Exam 1, 22S:161, Fall 2009

NAME: SOLUTIONS

This exam has 6 questions; some have several parts. You may use your "formula sheet," a table of percentage points of the t-distribution, a table of upper 5% points of the F distribution, and a calculator. All work should be done on the examination paper. Use backs of facing pages if you need extra room. Show your work as clearly and neatly as possible to receive partial credit. You have 90 minutes for this exam. Good luck!

Points awarded:

1. /14
2. /15
3. /15
4. /14
5. /14
6. /28

---

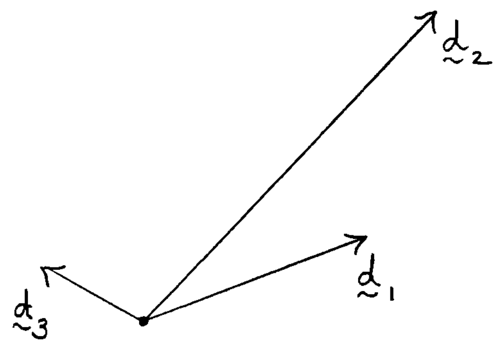
1. (14 pts.) Suppose that a random sample of size $n = 2$ is taken from a trivariate distribution. One geometric interpretation of the sample regards the columns of the 2×3 data matrix $\mathbf{X}$ as vectors in $\mathcal{R}^2$. That is,

$$\mathbf{X} = [\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3]$$

where each $\mathbf{y}_i$ is 2×1. The residual vectors are defined as

$$\mathbf{d}_i = \mathbf{y}_i - \bar{x}_i \mathbf{1} \quad (i = 1, 2, 3)$$

where $\bar{x}_i$ is the sample mean of the ith variable. Suppose that $\mathbf{d}_1$, $\mathbf{d}_2$, and $\mathbf{d}_3$ are as depicted in the following diagram:



Based on the information in this diagram, arrange the quantities in each of the following two lists in order, from smallest to largest.

(a) s_{11}, s_{22}, s_{33}

Since the squared lengths of the residual vectors are proportional to the sample variances, we have

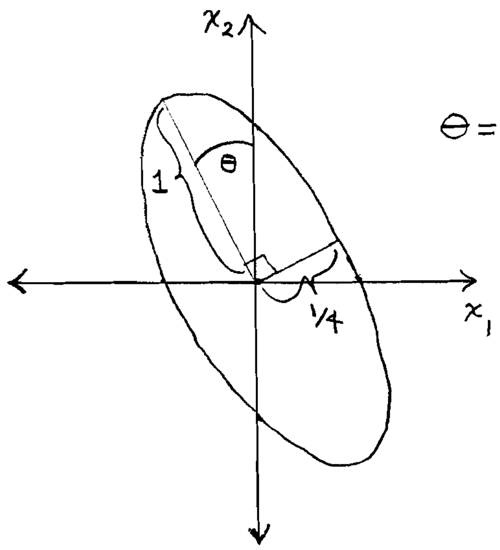
$$S_{33} < S_{11} < S_{22} .$$

(b) r_{12}, r_{13}, r_{23}

Since the cosine of the angle between two residual vectors is equal to the sample correlation coefficient between the two corresponding variables, we have

$$r_{13} < r_{23} < r_{12} .$$

2. (15 pts.) The following diagram depicts an ellipse in $\mathcal{R}^2$. The points on the ellipse satisfy the equation $\mathbf{x}'\mathbf{A}\mathbf{x} = 1$. Using the information given in the diagram, determine $\mathbf{A}$. (Hint: Use the spectral decomposition.)



$\theta = 30°$

Since $\theta = 30°$, we obtain $\underset{\sim}{e}_1 = \begin{pmatrix} -1/2 \\ \sqrt{3}/2 \end{pmatrix}$ and $\underset{\sim}{e}_2 = \begin{pmatrix} \sqrt{3}/2 \\ 1/2 \end{pmatrix}$.

Also, $\frac{1}{\sqrt{\lambda_1}} = 1 \Rightarrow \lambda_1 = 1$, $\frac{1}{\sqrt{\lambda_2}} = \frac{1}{4} \Rightarrow \lambda_2 = 16$.

