國立政治大學105學年度第1學期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	105年9月5日 星期一

1. (10%) Let $\mathbf{A} \in \Re^{n\times n}$ be idempotent. Prove that the eigenvalues of $\mathbf{A}$ are either 1 or 0.

2. (10%) Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu_X, \sigma^2)$, $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} N(\mu_Y, \sigma^2)$, and μ_0 be an arbitrary constant. Let $\bar{X} = \sum_{i=1}^n X_i/n$ and $\bar{Y} = \sum_{i=1}^n Y_i/n$. Please find the distributions of:

   (a) (5%)
   $$\frac{X_1 + X_2}{|Y_1 - Y_2|}$$

   (b) (5%)
   $$\frac{\bar{X} - \bar{Y} - \mu_0}{\sqrt{\sum_{i=1}^n \left[(X_i - Y_i) - (\bar{X} - \bar{Y})\right]^2}}$$

3. (20%) Let $J \sim Poisson(\phi)$ and $U|J = j \sim \chi^2_{p+2j}$, then unconditionally, U has the noncentral chi-square distribution with noncentrality parameter ϕ and denoted by $U \sim \chi^2_p(\phi)$.

   (a) (10%) Please derive the mean and variance of U.

   (b) (10%) Please prove that $\Pr(U > c)$ is strictly increasing in ϕ for fixed p and $c > 0$.

4. (25%) Consider the linear model in matrix form: $\mathbf{y}_{n\times 1} = \mathbf{X}_{n\times p}\boldsymbol{\beta}_{p\times 1} + \boldsymbol{\varepsilon}_{n\times 1}$ where $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$ where $\sigma^2 > 0$ and $\mathbf{I}_n$ is the $n \times n$ identity matrix. Moreover, let $p < n$. $\mathbf{X}$ is of full column rank, and $\mathbf{1}_n$ be an n-dimensional 1 vector. Define the projection matrix of an arbitrary matrix $\mathbf{A}$ as $\mathbf{P}_\mathbf{A}$.

   (a) (5%) Let $SSE = \mathbf{y}^T(\mathbf{I}_n - \mathbf{P}_\mathbf{X})\mathbf{y}$. $E(SSE) =$?

   (b) (5%) What is the distribution of SSE/σ^2?

   (c) (5%) Let $\hat{\mathbf{y}} = \mathbf{P}_\mathbf{X}\mathbf{y}$ and $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$. Please prove that $\hat{\mathbf{y}}$ and $\mathbf{e}$ are independent.

   (d) (5%) Assume that the first column of $\mathbf{X}$ is $\mathbf{1}_n$. Please prove that $\mathbf{y}^T\mathbf{P}_{\mathbf{1}_n}\mathbf{y}$ and $\mathbf{y}^T(\mathbf{I}_n - \mathbf{P}_{\mathbf{X}_n})\mathbf{y}$ are independent.

   (e) (5%) If 1_n does not fall in the column space of $\mathbf{X}$, please prove or disprove that $\mathbf{y}^T\mathbf{P}_{\mathbf{1}_n}\mathbf{y}$ and $\mathbf{y}^T(\mathbf{I}_n - \mathbf{P}_{\mathbf{X}_n})\mathbf{y}$ are independent.

5. (25%) Consider the following two-way design model

$$y_{ij} = \mu + \beta_i + \tau_j + \epsilon_{ij}, \quad \epsilon_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2) \tag{1}$$

where $i = 1, 2$ and $j = 1, 2, 3$. For convenience, we set the constraints $\beta_2 = 0$ and $\tau_3 = 0$. Denote $\bar{y}_{i.} = \sum_{j=1}^3 y_{ij}/3$, $\bar{y}_{.j} = \sum_{i=1}^2 y_{ij}/2$, and $\bar{y}_{..} = \sum_{i=1}^2 \sum_{j=1}^3 y_{ij}/6$.

