

UNIVERSITY OF MASSACHUSETTS
Department of Mathematics and Statistics
ADVANCED EXAM - LINEAR MODELS
Friday, Sept. 1 2006

Work all problems. 75 points are required to pass.

1. (10 points) Consider the standard linear model with $E(\mathbf{Y}) = \mathbf{X}\beta$ where \mathbf{Y} is a random vector with n elements; β is a r -vector of parameters; and \mathbf{X} is an $n \times r$ matrix not necessarily of full rank.
 - (a) Let $\phi(\beta)$ be a linear function of β . Give the definition that $\phi(\beta)$ is estimable.
 - (b) Consider a fixed effects one-way ANOVA model without restriction $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ with $\varepsilon_{ij} \sim N(0, \sigma^2)$, for $i = 1, \dots, I$, $j = 1, \dots, n_i$. In each of the following questions, if your answer is 'yes', then provide an estimator; if your answer is 'no', then prove it.
 - i. Is $\mu + \alpha_1$ estimable?
 - ii. Is $\alpha_1 - \alpha_2$ estimable?
 - iii. Is $\alpha_1 + \alpha_2$ estimable?
2. (25 points) Consider the following three-treatment analysis of covariance model:

$$y_{ij} = \mu + \tau_i + \gamma_i x_{ij} + \varepsilon_{ij}$$

for $i = 1, \dots, 3$, $j = 1, \dots, n_i$, where we assume $\tau_1 + \tau_2 + \tau_3 = 0$ and the ε_{ij} are i.i.d. $N(0, \sigma^2)$.

- (a) Making use of the constraint in τ_i by representing τ_3 in terms of τ_1 and τ_2 , write the model in the following matrix form

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon,$$

by specifying \mathbf{Y} , \mathbf{X} , β , and ε .

We will assume that the x_{ij} are such that with this constraint incorporated \mathbf{X} is of full column rank. There is no need in any of the following expressions to simplify $(\mathbf{X}'\mathbf{X})^{-1}$; you can leave it that way where needed.

- (b) Write down the expression (in matrix form) for the L.S. estimator $\hat{\beta}$ of β and state the distribution of $\hat{\beta}$.
- (c) Write down the expression for SSE (as some function of \mathbf{Y} , \mathbf{X} and/or $\hat{\beta}$) and specify the degrees of freedom of SSE.
- (d) Set-up a 90% confidence interval for $\tau_1 - \tau_3$.
- (e) Suppose we want to perform a statistical test with $\alpha = 0.05$ to see if the three groups have the same slope, i.e., to test the hypotheses

$$H_0 : \gamma_1 = \gamma_2 = \gamma_3 \text{ vs. } H_1 : \gamma_i \neq \gamma_{i'} \text{ for some } i \text{ and } i'$$

- i. Specify \mathbf{Y} , \mathbf{Z} and δ to write the reduced model (under H_0) in matrix form

$$\mathbf{Y} = \mathbf{Z}\delta + \varepsilon,$$

- ii. Write down an expression for the error sum of squares, say SSE_1 for the reduced model, and specify the degrees of freedom of SSE_1 .
- iii. Give the likelihood ratio test in terms of SSE and SSE_1 (and other constants as needed), and provide the rejection rule for the test.
3. (a) Let Z_1, \dots, Z_m be independent normally distributed random variables with $E(Z_i) = \mu_i$ and $V(Z_i) = 1$. Define the non-central chi-square distribution in terms of the distribution of a random variable C , which is some function of the Z_i 's. (You don't need to derive a density function, just state that the non-central chi-square distribution is defined as the distribution of $C = \dots$). What is the degrees of freedom? What is the non-centrality parameter?
- (b) Let $\mathbf{Y} \sim N(\boldsymbol{\mu}, \mathbf{I})$ where \mathbf{Y} is an $n \times 1$ random vector. Prove that if \mathbf{A} is an $n \times n$ symmetric idempotent matrix of rank d then the quadratic form has a non-central chi-square distribution. Give the degrees of freedom and the non-centrality parameter (which should be given as a function of \mathbf{A} and σ^2).
- (c) Consider the general linear model of full rank $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon$ assuming $E(\varepsilon) = \mathbf{0}$ and $Cov(\varepsilon) = \sigma^2\mathbf{I}$ and \mathbf{X} is $n \times p$ of rank p . Consider $S^2 = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})/(n-p)$. Assuming \mathbf{Y} is normally distributed, prove that $(n-p)S^2/\sigma^2$ is distributed chi-square. What is the degrees of freedom? What is the non-centrality parameter.
- (d) Using the setting of the previous problem **derive** $E(S^2)$ without assuming normality. (You can state and use without proof the result for the expected value of a quadratic form if you know it).
4. (15 points) Consider the standard two way analysis of variance model $Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$, ($i=1$ to I , $j=1$ to J and $k=1$ to n_{ij}) where μ_{ij} is the fixed mean associated with level i of factor A and level j of factor T and the ε_{ijk} are assumed iid $N(0, \sigma^2)$.
- (a) Let $\theta = \mathbf{c}'\boldsymbol{\mu} = \sum_i \sum_j c_{ij}\mu_{ij}$ Give a 95% confidence interval for θ .
- (b) Give simultaneous confidence intervals for all linear combinations of the form $\sum_i c_i\mu_i$, where $\mu_i = \sum_j \mu_{ij}/J$ using Scheffe's technique.
- (c) Consider testing the hypothesis $H_{0A} : \mu_{1.} = \dots = \mu_{I.}$ with the μ_i defined as in the preceding part. First express this hypothesis as $\mathbf{A}\boldsymbol{\mu} = \mathbf{0}$ and then provide the F-test for it. The test statistic can be left in matrix form but state explicitly what $\hat{\boldsymbol{\mu}}$ is. State the distribution of the F statistic under the null and under the alternative. Be sure to state the degrees of freedom involved.
5. (25 points) Consider the two way model in problem 4, but now assume the data is balanced, $n_{ij} = M$ for each (i,j) . The effects version of the model decomposes μ_{ij} and writes the model as

$$Y_{ijk} = \mu + A_i + T_j + G_{ij} + \varepsilon_{ijk},$$

where A_i is the effect of level i of factor A , T_j is the effect of level j of factor T and G_{ij} is the interaction. Consider the random effects version of the model where the A_i are iid $N(0, \sigma_A^2)$, the T_i are iid $N(0, \sigma_T^2)$, the G_{ij} are iid $N(0, \sigma_G^2)$ and all the random quantities are independent.

- (a) Find $V(Y_{ijk})$.
- (b) Find $Cov(Y_{ijk}, Y_{i'j'k'})$ for any two (i, j, k) and (i', j', k') .
- (c) Consider $\bar{Y}_{i..} = \sum_j \sum_k Y_{ijk} / JM$.
- Find $E(\bar{Y}_{i..})$, $V(\bar{Y}_{i..})$ and $Cov(\bar{Y}_{i..}, \bar{Y}_{i'..})$ for $i \neq i'$.
- (d) Consider $SSA = \sum_i (\bar{Y}_{i..} - \bar{Y}_{...})^2$ and $SSG = \sum_i (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$. Find c_a and c_g such that $c_a SSA$ and $c_g SSG$ are each distributed as a central chi-square.
- (e) Use the previous part and the fact that SSA and SSG are independent (you don't have to prove this) to derive an F-test of $H_0 : \sigma_A^2 = 0$. Give the test statistic, its distribution (in general, under the null or not) and describe the power function of the test.