STAT636 - Homework 3

Daniel Osorio - dcosorioh@tamu.edu
Department of Veterinary Integrative Biosciences
Texas A&M University

1. Consider the Auto data. These represent a two-factor experiment on cars. The two factors are (i) the number of cylinders (4 and 6 were considered) and (ii) origin (three origins were considered). We have 4 cars under each of the $2 \times 3 = 6$ factor combinations. For each car, we have measurements of three weight variables: X_1 = displacement, X_2 = horsepower, and X_3 = acceleration. So, in terms of a two-way MANOVA model, $g = 2$, $b = 3$, and $n = 4$.

   (a) Test for a location effect, a variety effect, and a location-variety interaction at $\alpha = 0.05$. Do this using the manova function in R. Overall, what do you conclude about these data?

   (b) Construct the two-way MANOVA table by computing SSPFAC 1, SSPFAC 2, SSPINT, SSPRES, and SSPCOR. Provide R code that matches the Wilks' statistics computed by manova. Note that your p-values (computed according to the notes) will not match those of manova, because the distributional results we have learned for two-way MANOVA are large-sample approximations. That said, how do your p-values compare to those of manova?

2. Conduct a simulation study to investigate the coverage probabilities of different confidence interval types with multivariate data. Let the sample size be $n = 30$, the number of variables $p = 4$, the number of simulations $B = 10000$, and

$$\mu' = (0, 0, 0, 0, 0)$$

   and

$$\Sigma = \begin{pmatrix} 1.0 & 0.6 & 0.6 & 0.6 & 0.6 \\ 0.6 & 1.0 & 0.6 & 0.6 & 0.6 \\ 0.6 & 0.6 & 1.0 & 0.6 & 0.6 \\ 0.6 & 0.6 & 0.6 & 1.0 & 0.6 \\ 0.6 & 0.6 & 0.6 & 0.6 & 1.0 \end{pmatrix}$$

   For each of B times, simulate a dataset of size n from the $N_p(\mu, \Sigma)$ distribution, compute 95% confidence intervals of types one-at-a-time, T^2 simultaneous, and Bonferroni simultaneous, and record whether each interval contains its corresponding population mean component value. Report a 3×5 matrix of estimated coverage probabilities. The rows of your matrix should correspond to the 3 different interval types, and the columns should correspond to the p mean components; be sure to clearly indicate which row goes with which interval type. Comment on the performance of the different interval types.