Masters Comprehensive Examination
Department of Statistics, University of Florida
May 16, 2003, 8:00 am - 12:00 noon

Instructions:

1. You have four hours to answer questions in this examination.

2. You must show your work to receive credit.

3. There are 10 problems of which you must answer 8.

4. Only your first 8 problems will be graded.

5. The last page contains two tables that you may need.

6. While the 10 questions are equally weighted, some problems are more difficult than others.

7. The parts within a given question are not necessarily equally weighted.

8. Write only on one side of the paper, and start each question on a new page.

9. You are allowed to use a calculator.

The following abbreviations are used throughout:

- ANOVA = analysis of variance

- CDF = cumulative distribution function

- CRD = completely randomized design

- iid = independent and identically distributed

- LRT = likelihood ratio test

- mgf = moment generating function

- ML = maximum likelihood

- pdf = probability density function

- pmf = probability mass function

- UMVUE = uniformly minimum variance unbiased estimator

- $\varepsilon_i \sim NID(0, \sigma^2)$ means that the ε_is are iid N$(0, \sigma^2)$.

1. Consider the CRD, stated as the cell means model with 3 fixed treatments, 5 replicates per treatment, and the following error structure:

$$y_{ij} = \mu_i + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim NID(0, \sigma^2) \qquad i = 1, 2, 3; \ \ j = 1, \ldots, 5$$

(a) Give the Treatment and Error sums of squares in symbolic form and derive the expected mean square for treatments. Hint: $\sum_i (x_i - \bar{x}_.)^2 = \sum x_i^2 - n\bar{x}_.^2$.

Now suppose the sample means (standard deviations) for the three treatments are: 50 (3), 60 (2), 40 (4).

(b) Obtain the Analysis of Variance.

(c) Test $H_0 : \mu_1 = \mu_2 = \mu_3$ vs H_A : Population means are not all equal at $\alpha = 0.05$ significance level.

(d) Use Bonferroni's method with an experiment-wise error rate of $\alpha = 0.05$ to compare all pairs of treatments.

2. An internet based firm has 2 sources of advertising: newspaper/magazines and television/radio. They vary the amounts of advertising of each type (X_1 and X_2, in thousands of dollars, respectively) across a sample of $n = 8$ similar sized markets, and obtain how many times their website is reached by people from each market (Y, in 1000s of "hits"). They fit the multiple regression model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \qquad \varepsilon_i \sim NID(0, \sigma^2)$$

and obtain the following estimated regression equation, residual mean square, and $(\mathbf{X'X})^{-1}$ matrix:

$$\hat{Y}_i = 1.8 + 4.5 X_{i1} + 3.5 X_{i2} \qquad s^2 = 1.87 \qquad (\mathbf{X'X})^{-1} = \begin{bmatrix} 0.4 & -0.12 & -0.12 \\ -0.12 & 0.096 & 0.016 \\ -0.12 & 0.016 & 0.096 \end{bmatrix}$$

(a) Give the predicted number of "hits" if they spend \$2000 on newspaper/magazines and \$2000 on television/radio (be careful of units).

(b) Test whether the mean number of hits when $X_1 = 2.0$ and $X_2 = 2.0$ is equal to 20.0. Set this up and conduct it as a test of the form: $H_0 : \mathbf{K'\beta} = \mathbf{m}$ and test at $\alpha = 0.05$ significance level.

(c) Suppose they fit a response surface and obtained the following least squares estimate of the regression equation containing an intercept, all linear and quadratic terms, and a cross-product term.

$$\hat{Y} = 0.5 + 6.0 X_1 + 4.0 X_2 - 1.0 X_1^2 - 0.5 X_2^2 + 1.0 X_1 X_2 \ .$$

Their budget is limited to \$5000 (recall units), and they must spend all of it. How should they allocate their budget between newspaper/magazine ads (X_1) and television/radio advertising (X_2) to maximize the predicted number of hits based on the estimated regression equation.

3. A clothing manufacturer buys three types of yarn. They are interested in comparing the three types of yarn in terms of mean breaking strength. They obtain random samples of each type from two suppliers and observe the sample mean breaking strengths (in psi) based on 10 replicates per yarn type from each supplier. A summary of the data follows.

| | Yarn Type | | | |
Supplier	1	2	3	Mean
1	35.0	55.0	75.0	55.0
2	45.0	45.0	105.0	65.0
Mean	40.0	50.0	90.0	60.0

(a) Obtain the Analysis of Variance, including the sources: yarn types, suppliers, interaction, and error $\left(\sum\sum (y_{ij} - \bar{y}_{..})^2 = 50,000 \right)$.

(b) Assuming that these are the only two suppliers for the manufacturer, test for (i) interaction, (ii) yarn type effects, and (iii) supplier effects by adding test statistics and critical values to your ANOVA table.

(c) Repeat the previous exercise under the assumption that these two suppliers are a random sample from a large number of suppliers.

4. In a study of the effect of Viagra, patients suffering from erectile dysfunction were randomly assigned to one of four doses of Viagra (in mg). The following table gives mean rating scores of erectile function after 24 weeks of treatment, as well as sample sizes (adjusted slightly for ease of computation).

Dose (X_i)	$\bar{Y}_{i.}$	r_i
0 (Placebo)	2.2	200
25	3.2	100
50	3.5	100
100	4.0	100

Consider two different models for these data:

$$\text{Model (a): } Y_{ij} = \beta_0 + X_{i1}\beta_1 + \varepsilon_{ij} \quad i = 1,\ldots,n; \quad j = 1,\ldots,r_i,$$

$$\text{Model (b): } \bar{Y}_{i.} = \beta_0 + X_{i1}\beta_1 + \bar{\varepsilon}_{i.} \quad i = 1,\ldots,n,$$

where ε_{ij} are uncorrelated random variables with mean 0 and variance σ^2.

