STAT 575 – Assignment 1 – due date is on course outline

When submitting R scripts and output, please include only the relevant parts, which you believe should be read. I will be quite unhappy if you merely recycle my own programs, from the course web site, with only the data sets changed.

1. Suppose that $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{A}_{q\times p}$ and $\mathbf{b}_{q\times 1}$ are a matrix and vector with constant elements.

   (a) Show that the affine transformation $\mathbf{Ax} + \mathbf{b} \sim N_q\left(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T\right)$.

   (b) Show that, if $q = p$ and $\mathbf{A}$ is nonsingular, and if the transformation in (a) is applied to each member of a sample $\mathbf{x}_1, ..., \mathbf{x}_n$, then the resulting T^2-statistic for testing $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ is unchanged.

2. Show that jointly normally distributed random vectors (not merely random variables) are independently distributed if and only if they are uncorrelated.

3. (a) Recall that a leading principal minor of a square matrix is the determinant of the submatrix formed from the first i rows and columns. Suppose that a matrix $\mathbf{M}_{n\times n}$ is such that *all* leading principal minors are non-zero. Thus if $\mathbf{M}$ is partitioned as

   $$\mathbf{M} = \begin{pmatrix} \mathbf{A}_{m\times m} & \mathbf{B} \\ \mathbf{C} & \mathbf{D}_{(n-m)\times(n-m)} \end{pmatrix},$$

   (for any $m \le n$), then $|\mathbf{A}| \neq 0$. In particular $|\mathbf{M}| \neq 0$. Show that we can write $\mathbf{M} = \mathbf{LU}$, where $\mathbf{L}$ and $\mathbf{U}$ are nonsingular, with $\mathbf{L}$ lower triangular and $\mathbf{U}$ upper triangular (both necessarily nonsingular). This is known as the LU-decomposition. (Hint: Verify that

   $$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \end{pmatrix} \begin{pmatrix} \mathbf{I}_m & \mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{I}_{n-m} \end{pmatrix},$$

   then use induction on n.)

   (b) Suppose that $\mathbf{M}$ is symmetric and positive definite. Show that the assumptions of (a) hold, and that there is then a decomposition (the 'Cholesky decomposition') $\mathbf{M} = \mathbf{U}^T\mathbf{U}$, with $\mathbf{U}$ upper triangular.

4. Suppose that $X_1, ..., X_m$ are independent, with $X_i \sim N(\nu_i, 1)$, so that $X^2 = \sum_{i=1}^m X_i^2 \sim \chi_n^2\left(\lambda^2\right)$, with $\lambda^2 = \sum_{i=1}^m \nu_i^2$.

   (a) Show that we can assume that $X_1 \sim N(\lambda, 1)$ and that $X_2, ..., X_m \sim N(0, 1)$. [Hint: Let $\mathbf{x}_{m\times 1}$ have elements X_i, and write $X^2 = \|\mathbf{x}\|^2 = \|\mathbf{Qx}\|^2$ for any orthogonal $\mathbf{Q}$. Choose $\mathbf{Q}$ to have first row $\boldsymbol{\nu}^T / \|\boldsymbol{\nu}\|$.]

... over

(b) Use (a) to show that $X^2 \sim X_1^2 + X_{m-1}^2$, where $X_1^2 \sim \chi_1^2(\lambda^2)$ independently of $X_{m-1}^2 \sim \chi_{m-1}^2$ (central).

(c) Show that X^2 is 'stochastically increasing in λ', in that the function $P(X^2 > c)$ is an increasing function of λ, for any $c > 0$. [Hint: Show that X_1^2 has this property, and then condition on X_{m-1}^2.]
Note: The same conditioning approach applies to a singly non-central F_n^m r.v. – it too is stochastically increasing in its ncp, implying that the power of the LR test of $H_0 : \lambda^2 = 0$, i.e. that all $\nu_i = 0$, increases as one moves away from the null hypothesis.

