP (sum of squares and cross-products) matrix, which when divided by $n - k = 150 - 5 = 145$ gives the estimated error covariance matrix. The error estimated variances for each of the four variables are the same as in the univariate ANOVAs. Minitab additionally prints the correlations for the error SSCP matrix, which equal the estimated correlations between errors. This matrix can be calculated from the error SSCP matrix in Stata as well.

To test equal variance (matrices) between the 5 periods, the assumption underlying the MANOVA and the pooled error matrix, one can use Box's M-test. The Manly text also discusses tests that generalize univariate Levene's tests, but the interested student is referred to the R code supplied with the book for those. Box's M-test generalizes the univariate Bartlett test for equal variances. (We note that these tests are non-significant for all 4 variables, and the recommended rule $s_{\max}/s_{\min} \leq 2$ is easily met as well for all 4 variables). For the multivariate test, Stata gives the same associated F-value (1.14) as in the textbook, also with the same degrees of freedom, and additionally a χ^2-approximation valid when the second degrees of freedom is large (it certainly is here). Box's M-test does not appear to be available in Minitab.

Further results: Stata's `mvreg` command allows to display the mean estimates in the familiar regression format, but both estimates and standard errors are exactly the same as in the univariate analyses. The same is true for the estimates obtained from the `margins` command. For diagnostics, Stata only offers raw residuals, whereas Minitab offers the entire suite of regression diagnostics. All of these values are however exactly the same as in the univariate models. Stata allows pairwise comparisons, but these are again the same as in the univariate models. However, both the `lincom` and `test` commands allow to estimate and test expressions involving multiple equations.

Stata also has a `manovatest` command, where (multiple) contrasts can be set up using matrices to test hypotheses across multiple variables. This feature is particularly helpful for MANOVA modelling of repeated measures; we won't go into details here.

Quantitative modelling of time: Once we have created a predictor with the time values for the five periods, it is straightforward to fit models with a quantitative effect of time. The time points are far from equidistant, but a linear relation still seems to work reasonably well for all variables. Indeed, quadratic and cubic terms are far from significant by the multivariate tests. In the linear model, the estimated slopes for time are positive for maximum breadth and nasal height, and negative for the other two variables. The interpretations are the same as in the univariate models; for example, a 1000-year change in time corresponds to change in maximum breadth of 1.10 units (*mm*). The significance for the linear term is moderate for basibregmatic height ($P = 0.026$) and nasal height ($P = 0.038$), and very strong ($P < 0.0005$) for the other two variables.