Math 585 HW 9 Spring 2024 Due Friday, April 12
Quiz 9 on Wed. April 10, covers HW9. 2 pages, problems A)-F)

For the following R problems perform the perform the *source("J:/mpack.txt")* and *source("J:/mrobdata.txt")* commands as described in homework 3. Also copy and paste commands from (http://parker.ad.siu.edu/Olive/mrsashw.txt) for the relevant problem into R.

A), 8.1 Assume the cases in each of the G groups are iid from a population with covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{x}}(j)$ Find $E(\boldsymbol{S}_{pool})$ assuming that the k groups have the same covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{x}}(j) \equiv \boldsymbol{\Sigma}_{\boldsymbol{x}}$ for $j = 1, ..., G$.

B), 5.10 Tests for covariance matrices are very nonrobust to nonnormality. Let a plot of x versus y have x on the horizontal axis and y on the vertical axis. A good diagnostic is to use the DD plot. So a diagnostic for $H_0 : \boldsymbol{\Sigma}_{\boldsymbol{x}} = \boldsymbol{\Sigma}_0$ is to plot $D_i(\overline{\boldsymbol{x}}, \boldsymbol{S})$ versus $D_i(\overline{\boldsymbol{x}}, \boldsymbol{\Sigma}_0)$ for $i = 1, ..., n$. If $n > 10p$ and H_0 is true, then the plotted points in the DD plot should cluster tightly about the identity line.

a) A test for sphericity is a test of $H_0 : \boldsymbol{\Sigma}_{\boldsymbol{x}} = d\boldsymbol{I}_p$ for some unknown constant $d > 0$. Make a "DD plot" of $D_i^2(\overline{\boldsymbol{x}}, \boldsymbol{S})$ versus $D_i^2(\overline{\boldsymbol{x}}, \boldsymbol{I}_p)$. If $n > 10p$ and H_0 is true, then the plotted points in the "DD plot" should cluster tightly about the line through the origin with slope d. Use the R commands for this part and paste the plot into *Word*. The simulated data set has $\boldsymbol{x}_i \sim N_{10}(\boldsymbol{0}, 100\boldsymbol{I}_{10})$ where $n = 100$ and $p = 10$. Do the plotted points follow a line through the origin with slope 100?

b) Now suppose there are k samples, and want to test $H_0 : \boldsymbol{\Sigma}_{\boldsymbol{x}_1} = \cdots = \boldsymbol{\Sigma}_{\boldsymbol{x}_k}$, that is, all k populations have the same covariance matrix. As a diagnostic, make a DD plot of $D_i(\overline{\boldsymbol{x}}_j, \boldsymbol{S}_j)$ versus $D_i(\overline{\boldsymbol{x}}_j, \boldsymbol{S}_{pool})$ for $j = 1, ..., k$ and $i = 1, ..., n_i$. If each $n_i > 10p$ and H_0 is true, what line will the plotted points cluster about in each of the k DD plots?

C), 9.3a) Suppose $S_2^2 = S_{22} = 126.05$, $\overline{x}_2 = 54.69$, $n = 87$, and $p = 3$. Find a large sample simultaneous 95% CI for μ_2.

D), 9.3b) Suppose a random sample of 50 bars of soap from method 1 and a random sample of 50 bars of soap from method 2 are obtained. Let X_1 = lather and X_2 = mildness with $\overline{\boldsymbol{x}}_1 = (8.4, 4.1)^T$ and $\overline{\boldsymbol{x}}_2 = (10.2, 3.9)^T$. Test $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ if $T_0^2 = 52.4722$.

E), 10.1 In the MANOVA model, $\hat{\boldsymbol{\beta}}_i = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}_i$, and $\boldsymbol{Y}_i = \boldsymbol{X}\boldsymbol{\beta}_i + \boldsymbol{e}_i$. Treating $\boldsymbol{X}\boldsymbol{\beta}_i$ as a constant, $\text{Cov}(\boldsymbol{Y}_i, \boldsymbol{Y}_j) = \text{Cov}(\boldsymbol{e}_i, \boldsymbol{e}_j) = \sigma_{ij}\boldsymbol{I}_n$. Using this information, show $\text{Cov}(\hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\beta}}_j) = \sigma_{ij}(\boldsymbol{X}^T\boldsymbol{X})^{-1}$.

F), 10.3 The Johnson and Wichern (1988, p. 262) turtle data gives the length, width and height of painted turtle shells. There is a sample of 24 female and a sample of 24 male turtles.

a) The *R* command for this part make the response and residual plots for each of the three variables. Click the rightmost mouse button and highlight *Stop* to advance the plot. When you have the response and residual plots for one variable on the screen, copy and paste the two plots into *Word*. Do this three times, once for each variable. The male turtles are smaller than the female turtles.

b) The *R* command for this plot makes a DD plot of the residuals and adds the lines corresponding to the three prediction regions of Section 5.2. The robust cutoff is larger than the semiparametric cutoff. Place the plot in *Word*. Do the residuals appear to follow a multivariate normal distribution?