國立政治大學100學年度博士班資格考考試命題紙

考 試 科 目	線性模式專題	卷　　別	第一卷	考 試 日 期	100 年 9 月 5 日 星期一

1. Let D be an $n \times n$ symmetric matrix. Suppose $\lambda_1, \ldots, \lambda_n$ are real numbers and $v_1, \ldots v_n$ are orthogonal vectors in $\mathcal{R}^n$. Denote by P_{v_i} the projection matrix onto the space spanned by v_i ($\mathcal{L}(v_i)$), $i = 1, \ldots, n$.

   (a) (12pts). Please show that $D = \Sigma_{i=1}^n \lambda_i P_{v_i}$ if and only if D has eigenvalues $\lambda_1, \ldots, \lambda_n$ with corresponding eigenvectors $v_1, \ldots v_n$ respectively.

   (b) (8pts). Suppose D is nonnegative definite. Let Z be a random vector in R^n having a standard multivariate normal distribution ($N_n(0, I_n)$). Please find an $n \times n$ matrix A in terms of $\lambda_1, \ldots, \lambda_n$ and $P_{v_1}, \ldots, P_{v_n}$ such that the random vector $Y = AZ$ has a multivariate normal distribution with mean vector 0 and covariance matrix D ($N_n(0, D)$).

   (c) (12pts). Please find the distribution of the quadratic form $Y'QY$, where $Q = \Sigma_{\{i: i=1,\ldots,n \text{ and } \lambda_i \neq 0\}} \frac{1}{\lambda_i} P_{v_i}$.

2. Let Y be a random vector in R^n. Suppose we wish to perform the following model:

$$Y = X_1\beta_1 + X_2\beta_2 + \epsilon, \qquad (\circledcirc)$$

   where X_1 and X_2 are constant $n \times p$ and $n \times q$ matrices, β_1 and β_2 are unknown constant p-vector and q-vector, and ϵ is a random n-vector. Suppose $X = (X_1, X_2)$ has full column rank (then so are X_1 and X_2).

   (a) (10pts). Consider the following equivalent model:

$$Y = X_1\beta_1^* + X_2^\perp\beta_2^* + \epsilon, \qquad (\circledast)$$

   where $X_2^\perp = (I - P_{X_1})X_2$, $P_{X_1} = X_1(X_1'X_1)^{-1}X_1'$ is the projection matrix onto $C(X_1)$, the column space of X_1, and β_1^* and β_2^* are unknown constant p-vector and q-vector. Please express β_1 and β_2 in terms of β_1^*, β_2^*, X_1, and X_2.

   (b) (16pts). Please derive that the least squares estimates of β_2 and β_1 are

$$\hat{\beta}_2 = (X_2^{\perp\prime}X_2^\perp)^{-1}X_2^{\perp\prime}Y^\perp \ \text{ (regression of } Y^\perp \text{ on } X_2^\perp \text{)}$$

   and

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'(Y - X_2\hat{\beta}_2) \ \text{ (regression of } Y - X_2\hat{\beta}_2 \text{ on } X_1 \text{),}$$

   where $Y^\perp = (I - P_{X_1})Y$. (Hint. Derive the least squares estimates of β_2^* and β_1^* first.)

   (c) (16pts). Consider the one-way model with one covariate: $y_{ij} = \alpha_i + \gamma x_{ij} + \epsilon_{ij}$, $i = 1, \ldots, k$, $j = 1, \ldots, m$. Using the idea of (b), find the least squares estimates of α_i, $i = 1, \ldots, k$, and γ (details are not required).