(a) The weighted least squares estimator of $(\beta_0, \beta_1)'$ for Model (b) is the same as the ordinary least squares estimator for Model (a). Write out the matrix form of $\hat{\beta}_w$ for Model (b) with numbers (no symbols) that you would enter into a matrix language to get the estimate. You need not multiply or invert any of the matrices as long as they are stated properly with correct numbers in the proper form.

The actual weighted least squares estimate is: $\hat{\beta}_w = (2.40, 0.02)'$. Further, the variances for the 4 dose groups are: 7.8, 4.0, 4.0, and 4.0, respectively.

(b) Give the fitted values for subjects in the 4 dose groups based on the weighted least squares estimator of β.

(c) Conduct the F-test for lack-of-fit to test whether the true relationship is linear at the $\alpha = 0.05$ significance level.

3

5. In scalar form for the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad \varepsilon_i \sim NID(0, \sigma^2) \,,$$

the least squares estimates of β_1 and β_0 are the commonly known values:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \qquad \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} \,.$$

(a) Derive the mean and variance of the sampling distribution of $\hat{\beta}_1$.

(b) Derive the mean of the sampling distribution of $\hat{\beta}_0$.

6. This question deals with moment generating functions and applications to normal random variables.

(a) Show that if $X \sim N(\mu, \sigma^2)$ then $Z = (X - \mu)/\sigma \sim N(0, 1)$.

(b) Let W be a random variable with mgf $M_W(t)$ and suppose that a and b are constants. Show that the mgf of $aW + b$ can be written as $e^{bt} M_W(at)$.

(c) Show that the mgf of X is

$$M_X(t) = \exp\left\{ \mu t + \frac{\sigma^2 t^2}{2} \right\} \,.$$

(d) Derive the distribution of $Y = Z^2$.

(e) Are Y and Z uncorrelated?

(f) Are Y and Z independent? (This requires more than a "yes" or "no".)

7. Toss n fair dice. Pick up only those *not* showing a 6 and toss them. Continue doing this until all dice show a 6. Let X denote the total number of tosses made in this experiment.

(a) How many tosses do you expect to make if $n = 1$?

(b) Compute the CDF of X for general n. (Hint: Think about what happens with each individual die.)

(c) Suppose that Y is a discrete random variable whose support is contained in $\{0, 1, 2, \dots\}$. Show that

$$\sum_{y=0}^{\infty} \left[1 - F_Y(y) \right] = EY \,,$$

where $F_Y(\cdot)$ is the CDF of Y. (Hint: Write the left-hand side as a double sum.)

(d) Use the result from (c) to find the expected value of X when $n = 2$.

8. Let $X_1, X_2, \ldots, X_n$ be iid with common pdf

$$f(x|\theta) = \begin{cases} \theta x^{\theta-1} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

where $\theta > 0$.

(a) Show that the ML estimator of $\frac{1}{\theta}$ is $\delta(X_1, \ldots, X_n) = -\frac{1}{n}\sum_{i=1}^{n} \log X_i$.

(b) Calculate the expected value and the variance of $\delta(X_1, \ldots, X_n)$.

(c) Find the Cramér-Rao lower bound for the variance of an unbiased estimator of $\frac{1}{\theta}$. What can you conclude?

(d) Find a pivotal quantity that depends on the data only through δ.

(e) Suppose that $n = 1$ and that $\alpha \in (0, 1)$. Use the result in (d) to construct an interval estimator for $\frac{1}{\theta}$ with confidence coefficient $1 - \alpha$. Give *explicit* formulas for the endpoints that involve *only* X_1 and α.

9. Suppose that $X_1, \ldots, X_n$ are iid $N(0, \sigma^2)$ and that $Y_1, \ldots, Y_m$ are iid $N(0, \tau^2)$. Assume further that $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_m)$ are independent. Define $\lambda = \sigma^2/\tau^2$.

(a) Find the ML estimator of σ^2.

(b) Construct the LRT statistic for testing $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda \neq \lambda_0$ where $\lambda_0 > 0$.

(c) Show that the LRT statistic can be written in such a way that it involves the data only through the statistic

$$F = \frac{\frac{1}{m}\sum_{i=1}^{m} Y_i^2}{\frac{1}{n\lambda_0}\sum_{i=1}^{n} X_i^2}.$$

(d) Find the distribution of F under H_0.

(e) The general LRT theory tells us to reject H_0 when the LRT statistic is small. Give an equivalent rejection rule in terms of F.

10. This question deals with the concept of completeness and its applications.

(a) Consider a family of pdfs (or pmfs) $\{f(t|\theta) : \theta \in \Theta\}$. Explain exactly what it means to say that this family is *complete*.

(b) Consider a binomial family where the number of trials, n, is a fixed positive integer and the success probability, θ, is restricted to the set

$$\Theta = \left\{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\right\}.$$

Is this family complete? (Remember to **show your work**!)

(c) Let $X_1, \ldots, X_k$ be iid Binomial(n, θ) where n is a fixed, known positive integer and $\theta \in (0, 1)$ is unknown. Find the best unbiased estimator of $n\theta(1 - \theta)^{n-1}$; that is, the UMVUE of $n\theta(1 - \theta)^{n-1}$.