First Year Examination
Department of Statistics, University of Florida

August 20, 2007, 1:00 - 5:00 pm

Instructions:

1. You have four hours to answer questions in this examination.
2. You must show your work to receive credit.
3. **Write only on one side of the paper, and start each question on a new page.**
4. There are 10 problems of which you must answer 8.
5. Only your first 8 problems will be graded.
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. You are allowed to use a calculator.

The following abbreviations and terminology are used throughout:

- ANOVA = analysis of variance
- cdf = cumulative distribution function
- iid = independent and identically distributed
- mgf = moment generating function
- ML = maximum likelihood
- MLR = monotone likelihood ratio
- MOM = method of moments
- MSE = mean squared error
- OLS = ordinary least squares
- pdf = probability density function
- pmf = probability mass function
- UMP = uniformly most powerful
- $\mathbb{Z}^+ = \{0, 1, 2, \dots\}$
- N(μ, σ^2) = normal distribution with mean μ and variance σ^2

You may use the following facts/formulas without proof (except in Problem 7a):

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 $\beta > 0$.

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

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

where $\alpha > 0$ and $\beta > 0$. Also, $\text{E}(X) = \alpha\beta$ and $\text{Var}(X) = \alpha\beta^2$. The mgf is given by $m_X(t) = (1-\beta t)^{-\alpha}$ for $t < 1/\beta$.

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

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

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

Iterated Expectation Formula: $\text{E}(X) = \text{E}\left[\text{E}(X|Y)\right]$.

Iterated Variance Formula: $\text{Var}(X) = \text{E}\left[\text{Var}(X|Y)\right] + \text{Var}\left[\text{E}(X|Y)\right]$.

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 $X/(X+Y) \sim \text{Beta}(\alpha_x, \alpha_y)$.

A Linear Combination of Normals: Let $X_1, \ldots, X_n$ be independent random variables such that $X_i \sim \text{N}(\mu_i, \sigma_i^2)$ for $i = 1, \ldots, n$. If $a_1, \ldots, a_n$ are constants, then the random variable $\sum_{i=1}^{n} a_i X_i$ has a normal distribution.

1. Consider the linear mixed model

$$y_{ijk} = \mu + \alpha_i + \beta_{ij} + \epsilon_{ijk}, \qquad i = 1, \ldots, a \quad j = 1, \ldots, b \quad k = 1, \ldots, n$$

$$\sum_{i=1}^{a} \alpha_i = 0, \qquad \beta_{ij} \sim \mathrm{N}(0, \sigma_\beta^2), \qquad \epsilon_{ijk} \sim \mathrm{N}(0, \sigma_\epsilon^2), \qquad \text{all } \beta_{ij}\text{'s and } \epsilon_{ijk}\text{'s independent}$$

where $a \geq 2$, $b \geq 2$, and $n \geq 2$. The parameters μ, α_i, σ_β^2, and σ_ϵ^2 are assumed to be unknown. Adopt the following notation:

$$\overline{y}_{ij\bullet} = \frac{1}{n}\sum_{k=1}^{n} y_{ijk} \qquad \overline{y}_{i\bullet\bullet} = \frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk} \qquad \overline{y}_{\bullet\bullet\bullet} = \frac{1}{abn}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}$$

(a) In terms of the parameters, find the *correlations* between (i) y_{111} and y_{112}, (ii) y_{111} and y_{121}, and (iii) y_{111} and y_{211}.

(b) For any given value of i, specify the joint distribution of $\overline{y}_{i1\bullet}, \ldots, \overline{y}_{ib\bullet}$.

(c) In terms of the data, write a formula for the usual unbiased estimator of $\alpha_1 - \alpha_2$. What is the exact distribution of this estimator?

(d) Show that

$$\mathrm{E}\left(\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{y}_{ij\bullet} - \overline{y}_{i\bullet\bullet})^2\right) \;=\; a(b-1)\left(\sigma_\beta^2 + \frac{1}{n}\sigma_\epsilon^2\right)$$

Justify all important steps. (Hint: Your answer to part (b) might be useful.)

(e) In terms of the data, write a formula for the usual *unbiased* (ANOVA) estimate of σ_β^2. (Define all new notation, if you use any.)

2. The weights (y_i, kilograms) and corresponding heights (x_i, centimeters) of 10 randomly-sampled adolescents ($i = 1, \ldots, 10$) are recorded, and the following summary statistics are computed:

$$\overline{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 165 \qquad \overline{y} = \frac{1}{10}\sum_{i=1}^{10} y_i = 59$$

$$\sum_{i=1}^{10}(x_i - \overline{x})^2 = 472 \qquad \sum_{i=1}^{10}(y_i - \overline{y})^2 = 731 \qquad \sum_{i=1}^{10}(x_i - \overline{x})(y_i - \overline{y}) = 274$$

You will perform a simple linear regression of weight on height, under the usual assumption of independent, identically distributed, normal errors.

(a) Compute the least squares estimates for the intercept and slope parameters.