國立政治大學100學年度博士班資格考考試命題紙

考 試 科 目	線性模式專題	卷　別	第一卷	考 試 日 期	100 年 9 月 5 日 星期一

3. Suppose Y, a random vector in $\mathcal{R}^n$, and X, a constant $n \times p$ matrix, are decomposed into $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$ and $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, where Y_1 and Y_2 are random vectors in $\mathcal{R}^{n_1}$ and $\mathcal{R}^{n_2}$ respectively, X_1 and X_2 are constant $n_1 \times p$ and $n_2 \times p$ matrix respectively, and $n = n_1 + n_2$. Suppose X_1 has full column rank. Then it can be shown (no need to prove here) that X has also full column rank and $I_{n_2} + X_2(X_1'X_1)^{-1}X_2'$ is positive definite. Denote $\hat{\beta}$ the best linear unbiased estimate (BLUE) of β for regression of Y_1 on X_1; Namely, $\hat{\beta} = X_1(X_1'X_1)^{-1}X_1'Y_1$. Furthermore, it can be verified (no need to prove here) that

$$\hat{\beta}^* = \hat{\beta} + (X_1'X_1)^{-1}X_2'(I_{n_2} + X_2(X_1'X_1)^{-1}X_2')^{-1}(Y_2 - X_2\hat{\beta})$$

satisfying the normal equation $(X'X)\hat{\beta}^* = X'Y$. Hence, $\hat{\beta}^*$ is the BLUE of β for regression of Y on X.

(a) (10pts). Let $\hat{Y} = X\hat{\beta} = \begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_2 \end{pmatrix}$ and $\hat{Y}^* = X\hat{\beta}^* = \begin{pmatrix} \hat{Y}_1^* \\ \hat{Y}_2^* \end{pmatrix}$. Please show that $||Y - \hat{Y}||^2 = ||Y - \hat{Y}^*||^2 + ||\hat{Y}^* - \hat{Y}||^2$.

(b) (16pts). Denote SSE and SSE* the error sums of squares for regression of Y_1 on X_1 and Y on X respectively. Please show that

$$\text{SSE}^* = \text{SSE} + (Y_2 - X_2\hat{\beta})'(I_{n_2} + X_2(X_1'X_1)^{-1}X_2')^{-1}(Y_2 - X_2\hat{\beta}).$$

(Hint. Use (a) and note that $||Y - \hat{Y}||^2 = SSE + ||Y_2 - \hat{Y}_2||^2$.)

國立政治大學 100 學年度博士班資格考考試命題紙

考 試 科 目	線性模式專題	卷 別	第二卷	考 試 日 期	100 年 9 月 5 日 星期一

1. (30%) Consider the linear regression model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \epsilon$ subject to $\boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{0}$.

   (1) Show the estimator of β is $\hat{\boldsymbol{\beta}}_c = \hat{\boldsymbol{\beta}} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{C}^T[\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{C}^T]^{-1}\boldsymbol{C}\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$.

   (2) Find $E(\hat{\boldsymbol{\beta}}_c)$.

   (3) Find $Var(\hat{\boldsymbol{\beta}}_c)$.

2. (25%) Consider a groupwise heterosedastic regression model

$$Y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \epsilon_i, \quad i = 1, \cdots, n,$$

   and $E(\epsilon_i|\boldsymbol{x}_i) = 0$. The n observations are grouped into G groups, each with n_g observations. The slope vector is the same in all group, but within group g:

$$Var(\epsilon_{jg}|\boldsymbol{x}_{jg}) = \sigma_g^2, \quad j = 1, \cdots, n_g, \ g = 1, \cdots, G.$$

   If the variances are known, please find the generalized least squares estimator, denoted by $\hat{\boldsymbol{\beta}}$, for $\boldsymbol{\beta}$ and $Var(\hat{\boldsymbol{\beta}})$.

3. (25%) Let $\boldsymbol{y}$ be $MN_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 I)$, where $\boldsymbol{X}$ is $n \times p$ of rank $k < p \leq n$, and let C be $m \times p$ of rank $m < k$ such that $\boldsymbol{C}\boldsymbol{\beta}$ is a set of m linearly independent estimable functions, and $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$. Please show the test statistic for $H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{0}$.

