Masters Comprehensive Examination
Department of Statistics, University of Florida
August 23, 2002, 8:00am - 12:00 noon

Instructions:

1. You have exactly four hours to answer questions in this examination.
2. There are 10 problems of which you must answer 8.
3. Only your first 8 problems will be graded.
4. Write only on one side of the paper, and start each question on a new page.
5. Write your **number** on every page.
6. Do not write your name anywhere on your exam.
7. You must show your work to receive credit.
8. While the ten questions are equally weighted, within a given question, the parts may have different weights.
9. You are allowed to use a calculator.

- iid = independent and identically distributed
- pdf = probability density function
- pmf = probability mass function
- mgf = moment generating function
- MLE = maximum likelihood estimator
- MOM = method of moments
- MSE = mean squared error
- UMVUE = uniformly minimum variance unbiased estimator

You may use the following facts/formulas without proof:

Iterated expectation: For any two random variables X and Y,

$$\mathrm{E}(X) = \mathrm{E}\left[\mathrm{E}(X|Y)\right] \;,$$

provided the expectations exist.

Beta Density: $X \sim \text{Beta}(\alpha, \beta)$ means X has pdf

$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, x^{\alpha-1}\,(1-x)^{\beta-1}\, I_{(0,1)}(x)$$

where $\alpha > 0$ and $\beta > 0$. $\mathrm{E}(X) = \alpha/(\alpha+\beta)$ and $\text{Var}(X) = \alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$.

1. Let $X_1, \dots, X_n$ be iid Uniform$(0, \theta)$ where $\theta > 0$; that is, the common pdf is

$$f(x|\theta) = \theta^{-1} I(0 \le x \le \theta) \ .$$

(a) Find the MLE of θ, call it $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$.

(b) Find the pdf of $\hat{\theta}$ and show that $\hat{\theta}/\theta$ has a beta distribution.

(c) Show that $n\left(1 - \frac{\hat{\theta}}{\theta}\right)$ converges in distribution and find the limiting distribution.

(d) Find the method of moments estimator of θ, call it $\tilde{\theta} = \tilde{\theta}(X_1, \dots, X_n)$.

(e) Compare the two estimators using MSE.

2. (a) Let $Z|\theta \sim$ Geometric(θ); that is,

$$P(Z = z|\theta) = \theta(1-\theta)^z$$

for $z \in \{0, 1, 2, \dots\}$ and $\theta \in (0, 1)$. Note that E$(Z|\theta) = \frac{1-\theta}{\theta}$ and Var$(Z|\theta) = \frac{1-\theta}{\theta^2}$. Find the *marginal* pmf of Z assuming that $\theta \sim$ Beta(α, β).

(b) Suppose that $Z_1, \dots, Z_n$ are iid with pmf given by

$$P_\alpha(Z = z) = \frac{\alpha^2\, \Gamma(\alpha)\, z!}{\Gamma(z + \alpha + 2)}$$

for $z \in \{0, 1, 2, \dots\}$ and $\alpha > 2$. Find the MOM estimator of α, call it $\tilde{\alpha}_n = \tilde{\alpha}_n(Z_1, \dots, Z_n)$. (Hint: Iterated expectation.)

(c) Show that, in general, for any two random variables X and Y,

$$\text{Var}(X) = \text{E}\left[\text{Var}(X|Y)\right] + \text{Var}\left[\text{E}(X|Y)\right] \ ,$$

provided the expectations exist.

(d) Show that

$$\sqrt{n}\left(\tilde{\alpha}_n - \alpha\right) \xrightarrow{d} \text{N}\left(0, \frac{\alpha^2(\alpha-1)^2}{\alpha-2}\right) .$$

3. It is sometimes reasonable to think of the lifetime of a piece of equipment as a random variable following an Exp(λ) distribution where λ is unknown. Suppose n identical pieces of equipment are to be run until they fail and the (random) failure times are given by $X_1, \dots, X_n$. Assume that these are a random sample from the Exp(λ) distribution. Consider the probability of early failure given by

$$e(\lambda) = P_\lambda(X_1 < x) = 1 - e^{-x/\lambda}$$

for some fixed $x > 0$.

(a) Find the MLE of $e(\lambda)$.