5. Let $\mathbf{A}$ and $\mathbf{B}$ be $n \times n$ symmetric matrices, with $\mathbf{B} > \mathbf{0}$.

(a) Show that

$$\max_{\mathbf{x}} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{x}} = ch_{\max}\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2},$$

and that if this maximum characteristic root is denoted by $\lambda_{\max}$, then the maximum is attained by any non-zero vector $\mathbf{x}_0$ satisfying $(\mathbf{A} - \lambda_{\max}\mathbf{B})\,\mathbf{x}_0 = \mathbf{0}$. [Note: A result which is of some interest here, and very useful elsewhere, is that if $\mathbf{P}$ is $m \times n$ and $\mathbf{Q}$ is $n \times m$, with $m \leq n$, then the eigenvalues of $\mathbf{QP}$ are those of $\mathbf{PQ}$ together with $n - m$ zeros. In this current problem, $m = n$ and so $\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}$, $\mathbf{B}^{-1}\mathbf{A}$ and $\mathbf{A}\mathbf{B}^{-1}$ all have the same eigenvalues, hence the same maximum eigenvalue.]

(b) Using (a) – or otherwise (e.g. the Cauchy-Schwarz Inequality) show that, as in Lecture 5, $\left|\frac{\sqrt{n}\left(\mathbf{a}^T(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)\right)}{\sqrt{\mathbf{a}^T\boldsymbol{S}\mathbf{a}}}\right|$ is maximized by $\mathbf{a} = \mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)$.

6. The following is a partial description of a statistical model relating fetal weight (in grams) to gestational age (weeks following conception). For a subject sampled at random from a population of pregnant women, let w_1 and w_2 denote the fetal weights determined from ultrasound examinations at gestational ages a_1 and a_2. The model is based on a transformation to a z-score $z_j = f(w_j, a_j)$ for $j = 1, 2$. The model implies the following assumptions for $a_1 = 20$ and $a_2 = 28$. Each z-score is the sum of a 'true' score and a measurement error: $z_j = t_j + u_j$. The joint distribution of (t_1, t_2, u_1, u_2) is multivariate normal with mean $(0, 0, 0, 0)$. For the 'true' scores, we have $\text{VAR}[t_1] = .70$, $\text{VAR}[t_2] = .84$ and $\text{COV}[t_1, t_2] = .67$. For the errors – which are uncorrelated with the true scores – we have $\text{VAR}[u_1] = .30$, $\text{VAR}[u_2] = .16$ and $\text{COV}[u_1, u_2] = 0$.

(a) What is the joint (trivariate) distribution of (z_1, z_2, t_2)?

... over

(b) What is the conditional distribution of t_2 given $z_1 = -2.0$?

(c) Suppose you obtain the z-scores z_1 and z_2 from ultrasounds taken at 20 and 28 weeks. How would you use these values to predict the unknown 'true' score t_2 at 28 weeks gestation? Is the z_1 value useful here, or is the best prediction based on z_2 alone? Provide a rigorous justification for your answer.

7. A wildlife ecologist measured X_1 = tail length (in mm.) and X_2 = wing length (in mm.) for a sample of $n = 45$ female hook-billed kites. The data are in T5-12.dat on the course web site. For instance the first bird had (tail length, wing length) = $(191, 284)$. Using R:

   (a) Plot a 95% confidence ellipse for the population mean vector. Suppose it is known that *male* hook-billed kites have mean tail and wing lengths of 190 mm. and 275 mm. Are these plausible values for the mean tail and wing lengths of the female birds? Why or why not?

   (b) Formulate and test the hypothesis implied by (a). But don't restrict to a particular value of α – report the p-value instead.

   (c) Construct simultaneous 95% confidence intervals for the two means, and the corresponding Bonferonni intervals. Compare them with each other and with the ellipse in (a).

   (d) Is the bivariate normal distribution a viable population model? Justify your answer with appropriate plots. When testing the marginal normality, use the R function `shapiro.test()` to get the p-values.

   (e) Using appropriate large sample approximations (explain what they are), give a 95% confidence interval on the *ratio* of the mean lengths (i.e. tail/wing).