4. (20%) Let $Y_{ijk} = \mu_{ij} + \epsilon_{ijk}$, where

$$\begin{aligned} \mu_{ij} &= \bar{\mu}_{..} + (\bar{\mu}_{i.} - \bar{\mu}_{..}) + (\bar{\mu}_{ij} - \bar{\mu}_{i.}) \\ &= \mu + \alpha_i + \beta_{ij}, \end{aligned}$$

   where $\epsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2)$, $i = 1, \cdots, I$, $j = 1, \cdots, J$, and $k = 1, \cdots, K$. Please obtain the test statistics for testing hypotheses $H_0 : \beta_{ij} = 0$ for all i, j.

國立政治大學 100 學年度博士班資格考考試命題紙

考試科目	線性模式專題	卷別	第三卷	考試日期	100 年 9 月 5 日 星期一

1.(12%) Assume that $\mathbf{X}=\begin{bmatrix}X_1\\X_2\\X_3\end{bmatrix} \sim N\left(\begin{bmatrix}-1\\1\\2\end{bmatrix}, \begin{bmatrix}4&3&-3\\3&5&-2\\-3&-2&8\end{bmatrix}\right)$.

a(6%) Obtain the best linear predictor $\hat{X}_1$ of X_1 in terms of X_2 and X_3.

b(6%) Calculate $\mathrm{var}(X_1 - \hat{X}_1)$.

2.(10%) Assume that $\mathbf{X}=\begin{bmatrix}X_1\\X_2\\X_3\end{bmatrix} \sim N\left(\begin{bmatrix}-1\\4\\0\end{bmatrix}, \begin{bmatrix}6&-3&0\\-3&6&3\\0&3&9\end{bmatrix}\right)$, and that $Q_1(\mathbf{X})$ =

$(X_1-2X_2)^2+(X_2-3X_3)^2+(2X_1-X_3)^2$, and $Q_2(\mathbf{X}) = 2(X_1-\bar{X})^2+(X_2-\bar{X})^2+3(X_3-\bar{X})^2$. Compute $E[Q_1(\mathbf{X})]$ and $E[Q_2(\mathbf{X})]$.

3.(12%) Suppose the following pairs (X_{ijk}, Y_{ijk}) are recorded for i =1, 2, j =1, 2, and k =1, 2.

	$\mathbf{B}_1$	$\mathbf{B}_2$
$\mathbf{A}_1$	(1,6)	(2,14)
	(5,11)	(8,23)
$\mathbf{A}_2$	(8,18)	(11,19)
	(6,17)	(15,28)

Complete a two-way analysis of covariance according to the following model.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma X_{ijk} + \varepsilon_{ijk}.$$

The analysis should include a number of F-tests and conclusions.

國立政治大學100學年度博士班資格考考試命題紙

考試科目	線性模式專題	卷別	第三卷	考試日期	100年9月5日 星期一

4.(12%) The effect of anesthetics on dogs are studied using a design described as follows. Five dogs were initially given the drug pentobarbital to put to sleep. Each dog was then administered carbon dioxide (CO_2) at each of three different pressure levels (low, medium, and high). The response, time between heartbeats of each dog, were recorded and shown in the table below.

Dog	**CO_2 pressure level** low	**med**	**high**
1	**278**	**442**	**449**
2	**359**	**470**	**455**
3	**334**	**414**	**489**
4	**312**	**436**	**427**
5	**375**	**423**	**503**

Use the appropriate method to test to see whether there is a significant effect on the heartbeat rate of the dogs due to different pressure levels of CO_2. State your conclusion.

