STAT 575 Assignment 1

Due Date: January 28, 2019 by 14:00 in class

There are seven equally weighted questions. Please have a cover page for your assignment and put only your name on it. Write down your name and your ID at the header of every single page of your assignment except for the cover page. Send me your R code in a zip file named by stat575_Assign1_FirstName_LastName.zip. In the zip file, you need have one individual Assign1_Q#.R for each question that needs R code. In the subject line of the email, write "stat 575 HW1 R code".

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 three definitions in lecture are equivalent. In the class, we have shown that definition (i) $\Leftrightarrow$ definition (iii). Now you only need to prove that definition (ii) is equivalent to either definition (i) or definition (iii).

   (b) Show that the affine transformation $\mathbf{Ax}+\mathbf{b} \sim N_q(\mathbf{A}\boldsymbol{\mu}+\mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$ in at least three ways.

   (c) Show that, if $q = p$ and $\mathbf{A}$ is *nonsingular*, and if the transformation in (b) 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. (a) Suppose that $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. If we have partition,

   $$\mathbf{x} = \begin{pmatrix} \mathbf{x}_{(1)} : p_1 \times 1 \\ \mathbf{x}_{(2)} : p_2 \times 1 \end{pmatrix}, \ \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{(1)} \\ \boldsymbol{\mu}_{(2)} \end{pmatrix}, \ \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix};$$

   then

   $$\mathbf{x}_{(1)}|\mathbf{x}_{(2)} \sim N_{p_1}\left(\boldsymbol{\mu}_{(1)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left(\mathbf{x}_{(2)} - \boldsymbol{\mu}_{(2)}\right), \boldsymbol{\Sigma}_{11\cdot 2}\right),$$

   where $\boldsymbol{\Sigma}_{11\cdot 2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$. In the lass, we have prove this by construction method. Can we use conditional density method to prove it? Add necessary condition, and prove it by conditional density method.

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

3. Suppose that $X_1, ..., X_m$ are independent, with $X_i \sim N(\nu_i, 1)$, so that $X^2 = \sum_{i=1}^m X_i^2 \sim \chi_n^2(\lambda^2)$, 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 $\boldsymbol{\nu}^T/\|\boldsymbol{\nu}\|$.]

   (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).

4. 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 $(\mathbf{A}-\lambda_{\max}\mathbf{B})\mathbf{x}_0 = \mathbf{0}$.

(b) 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 $m \le n$, then the eigenvalues of $\mathbf{QP}$ are those of $\mathbf{PQ}$ together with $n-m$ zeros. Prove the statement. In (a), $m = n$ 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.

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

5. 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 $z_j = f\left(w_j, a_j\right)$ 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)?

(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.

6. (a) Suppose that Suppose that $\mathbf{x} \sim N_p\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$, where $p = 5$, $\boldsymbol{\mu} = (1,2,3,4,5)$ and $\boldsymbol{\Sigma} = (\sigma_{ij}) = (\rho^{|i-j|})$ for $i, j = 1,\ldots,5$ with $\rho = 0.8$. Find the conditional distribution of $Y = X_1|X_2\cdots X_5$ using R.

(b) Generate 50000 observations following the distribution in (a) (please do you report all the observations, R would be enough). Using the the generated sample, check if X_1 follows normal distribution by normal QQ plot and if $\mathbf{x}$follows multivariate normal distribution by chi-square QQ plot. Please do report the two plots.

(c) Using the the generated sample in (b), fit linear regression $X_1 \sim 1+X_2+\cdots+X_5$ compare the results with the conditional mean in (a).

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).