(b) Compute the usual unbiased estimate of the error variance.

(c) Compute unbiased estimates of the variances of the least squares estimates in part (a).

(d) Perform a two-sided test for whether or not height and weight are related (assuming the simple linear regression model holds). State the null and alternative hypotheses, and use $\alpha = 0.05$.

(e) Compute 95% *simultaneous* two-sided confidence intervals for the intercept and slope parameters, using the Bonferroni method.

3. Consider a randomized complete block design with 12 blocks and a single treatment factor having 3 levels. Let Y_{ij} denote the response measured for the experimental unit in block j that receives treatment i for $i = 1, 2, 3$ and $j = 1, \ldots, 12$. Suppose there is also a covariate whose value X_{ij} is measured for each experimental unit.

The following four models are fit to the data (using least squares), with the resulting residual (error) sums of squares as specified:

Model 1:	$Y_{ij} = \mu + \gamma_j + \epsilon_{ij}$	$\mathrm{SS(Res)} = 660$
Model 2:	$Y_{ij} = \mu + \alpha_i + \gamma_j + \epsilon_{ij}$	$\mathrm{SS(Res)} = 550$
Model 3:	$Y_{ij} = \mu + \alpha_i + \gamma_j + \beta X_{ij} + \epsilon_{ij}$	$\mathrm{SS(Res)} = 300$
Model 4:	$Y_{ij} = \mu + \gamma_j + \beta X_{ij} + \epsilon_{ij}$	$\mathrm{SS(Res)} = 420$

The treatment effects are $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ and the block effects are $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{12})$.

The corrected total sum of squares is 820.

(a) Find the *sequential* sums of squares for $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$, and β, in that order.

(b) Form an ANOVA table for the randomized complete block design *without* the covariate X_{ij}, that is, based on Model 2. The table should include all appropriate sources of variation (including the corrected total), with degrees of freedom, sums of squares, and mean squares where appropriate. Then test whether or not there is any treatment effect based on this model. Use $\alpha = 0.05$.

(c) Test whether there is any treatment effect, after accounting for both blocking and the covariate. Use $\alpha = 0.05$.

(d) Suppose the (possibly incorrect) model $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ is fit to the data. Compute the residual sum of squares for this model.

4. City planners are evaluating the effectiveness of a new “intelligent” traffic control system in reducing the amount of time motorists must spend on city streets. A total of 24 simulations are run: 4 simulations for each of the 6 combinations of control system (old or new) and traffic intensity (light, moderate, or heavy). All simulations use different random seeds, the combinations are run in a completely random order, and the median travel time (minutes) is recorded for each simulation. For each combination, the following table gives the average and sample *standard deviation* of the median travel times from the 4 simulations assigned that combination:

	Old System			New System		
	Light	Moderate	Heavy	Light	Moderate	Heavy
Sample Mean	13	14	15	5	8	17
Sample Standard Deviation	1	2.5	3.5	2.5	2	3.5

(a) Write a (univariate) linear model equation of the usual full form for data from this experiment, with median travel time as the response. Explain each term and specify any conditions it satisfies. What crucial assumption are you making about the error variances?

(b) Produce an ANOVA table with all appropriate sources of variation, including the (corrected) total. Include sums of squares, degrees of freedom, and appropriate mean squares.

(c) Test whether your model in part (a) may be reduced to a model in which the effects of system and traffic intensity are purely additive. Remember to state the null and alternative hypotheses. Use $\alpha = 0.05$.

(d) Form a two-sided 95% confidence interval for the difference in median travel time between the new system and the old system under moderate traffic conditions.

5. For a series of data $y_t,\ t = 1, \ldots, n$, consider the model

$$y_t = \alpha_0 + \alpha_1 \cos(\pi t/2) + \beta_1 \sin(\pi t/2) + \epsilon_t, \qquad \epsilon_t \sim \text{iid N}(0, \sigma^2), \qquad t = 1, \ldots, n \tag{1}$$

where α_0, α_1, β_1, and σ^2 are unknown parameters, and the arguments of $\cos$ and $\sin$ are in radians (so that $\cos(\pi/2) = 0$, $\sin(\pi/2) = 1$, $\cos(\pi) = -1$, $\sin(\pi) = 0$, and so forth). Suppose $n = 8$.

(a) Is (1) a linear model? Explain.

(b) Write out the full matrix $\boldsymbol{X}$ in the matrix representation $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$ of this data model, where $\boldsymbol{Y} = (y_1, \ldots, y_8)'$, $\boldsymbol{\beta} = (\alpha_0, \alpha_1, \beta_1)'$, and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_8)'$.

(c) The data values are

$$y_1 = 21 \qquad y_2 = 20 \qquad y_3 = 7 \qquad y_4 = 10 \qquad y_5 = 21 \qquad y_6 = 18 \qquad y_7 = 3 \qquad y_8 = 20$$

(i) Compute the OLS estimates of α_0, α_1, and β_1 under model (1).