考 試 科 目	線性模式專題	卷 別	第一卷	考 試 日 期	105 年 9 月 5 日 星期一

(a) (10%) Prove or disprove that the following terms are estimable.

i. (2%) μ

ii. (2%) $\mu + \beta_1$

iii. (2%) $\mu + \beta_1 + \tau_2$

iv. (2%) $\beta_1 + \tau_2$

v. (2%) $\beta_1 - \tau_2$

(b) (5%) In order to examine that $H_0 : \beta_1 = \beta_2$ versus $H_a : \beta_1 \neq \beta_2$, we use the statistic

$$F_\beta = \kappa \frac{\sum_{i=1}^{2} (\bar{y}_{i.} - \bar{y}_{..})^2}{\sum_{i=1}^{2} \sum_{j=1}^{3} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2} \tag{2}$$

where $F_\beta \sim F_{1,2}$ under H_0. Please find κ.

(c) (10%) Now suppose that we have the two-way model (1) and assume that $\beta_2 = 0$ and $\tau_j \overset{i.i.d.}{\sim} N(0, \sigma_\tau^2)$ where τ_j's and ϵ_{ij}'s are mutually independent. Given the κ derived in (b), what is the distribution of F_β defined in (2) under alternative hypothesis? Is this test statistic valid to examine $H_0 : \beta_1 = \beta_2$? Please address your reason in detail.

6. (10%) Consider a student opinion poll of instructors' classroom performance. The experimental design is as follows. Three instructors, A, B, and C, are participated in the experiment. All of them are going to teach Statistics in the coming semester. A will teach the classes in the Department of Statistics and the Department of Business Managements; B will teach the classes in the Department of Accounting and the Department of Economics; and C will teach the class in the Department of Epidemiology and the Department of Civil Engineering. In each class, 10 students will be randomly chosen to participate in the poll. Please define a model for this design which can tell the differences of performances among these instructors. Define your notation clearly.

國立政治大學 105 學年度第 1 學期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	105 年 9 月 5 日 星期一

1. (20pts). Let $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where X is an $n \times k$ constant matrix, $\boldsymbol{\beta}$ is a unknown k-vector, and $\boldsymbol{\epsilon}$ is a random vector in $\mathcal{R}^n$. We say $\hat{\boldsymbol{\beta}}$ is a least-squares estimator (LSE) of $\boldsymbol{\beta}$ if it minimizes the sum of squares of deviations of the n observed y's from their predicted values $\hat{y}$. Usually, the estimator $\hat{\boldsymbol{\beta}}$ is linear in $\mathbf{Y}$. However, it can be nonlinear under some circumstances. Please give an example where $\hat{\boldsymbol{\beta}}$ is not linear in Y.

2. Consider two linear regression models. Model (A) is $\mathbf{Y} = X_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}$ and Model (B) is $\mathbf{Y} = X_1\boldsymbol{\beta}_1 + X_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$ where X_1 and X_2 are $n \times k_1$ and $n \times k_2$ constant matrices, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are unknown k_1-vector and k_2-vector, and $\boldsymbol{\epsilon}$ is a random vector in $\mathcal{R}^n$ with mean $\mathbf{0}_n$ and covariance matrix $\sigma^2 I_{n\times n}$. Suppose matrix $X_{n\times(k_1+k_2)} = [X_1 \ \ X_2]$ has full column rank. Denote $\hat{\boldsymbol{\beta}}_1$ the unique LSE under model (A) and $\tilde{\boldsymbol{\beta}}_1$ and$\tilde{\boldsymbol{\beta}}_2$ the unique LSE under model (B).

   (a) (20pts). If (A) is the true model, are $\hat{\boldsymbol{\beta}}_1$ and $\tilde{\boldsymbol{\beta}}_1$ both unbiased? Justify your answer.

   (b) (10pts). Again, suppose (A) is the true model. Which of $\hat{\boldsymbol{\beta}}_1$ and $\tilde{\boldsymbol{\beta}}_1$ has smaller mse? Simply give your reason.

   (c) (20pts). Please give a necessary and sufficient condition (on X) such that $\hat{\boldsymbol{\beta}}_1 = \tilde{\boldsymbol{\beta}}_1$. Justify your answer.