(b) Find the best unbiased estimator (or UMVUE) of $e(\lambda)$. (Your answer must be a closed form function of the data.) Hint:

$$P(X < x|Y = y) = P\left(\frac{X}{Y} < \frac{x}{y}\Big| Y = y\right) \ .$$

4. Let X_1, X_2 be iid Uniform$(\theta, \theta+1)$. In this question, we consider testing $H_0 : \theta = 0$ against $H_1 : \theta > 0$ using the following two tests:

$$\phi_1(X_1) : \text{ Reject } H_0 \text{ if } X_1 > .95$$
$$\phi_2(X_1, X_2) : \text{ Reject } H_0 \text{ if } X_1 + X_2 > C$$

where $C > 0$.

(a) Show that $X_1 - \theta$ and $X_2 - \theta$ are iid Uniform$(0, 1)$.

(b) Derive the power function of ϕ_1; that is, find $\beta_1(\theta) = P_\theta(X_1 > .95)$ for $\theta \geq 0$. (Hint: What happens if $\theta > .95$?)

(c) Let U_1, U_2 be iid Uniform$(0, 1)$. Derive the pdf of $V = U_1 + U_2$.

(d) Derive the power function of ϕ_2; that is, find $\beta_2(\theta) = P_\theta(X_1 + X_2 > C)$ for $\theta \geq 0$. (Hint: This function will take three different forms for three different ranges of θ.)

(e) What value of C should we use if we want the two tests to have the same size? With this value of C, is ϕ_2 more powerful than ϕ_1?

5. (a) Derive the mgf of a standard normal random variable.

(b) Let Z be a random variable with mgf $M_Z(t)$ and let a and b be constants. Show that the mgf of $aZ + b$ is $e^{tb} M_Z(at)$. Use this result to write down the mgf of a N(μ, σ^2) random variable.

(c) Suppose that $W_1, \ldots, W_m$ are independent random variables such that $W_i \sim$ N(μ_i, σ^2). Find the distribution of $\sum_{i=1}^n a_i W_i$ where $a_1, \ldots, a_m$ are known constants.

(d) Suppose that the random variables $Y_1, \ldots, Y_n$ satisfy

$$Y_i = \beta x_i + \varepsilon_i \ ,$$

for $i = 1, \ldots, n$ where $x_1, \ldots, x_n$ are known constants and $\varepsilon_1, \ldots, \varepsilon_n$ are iid N$(0, \tau^2)$. Find a two-dimensional sufficient statistic for the unknown parameter (β, τ^2).

(e) Find the MLE of β, call it $\hat{\beta} = \hat{\beta}(Y_1, \ldots, Y_n)$. Show that $\hat{\beta}$ is unbiased.

(f) Find the distribution of $\hat{\beta}$.

(g) Find the distribution of the alternative estimator of β given by

$$\tilde{\beta} = \tilde{\beta}(Y_1, \ldots, Y_n) = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n x_i}$$

(h) Which of these estimators is better?

6. A regression model is fit, relating a response variable Y to X in a nonlinear (quadratic) fashion through the origin:

$$Y_i = \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i \ ,$$

for $i = 1, \ldots, 8$ where $\varepsilon_1, \ldots, \varepsilon_8$ are iid $\mathrm{N}(0, 4)$. Of course, in matrix terms, these equations can be written simultaneously as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. The data pairs (X_i, Y_i) are: (1,6), (1,4), (2,8), (2,12), (3,18), (3,22), (4,32), (4,36). **Keep all answers in all parts of this problem in rational form to avoid variations due to round-off, no need to simplify fractions.**

(a) Give the least squares estimate of $\boldsymbol{\beta}$.

(b) What is the rate of change of $E(Y)$ as a function of X?

(c) Give the least squares estimate and standard error for your answer in part (b).

7. Suppose we obtain $Y_i =$ *yearly salary* and $X_{i1} =$ *years of experience* for a random sample of n professors at the University of Florida. Suppose further that we note which of the professors are male and which are female. If we are interested in establishing a gender-based difference in salaries, after controlling for experience, we might use the model

$$\mathrm{E}(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} \tag{1}$$

where X_{i2} is an indicator that takes the value 0 for males and 1 for females. Under this model, the expected salaries for men and women are $\beta_0 + \beta_1 X_{i1}$ and $\beta_0 + \beta_2 + \beta_1 X_{i1}$, respectively. Of course, the way we defined the indicator in model (1) was completely arbitrary. In particular, the model

$$\mathrm{E}(Y_i) = \beta_0^* + \beta_1^* X_{i1} + \beta_2^* X_{i3} \tag{2}$$

where X_{i3} is an indicator that takes the value 1 for males and 0 for females seems equally legitimate.

