Masters Comprehensive Examination
Department of Statistics, University of Florida
May 10, 2002, 8:00am - 12:00 noon

Instructions:

1. You have 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. You must show your work to receive credit.
5. While the questions are equally weighted, some problems are more difficult than others.
6. Write only on one side of the paper, and start each question on a new page.
7. You are allowed to use a calculator.

The following abbreviations are used throughout:

- ANOVA = analysis of variance
- iid = independent and identically distributed
- LRT = likelihood ratio test
- pdf = probability density function
- pmf = probability mass function
- MLE = maximum likelihood estimator
- MOM = method of moments
- UMP = uniformly most powerful
- UMVUE = uniformly minimum variance unbiased estimator
- $\varepsilon_i \sim NID(0, \sigma^2)$ means that $\varepsilon_1, \ldots, \varepsilon_n$ are iid $\mathrm{N}(0, \sigma^2)$.

You may use the following facts/formulas without proof:

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

$$f(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\, \beta^\alpha}\, x^{\alpha-1}\, e^{-x/\beta}\, I(x > 0)$$

where $\alpha > 0$ and $\beta > 0$.

Distributional Result: If $X \sim \text{Gamma}(\alpha_x, \beta)$ and $Y \sim \text{Gamma}(\alpha_y, \beta)$ and X and Y are independent, then $X/(X+Y) \sim \text{Beta}(\alpha_x, \alpha_y)$.

1. Let X and Y be iid Exp(λ).

(a) Find the joint distribution of $V = X + Y$ and $W = X$.

(b) Are V and W independent?

(c) Express $P(V + W < \pi)$ as a double integral. (You do not need to evaluate it.)

(d) Find Cov(V, W).

(e) Show that, in general, for any two random variables S and T,

$$\mathrm{E}(S) = \mathrm{E}\left[\mathrm{E}(S|T)\right] \ ,$$

provided the expectations exist.

(f) For our specific (V, W), find $\mathrm{E}(W|V)$ and show that $\mathrm{E}\left[\mathrm{E}(W|V)\right] = \mathrm{E}(W)$.

2. Suppose that $X_1, \ldots, X_n$ are iid Exp(θ) and that $Y_1, \ldots, Y_m$ are iid Exp(μ). Assume further that $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_m)$ are independent.

(a) Find the MLE of θ.

(b) Construct the LRT statistic for testing $H_0 : \theta = \mu$ against $H_1 : \theta \neq \mu$.

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

$$T = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i + \sum_{i=1}^{m} Y_i} \ .$$

(d) The general LRT theory tells us to reject H_0 when $\lambda(x, y) < c$. Give an equivalent rejection rule in terms of T.

(e) Find the distribution of T under H_0.

(f) Suppose that $n = m = 1$. Give the exact rejection region for the size 0.05 LRT of $H_0 : \theta = \mu$ versus $H_1 : \theta \neq \mu$.

3. Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ and (finite) variance σ^2. Let $a_1, \ldots, a_n$ be constants.

(a) Show that the estimator $\sum_{i=1}^{n} a_i X_i$ is an unbiased estimator of μ if and only if $\sum_{i=1}^{n} a_i = 1$.

(b) Among all unbiased estimators of this form (called *linear unbiased estimators*) find the one with minimum variance *and* calculate the variance.

Now let $W_1, \ldots, W_k$ be unbiased estimators of a parameter θ with Var$(W_i) = \sigma_i^2$ and Cov$(W_i, W_j) = 0$ if $i \neq j$. Assume that the σ_i^2s are all known and finite and, again, let $a_1, \ldots, a_k$ be constants.

(c) Show that, of all the unbiased estimators of θ having the form $\sum_{i=1}^{k} a_i W_i$, the one with the smallest variance is

$$\phi(W_1, \ldots, W_k) = \frac{\sum_{i=1}^{k} \frac{W_i}{\sigma_i^2}}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2}} \ ,$$

and that

$$\mathrm{Var}\left[\phi(W_1, \ldots, W_k)\right] = \frac{1}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2}} \ .$$

4. 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 $X_1, \ldots, X_n$. 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:

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

5. Suppose that $X|\lambda \sim$ Poisson(λ) and that $\lambda \sim$ Gamma(α, β).

(a) Show that the marginal mass function of X is given by

$$P_{\alpha,\beta}(X = x) = \frac{\Gamma(\alpha + x)}{\Gamma(\alpha)\, x!\, \beta^\alpha} \left(\frac{\beta}{\beta + 1}\right)^{x+\alpha} \tag{1}$$

for $x \in \{0, 1, 2, \ldots\}$.

(b) Find the *marginal* mean and variance of X.

(c) Are there any (α, β) pairs for which X has a Geometric distribution? If yes, identify the success probability in terms of α and β.

(d) Suppose that $X_1, \ldots, X_n$ are iid from the pmf in (1). Find the MOM estimator of (α, β).

(e) Why might one prefer the MLE of (α, β) over the MOM estimator derived in (d)?

(f) Suppose that X_1, X_2 are iid from the following pmf

$$P_\beta(X = x) = \frac{\Gamma(2 + x)}{\Gamma(2)\, x!\, \beta^2} \left(\frac{\beta}{\beta + 1}\right)^{x+2}$$

for $x \in \{0, 1, 2, \ldots\}$. (This is, of course, the pmf in (1) with α known and equal to 2.) Find a UMP test of $H_0 : \beta \leq 1$ against $H_1 : \beta > 1$ with level equal to $\frac{13}{16}$.