By the spectral decomposition,

$$\underset{\sim}{A} = \lambda_1 \underset{\sim}{e}_1 \underset{\sim}{e}_1' + \lambda_2 \underset{\sim}{e}_2 \underset{\sim}{e}_2'$$

$$= 1 \begin{pmatrix} 1/4 & -\sqrt{3}/4 \\ -\sqrt{3}/4 & 3/4 \end{pmatrix} + 16 \begin{pmatrix} 3/4 & \sqrt{3}/4 \\ \sqrt{3}/4 & 1/4 \end{pmatrix}$$

$$= \begin{pmatrix} 49/4 & \frac{15\sqrt{3}}{4} \\ \frac{15\sqrt{3}}{4} & 19/4 \end{pmatrix}.$$

3. (15 pts.) The following partial information is given:

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \text{where} \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & 2 & 1 \\ 2 & \sigma_{22} & \sigma_{23} \\ 1 & \sigma_{32} & \sigma_{33} \end{bmatrix}.$$

Use the following 4 additional pieces of information to completely determine $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

(a) The distribution of $\begin{bmatrix} X_2 \\ X_3 \end{bmatrix}$ is bivariate normal with mean $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$ and covariance matrix $\begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix}$.

(b) The conditional variance of X_1, given $X_2 = 3$ and $X_3 = 0$, is $\frac{13}{7}$.

(c) $E(X_1^2) = 4$.

(d) $\mu_1 < 0$.

$$\text{var}(X_1 \mid X_2 = 3, X_3 = 0) = \frac{13}{7}$$

$$\Rightarrow \sigma_{11} - (2, 1)\begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}^{-1}\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \frac{13}{7}$$

$$\Rightarrow \sigma_{11} = \frac{13}{7} + (2, 1) \cdot \frac{1}{7}\begin{pmatrix} 2 & -1 \\ -1 & 4 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \frac{13}{7} + \frac{1}{7}(3, 2)\begin{pmatrix} 2 \\ 1 \end{pmatrix} = 3$$

Thus, (a) and (b) imply that $\underset{\sim}{\Sigma} = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 4 & 1 \\ 1 & 1 & 2 \end{pmatrix}$.

Next, $E(X_1^2) = 4$

$$\Rightarrow \text{var}(X_1) + \mu_1^2 = 4$$

$$\Rightarrow 3 + \mu_1^2 = 4$$

$$\Rightarrow \mu_1^2 = 1$$

Since $\mu_1 < 0$ according to (d), $\mu_1 = -1$.

Thus, $\underset{\sim}{\mu} = \begin{pmatrix} -1 \\ 3 \\ 4 \end{pmatrix}$.

4. (14 pts.) Suppose that

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \text{where} \quad \boldsymbol{\mu} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 4 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

(a) Determine the distribution of $\begin{bmatrix} X_1 + X_2 \\ X_2 + X_3 \end{bmatrix}$. $\left(= \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \underset{\sim}{X} \right)$

$N_2(\underset{\sim}{\mu}^*, \Sigma^*)$, where

$$\underset{\sim}{\mu}^* = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix},$$

$$\underset{\sim}{\Sigma}^* = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} 4 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 5 & 4 & 2 \\ 2 & 4 & 3 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 9 & 6 \\ 6 & 7 \end{pmatrix}.$$

(b) Determine a function of X_1 and X_3 that has a chi-square distribution with two degrees of freedom.

$$\begin{pmatrix} X_1 \\ X_3 \end{pmatrix} \sim N_2\left(\begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}\right),$$

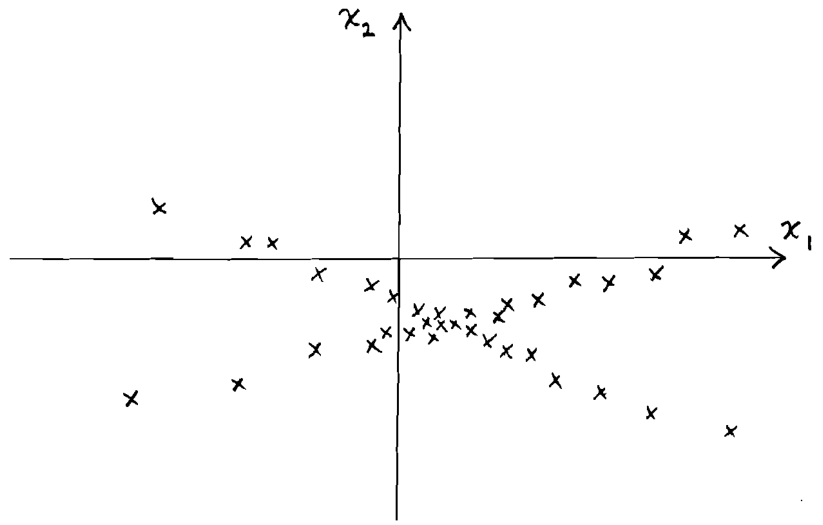
So

$$(X_1+1,\ X_3-1)\begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}^{-1}\begin{pmatrix} X_1+1 \\ X_3-1 \end{pmatrix} \sim \chi^2_2, \quad \text{i.e.}$$