(a) What are the expected salaries for men and women under model (2)? Use this to write the β^*'s as functions of the β's.

(b) Let $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ and let $\boldsymbol{\beta}^* = (\beta_0^*, \beta_1^*, \beta_2^*)'$. Find a 3×3 matrix $\mathbf{M}$ such that $\boldsymbol{\beta}^* = \mathbf{M}\boldsymbol{\beta}$.

(c) Show that $\mathbf{M}^{-1} = \mathbf{M}$.

(d) Let $\mathbf{X}$ and $\mathbf{X}^*$ denote the design matrices corresponding to models (1) and (2), respectively; that is,

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{12} \\ 1 & X_{21} & X_{22} \\ \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} \end{bmatrix} \qquad \mathbf{X}^* = \begin{bmatrix} 1 & X_{11} & X_{13} \\ 1 & X_{21} & X_{23} \\ \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n3} \end{bmatrix}.$$

How can $\mathbf{M}$ be used to relate $\mathbf{X}^*$ to $\mathbf{X}$?

(e) Use matrix algebra to find the least squares estimate of $\boldsymbol{\beta}^*$ in terms of the least squares estimate of $\boldsymbol{\beta}$. (Hint: Recall that $(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\,\mathbf{B}^{-1}\,\mathbf{A}^{-1}$ as long as $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are square and full rank.)

(f) Is the result surprising? Why?

8. A chemical production process consists of a first reaction with an alcohol and a second reaction with a base. A 3×2 factorial with three alcohols and 2 bases (these are the only alcohol and base types of interest to the researcher) was conducted with three replicate reactions per treatment. The design was completely randomized. The data are given in the following table:

	Base	
Alcohol	1	2
1	93, 90, 87	77, 83, 80
2	80, 78, 82	91, 88, 91
3	81, 89, 85	84, 84, 87

(a) Write a linear model for the experiment, explaining all terms. Compute the Analysis of Variance.

(b) Test whether there is a base by alcohol interaction with respect to yields. Use the $\alpha = 0.05$ significance level. Fully state null and alternative hypotheses with respect to your model parameters, give test statistic, rejection region, and a sketch representing the P-value of the test.

(c) Compare the two bases under each alcohol type using Bonferroni's method with simultaneous 95% confidence intervals (how many comparisons are being made?). Clearly describe your conclusions. Give an interaction plot of results in terms of the sample means.

9. An experiment to compare five drugs for reducing blood pressure was conducted in a completely randomized design with equal replication. The sample means and standard deviations of reduction are given below.

Drug	$\overline{Y}_i$	S_i
1	10	6
2	12	8
3	8	6
4	14	$7.07 = \sqrt{50}$
5	16	8

(a) Assuming these are the only five drugs of interest, what would be the minimum common replicate size so that these treatment means can be considered not all equal at the $\alpha = 0.05$ significance level? Use linear interpolation whenever necessary.

(b) Now assume that these are five of a virtually infinite number of potential drugs. (i) Give the estimate of the variance component corresponding to treatment effects as a function of common replicate size. (ii) What would be the replicate size corresponding to this variance component estimate going from negative to positive?

10. A regression model was fit relating Y (log sales among stores in a retail chain) to X_1 (log area in square feet of the store), X_2 (log inventory), and X_3 (log total income of households within 5 miles) as well as an intercept term based on a sample of $n = 20$ stores in the chain. The model is

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \ ,$$

where $i = 1, \ldots, 20$ and $\varepsilon_1, \ldots, \varepsilon_{20}$ are iid $\mathrm{N}(0, \sigma^2)$. The sequential (Type I) and partial (Type II) sums of squares are given below for the model where $SS(\text{Total Corrected}) = 2000.0$:

Term	Type I SS	Type II SS
Intercept	5000	20
X_1 (log area)	1300	400
X_2 (log inventory)	100	100
X_3 (log income)	200	200

(a) Give the following partial sum of squares: $R(\beta_2, \beta_3 | \beta_0)$.

(b) Test $H_0 : \beta_2 = \beta_3 = 0$ vs $H_A : \beta_2 \neq 0$ and/or $\beta_3 \neq 0$ under two different assumptions. First do the test conditional on X_1 being in the model, and then do the test assuming X_1 is not in the model. Do each test at $\alpha = 0.05$. (Note that it makes sense to do both of these tests since the β's have different interpretations depending on whether X_1 is in the model.)