Exercise #26  Mardia’s test of multivariate normality

Consider an $n \times p$ data matrix $\mathbf{Y}$ of a random sample of size $n$ from a $p$-variate population. Then the $i$-th observation, represented by a $p \times 1$ vector $\mathbf{y}_i$, is the transpose of the $i$-th row of $\mathbf{Y}$. Define

$$ \hat{\beta}_1 = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij}^3, $$

$$ \hat{\beta}_2 = \frac{1}{n} \sum_{i=1}^{n} g_{ii}^2, \text{ where} $$

$$ g_{ij} = (\mathbf{y}_i - \bar{\mathbf{y}})\mathbf{S}_n^{-1}(\mathbf{y}_j - \bar{\mathbf{y}}) \text{ and} $$

$$ \mathbf{S}_n = \frac{1}{n} \mathbf{Y}' \left( \mathbf{I}_n - \frac{1}{n} \mathbf{J}_n \right) \mathbf{Y}. $$

Note that $\hat{\beta}_1$ and $\hat{\beta}_2$ are respectively, the multivariate sample skewness and kurtosis. Mardia shows, for large sample ($n - p$ is large), $\kappa_1 = n \hat{\beta}_1/6$ follows a chisquare distribution with $p(p + 1)(p + 2)/6$ degrees of freedom, and $\kappa_2 = \left\{ \hat{\beta}_2 - p(p + 2) \right\} / \left\{ 8p(p + 2)/n \right\}^{1/2}$ follows a normal distribution. Hence these two quantities can be used to test null hypothesis of multivariate normality. Write an IML module to implement this and find a test data set of your own to test on it.