6. Two statistical analysts are given the same data for a single response variable Y, and a single independent variable. The independent variable has been labeled with levels 1,2,3,4. John treats the independent variable as interval scale, and assumes that the relationship between the dependent and independent variables is linear (he has no reason to believe the mean response is 0 when the independent variable is 0). Jane treats the independent variable as nominal scale (no distinct ordering among the levels), making no assumption about the relationship between the dependent and independent variables. Both John and Jane believe that error terms are independent and normally distributed with constant variance.

(a) Write out John's statistical model.

(b) Write out Jane's statistical model.

(c) The following data were obtained. Give least squares estimates of all model parameters for John and Jane.

Trt (X)	Responses (Y)		
1	17	20	23
2	29	21	25
3	31	29	30
4	45	43	47

(d) Give John's and Jane's Analyses of Variance.

(e) State the null and alternative hypotheses for John and Jane to determine whether there is an association between treatment (X) and response (Y).

(f) Conduct your tests in part (e), each at $\alpha = 0.05$. Note, there is no need to adjust for simultaneous tests, as John and Jane are working independently.

(g) Use the F-test for lack of fit to determine whether John's model is appropriate (H_0) or Jane's is (H_A), with $\alpha = 0.05$.

7. A plant foreman is interested in the effects of changes in labor (X_1) and capital (X_2) on the plant's output. She designs an experiment where she varies labor and capital, observing the output. She believes that output should increase linearly over these ranges of labor and capital, and that there is no interaction between these two factors. She fits the model:

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

She obtains the following data:

i	Y_i	X_{i1}	X_{i2}
1	6	2	2
2	9	2	4
3	12	2	6
4	10	5	2
5	15	5	4
6	17	5	6
7	14	8	2
8	15	8	4
9	19	8	6

(a) Give $\boldsymbol{X}'\boldsymbol{X}$, $\boldsymbol{X}'\boldsymbol{Y}$, and $\boldsymbol{Y}'\boldsymbol{Y}$.

(b) The least squares estimate ($\hat{\boldsymbol{\beta}}$), $(\boldsymbol{X}'\boldsymbol{X})^{-1}$, and MS(Residual) are given below. Test whether the true effects of increasing labor by one unit and increasing capital by one unit are the same. Clearly state the null and alternative hypotheses, test statistic, and rejection region (based on $\alpha = 0.05$ significance level).

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 1.1667 \\ 1.1667 \\ 1.5000 \end{bmatrix} \qquad (\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{bmatrix} 1.2407 & -0.0926 & -0.1667 \\ -0.0926 & 0.0185 & 0.0000 \\ -0.1667 & 0.0000 & 0.0417 \end{bmatrix} \qquad MS(\text{Residual}) = 1.4167$$

(c) Give the predicted value, residual, and their variances for the first observation; that is, find $\hat{Y}_1$, e_1, $s^2(\hat{Y}_1)$ and $s^2(e_1)$.

8. While lost in the woods, you find the following partial ANOVA table. The response being measured was the bird's height. The researcher has written "2-factor ANOVA" (which is barely legible next to the sabre-tooth bite marks on the tattered output). The species that (s)he has analyzed are a random sample of species from this region.

Source	df	SS
Species		150.0
Gender	1	20.0
Interaction	5	
Error		25.0
Total	35	200.0

(a) How many birds were there in the sample?

(b) How many species were there in the sample?

(c) What was the number of replicates per species/gender combination?

(d) Find MS(Species), MS(Gender), MS(Interaction) and MSE.

(e) Find the test statistics for testing for interaction, species effects, and gender effects; that is, find $F_{Interaction}$, $F_{Species}$ and F_{Gender}.

(f) Find the rejection regions for the tests in (e).

9. A shoe company executive is interested in comparing her company's (NIKE) top of the line shoe with those of her 2 biggest competitors (ADIDAS, REEBOK). The response she's interested in is the vertical leap (distance off the ground, in inches) for basketball players. Because she knows that there is a large amount of variation in players, she takes a sample of players, treats them as blocks, having each player wear each brand (in random order). She has no reason to believe the effects of the brands differ among the players.

(a) Write out the statistical model for the randomized complete block design, very briefly defining all terms.

(b) She obtains the following sample data. Give the Analysis of Variance. You may use the fact that $\sum_{i=1}^{n}\sum_{j=1}^{r}(Y_{ij} - \overline{Y}_{..})^2 = 660$.

Player	Nike	Adidas	Reebok
Pat	37	31	34
Terry	23	21	19
Kim	28	20	24
Mel	16	15	14
Chris	26	23	29

(c) Test whether the true mean vertical leaps differ among the brands of shoes (use $\alpha = 0.05$). Be sure to explicitly state the hypotheses, test statistic, rejection region and conclusion.

(d) Give the decision rule based on Tukey's and Bonferroni's methods to determine whether the true means differ among all pairs of brands. That is, we should conclude $\mu_i \neq \mu_j$ if what? Be precise with your notation.

10. Consider a linear regression model with 2 independent variables and no intercept:

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

(a) Derive the least squares estimates of β_1 and β_2 in terms of the actual data, (Y_i, X_{i1}, X_{i2}). (Your answer to this question must be a closed form function of the data.)

(b) Give $s^2\{\hat{\beta}_1 - \hat{\beta}_2\}$ in terms of the actual data, (Y_i, X_{i1}, X_{i2}). (Again, your answer must be a closed form function of the data.)