5.(18%) Suppose the model $\mathbf{y}=\sum_{j=1}^{3}\beta_j x_j+\varepsilon$ is to be fitted to $\mathbf{x}_1=[1, 1, 1, 1, 1]'$, $\mathbf{x}_2=[1, 1, 1, 0, 0]'$, $\mathbf{x}_3=[1, 0, 0, 1, 1]'$, and $\mathbf{y}=[-15, 11, 19, -3, -2]'$.

a(6%) Compute the estimates for β_j, j =1, 2, 3.

b(6%) Compute $\mathbf{x}_j^{\perp}$, j =1, 2, 3; where $x_j^{\perp}=x_j-P(x_j\mid x_1,\cdots,x_{j-1})$, which is the part of $\mathbf{x_j}$ orthogonal to the column space of $\mathbf{x}_1,\ldots,\mathbf{x}_{j-1}$.

c(6%) Use the alternative model $\mathbf{y}=\sum_{j=1}^{3}\beta_j^{*}x_j^{\perp}+\varepsilon$ to estimate β_j^{*}, and thus β_j. Compare the results with what you got in part a.

(Matrices for you to use: $X'X=\begin{bmatrix}5&3&3\\3&3&1\\3&1&3\end{bmatrix}$, $(X'X)^{-1}=\frac{1}{2}\begin{bmatrix}4&-3&-3\\-3&3&2\\-3&2&3\end{bmatrix}$.)

6.(18%) Suppose $\Omega=R_4$, $\mathbf{x}_1=[1, 1, 1, 1]'$, $\mathbf{x}_2=[0, 1, 0, 1]'$, $\mathbf{x}_3=[1, 0, -1, 0]'$, $\mathbf{x}_4=[0, 3, 1, 3]'$, $V=L(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)$ be a supspace of Ω, and $\eta=\beta_1+\beta_2+2\beta_3+2\beta_4$.

a(6%) Verify that η is estimable. Compute the vector **a** such that $\hat{\eta}=(\mathbf{a}, \mathbf{y})$ is the BLUE of η

b(6%) Give a vector **d** different from the vector **a** such that $\eta^{*}=(\mathbf{d}, \mathbf{y})$ be another linear unbiased estimator of η Show that $p(\mathbf{d}|V)=\mathbf{a}$, and $var(\eta^{*})-var(\hat{\eta})>0$.

c(6%) Now it is given that $\mathbf{y}=[4, 6, 5, 8]'$, compute $\hat{\eta}$ and η^{*}.

考 試 科 目	線性模式專題	卷　　別	第三卷	考 試 日 期	100 年 9 月 5 日 星期一

7.(18%) We are modeling yields of blueberry of two different conditions with several plots within each condition. Let $\mathbf{y} = \begin{bmatrix} y_{11} & y_{21} \\ y_{12} & y_{22} \\ y_{13} & y_{23} \\ y_{14} & \end{bmatrix}$, in which the entry y_{ij} denotes the yield of blueberry of the ith condition on the jth plot, where j=1, 2,…, n_i, i=1, 2, and $n_1 = 4$, $n_2 = 3$. Furthermore, let $\mathbf{J}_1 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & \end{bmatrix}$ and $\mathbf{J}_2 = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & \end{bmatrix}$ denote the condition effect indicator matrices, and the entry x_{ij} in the matrix $\mathbf{x} = \begin{bmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \\ x_{13} & x_{23} \\ x_{14} & \end{bmatrix}$ denotes the amount of fertilization for the blueberry of ith condition on the jth plot. We are now fitting the model

$$\mathbf{y} = \sum_{j=1}^{2} \beta_j \mathbf{J}_j + \beta \mathbf{x} + \boldsymbol{\varepsilon}$$

to the yield of blueberry.

And now we have $\mathbf{y} = \begin{bmatrix} 84 & 35 \\ 105 & 167 \\ 146 & 91 \\ 120 & \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} 3 & 2 \\ 4 & 8 \\ 6 & 4 \\ 5 & \end{bmatrix}$.

a(6%) Estimate β_1, β_2, and β.

b(6%) Compute S^2 and the standard error of $\hat{\beta}$.

c(6%) Use α=0.05 to test H_0: $\beta_1 = \beta_2$ against the appropriate alternative.