3. Let V_1, V_2, V_3 be subspaces in $\mathcal{R}^n$ of dimension r_1, r_2, r_3 respectively and P_{V_1}, P_{V_2}, P_{V_3} be the corresponding projection matrices. It can be shown that (L1): if $V_1 \bigcap V_2 = \{\mathbf{0}_n\}$ and $P_{V_1}P_{V_2} + P_{V_2}P_{V_1} = O_{n\times n}$, then $V_1 \perp V_2$

   (a) (15pts). Please use (L1) to show that if $P_{V_1} + P_{V_2} + P_{V_3} = I_{n\times n}$, then V_1, V_2, V_3 are mutually orthogonal. (Hint: First prove that $V_1 \bigcap V_2 = \{\mathbf{0}_n\}$. Next obtain $P_{V_1}P_{V_2} + P_{V_2}P_{V_1} = O_{n\times n}$ by looking at $P_{V_1} + P_{V_2} = I - P_{V_3}$.

   (b) (15pts). Suppose $\mathbf{Y}$ has a multivariate normal distribution with mean $\mathbf{0}_n$ and covariance matrix $I_{n\times n}$. It can be shown that (L2): $\mathbf{Y}'A\mathbf{Y}$ for symmetric matrix $A_{n\times n}$ is chi-square distributed with df r if and only if $A = P_{V_A}$ for some r-dimensional subspace V_A in $\mathcal{R}^n$. Furthermore, (L3): $\mathbf{Y}'A\mathbf{Y}, \mathbf{Y}'B\mathbf{Y}, \mathbf{Y}'C\mathbf{Y}$ are independent if and only if V_A, V_B, V_C are mutually orthogonal. Please simply explain how you would use (a) to prove that " if $\mathbf{Y}'P_{V_1}\mathbf{Y} + \mathbf{Y}'P_{V_1}\mathbf{Y} + \mathbf{Y}'P_{V_1}\mathbf{Y}$ is chi-square distributed with df n if and only if $\mathbf{Y}'P_{V_1}\mathbf{Y}$, $\mathbf{Y}'P_{V_2}\mathbf{Y}$, $\mathbf{Y}'P_{V_3}\mathbf{Y}$ are independent and $r_1 + r_2 + r_3 = n$".

國立政治大學 105 學年度第 1 學期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	105 年 9 月 5 日 星期一

1. (25%) Given the linear model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{X}$ is $n \times p$ of rank p and $\boldsymbol{\epsilon} \sim MN(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$, we wish to test

$$H_0 : \boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{c}, \tag{1}$$

where $\boldsymbol{A}$ is $q \times p$ of rank q.

(1) Please derive a likelihood ratio test for (1).

(2) Please derive an F-test for (1).

2. (25%) Let $\boldsymbol{Y}$ satisfy the independent Poisson model with parameter vector $\boldsymbol{\Lambda}$. Let S be the sum of the components of $\boldsymbol{Y}$, and λ be the sum of the components of $\boldsymbol{\Lambda}$.

(1) Please show that conditional on $S = s$, $\boldsymbol{Y}$ has the multinomial distribution with parameters $n = s$, and $\boldsymbol{p} = \boldsymbol{\Lambda}/\lambda$.

(2) Suppose that X_1 and X_2 have Poisson distributions with parameters λ_1 and λ_2, and X_1 and X_2 are independent. Please construct a confidence interval on $R = \lambda_1/\lambda_2$.

3. (30%) Consider the log of the expected frequency in cell ijk in a three-way contigency table as follows:

$$\mu_{ijk} = log(m_{ijk}) = \lambda + \lambda_i^1 + \lambda_j^2 + \lambda_k^3 + \lambda_{ij}^{12} + \lambda_{ik}^{13} + \lambda_{jk}^{23} + \lambda_{ijk}^{123},$$

for $i = 1, \cdots, I$, $j = 1, \cdots, J$, and $k = 1, \cdots, K$.

(1) Which model corresponds to conditional independence of factors 1 and 3, for each level of factor 2?

(2) For which model are factors 1 and 3 jointly independent of factor 2?

(3) Give two $2 \times 3 \times 4$ tables $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, consisting only of -1's, 0's and 1's, which span the interaction space V_{12}, corresponding to the term λ_{ij}^{12}. Express λ_{11}^{12}, λ_{12}^{12}, and λ_{13}^{12} in terms of $(\boldsymbol{\mu}, \boldsymbol{x})$ and $(\boldsymbol{\mu}, \boldsymbol{x}_2)$.

國立政治大學 105 學年度第 1 學期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	105 年 9 月 5 日 星期一

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.