$$\frac{1}{7}(X_1+1,\ X_3-1)\begin{pmatrix} 2 & -1 \\ -1 & 4 \end{pmatrix}\begin{pmatrix} X_1+1 \\ X_3-1 \end{pmatrix} \sim \chi^2_2, \quad \text{i.e.}$$

$$\frac{2}{7}(X_1+1)^2 + \frac{4}{7}(X_3-1)^2 - \frac{2}{7}(X_1+1)(X_3-1) \sim \chi^2_2.$$

5·~~4~~. (14 pts.) A random sample of size 35 was taken from a bivariate distribution. The bivariate scatter diagram of these data is displayed below:



(a) On the basis of this diagram, is the hypothesis that the distribution is bivariate normal plausible? Explain.

No. The scatter diagram is not even close to being elliptical in shape.

(b) Would you have reach the same conclusion you reached in part (a) if you had examined histograms of only the two marginal distributions. Explain.

No. Each of the two marginal histograms would be approximately bell-shaped, as is seen by mentally projecting all the points in the scatter diagram to either the x_1-axis or x_2-axis.

6 ~~8~~. (28 pts.) Suppose that a random sample of size 16 was taken from a $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population, resulting in the following sample mean and covariance matrix:

$$\bar{\mathbf{X}} = \begin{bmatrix} 10 \\ 11 \\ 13 \end{bmatrix}, \qquad \mathbf{S} = \begin{bmatrix} 4 & 1 & 3 \\ 1 & 9 & 2 \\ 3 & 2 & 4 \end{bmatrix}.$$

(a) Test $H_0 : \boldsymbol{\mu} = \begin{bmatrix} 11 \\ 11 \\ 11 \end{bmatrix}$ versus $H_1 : \boldsymbol{\mu} \neq \begin{bmatrix} 11 \\ 11 \\ 11 \end{bmatrix}$ at the .05 level of significance. You may use the following result:

$$\mathbf{S}^{-1} = \begin{bmatrix} .58 & .04 & -.45 \\ .04 & .13 & -.09 \\ -.45 & -.09 & .64 \end{bmatrix}$$

$$T^2 = n(\bar{\underset{\sim}{X}} - \mu_0)' \underset{\sim}{S}^{-1}(\bar{\underset{\sim}{X}} - \mu_0)$$

$$= 16 [-1, 0, 2] \begin{bmatrix} .58 & .04 & -.45 \\ .04 & .13 & -.09 \\ -.45 & -.09 & .64 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$$

$$= 16 [-1.48, -.22, 1.73] \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$$

$$= 79.04$$

$$\frac{(n-1)p}{n-p} F_{.05, p, n-p} = \frac{(15)(3)}{13} F_{.05, 3, 13} = \frac{45}{13}(3.41) = 11.80.$$

Since $79.04 > 11.80$, we reject H_0.

/.

(b) Obtain Bonferroni-based 94% simultaneous confidence intervals for the component means and all pairwise difference of the component means. (That is, use the Bonferroni approach to obtain intervals for the component means and their pairwise differences, such that the simultaneous coverage probability of the intervals is at least 94%.)

We need intervals for the following 6 quantities:

$\mu_1, \mu_2, \mu_3, \mu_1-\mu_2, \mu_1-\mu_3, \mu_2-\mu_3$

Therefore, the t-value we need is $t_{.06/(2\cdot 6), 15} = 2.947$.

The desired confidence intervals are:

$$\left.\begin{array}{l} \bar{X}_1 \pm 2.947\sqrt{4/16} \\ \bar{X}_2 \pm 2.947\sqrt{9/16} \\ \bar{X}_3 \pm 2.947\sqrt{4/16} \\ (\bar{X}_1-\bar{X}_2) \pm 2.947\sqrt{\frac{4+9-2}{16}} \\ (\bar{X}_1-\bar{X}_3) \pm 2.947\sqrt{\frac{4+4-6}{16}} \\ (\bar{X}_2-\bar{X}_3) \pm 2.947\sqrt{\frac{9+4-4}{16}} \end{array}\right\} = \begin{array}{l} 10 \pm 1.47 \\ 11 \pm 2.21 \\ 13 \pm 1.47 \\ -1 \pm 2.44 \\ -3 \pm 1.04 \\ -2 \pm 2.21 \end{array}$$