(ii) Compute the corrected total sum of squares and the residual (error) sum of squares. What are their degrees of freedom?

(iii) Test $\text{H}_0 : \ \alpha_1 = \beta_1 = 0$ versus the alternative $\text{H}_a : \ \alpha_1 \neq 0$ or $\beta_1 \neq 0$. Use $\alpha = 0.05$.

(d) Suppose you tried to fit the higher-degree expansion of the form

$$y_t = \alpha_0 + \alpha_1 \cos(\pi t/2) + \beta_1 \sin(\pi t/2) + \alpha_2 \cos(\pi t) + \beta_2 \sin(\pi t) + \epsilon_t$$

using OLS. What problem would you encounter? Explain specifically.

6. Let $X_1, \ldots, X_n$ be iid random variables from a continuous population with cdf $F(x)$ and pdf $f(x)$.

(a) Derive the cdf of $X_{(j)}$ for $1 \leq j \leq n$ (in terms of F).

(b) Write down the pdfs of $X_{(1)}$ and $X_{(n)}$.

(c) Suppose that $Z \sim \text{Beta}(\alpha, \beta)$. Derive the mean and variance of Z.

Now suppose that X_1 and X_2 are iid Uniform$(0, 1)$; that is, the common pdf is given by

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

(d) Compute the correlation between $X_{(1)}$ and $X_{(2)}$.

7. Let X and Y be two random variables with finite second moments.

(a) Prove that $\text{Var}(X) = \text{Var}\big[\text{E}(X|Y)\big] + \text{E}\big[\text{Var}(X|Y)\big]$.

(b) Use the function $h(t) = \text{E}\Big\{\big[(X - \text{E}X)t + (Y - \text{E}Y)\big]^2\Big\}$ to prove that $-1 \leq \rho_{XY} \leq 1$ where ρ_{XY} denotes the correlation between X and Y.

(c) Prove or disprove the following statement: If $\text{Cov}(X, Y) = 0$, then X and Y are independent.

Let $X_1, X_2, \cdots, X_n$ be random variables with finite second moments.

(d) Show that

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j) \ .$$

where $\sum_{1 \leq i < j \leq n}$ means the sum over all (i, j) pairs such that $1 \leq i \leq n$, $1 \leq j \leq n$ and $i < j$.

8. Suppose that the random variables $Y_1, \ldots, Y_n$ satisfy

$$Y_i = x_i \beta + \varepsilon_i \ , \quad i = 1, \ldots, n$$

where $x_1, \ldots, x_n$ are fixed constants, $\varepsilon_1, \ldots, \varepsilon_n$ are iid $\text{N}(0, \sigma^2)$ and σ^2 is known.

(a) Find the ML estimator of β, call it $\hat{\beta} = \hat{\beta}(Y)$.

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

(c) Find the posterior distribution of β under a normal prior with mean 0 and variance $\tau^2 / \big(\sum_{i=1}^{n} x_i^2\big)$.

(d) Show that the posterior expectation of β, call it $\tilde{\beta} = \tilde{\beta}(Y)$, can be written as a simple function of $\hat{\beta}$.

(e) Find the MSE of $\tilde{\beta}$.

9. 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

$$f(x|\alpha,\beta) = P_{\alpha,\beta}(X = x) = \frac{\Gamma(\alpha + x)}{\Gamma(\alpha)\, x!\, \beta^{\alpha}} \left(\frac{\beta}{\beta+1}\right)^{x+\alpha} I_{\mathbb{Z}^+}(x)\ . \tag{2}$$

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

Suppose that $X_1, \dots, X_n$ are iid with common pmf given by

$$f(x|\theta) = \frac{(x+1)}{\theta^2}\left(\frac{\theta}{\theta+1}\right)^{x+2} I_{\mathbb{Z}^+}(x)\ , \tag{3}$$

where $\theta > 0$.

(c) Find the Cramér-Rao lower bound for the variance of an unbiased estimator of θ.

(d) Find the MOM estimator of θ, call it $\tilde{\theta}(X)$. Is the MOM estimator unbiased? Is it best unbiased?

(e) Find the ML estimator of θ, call it $\hat{\theta}(X)$.

10. Suppose that X_1, X_2, X_3 are iid from the following pmf

$$P_\theta(X = x) = \frac{(x+1)}{\theta^2}\left(\frac{\theta}{\theta+1}\right)^{x+2} I_{\mathbb{Z}^+}(x)\ ,$$

where $\theta > 0$.

(a) Show that $W = X_1 + X_2 + X_3$ is a sufficient statistic for θ.

(b) Derive the Law of Total Probability. More specifically, suppose that S is a sample space, A is a subset of S and $\{B_0, B_1, B_2, \dots\}$ is a partition of S and show that

$$P(A) = \sum_{i=0}^{\infty} P(A|B_i)\, P(B_i)\ .$$

(c) Use the Law of Total Probability to derive the pmf $X_1 + X_2$. (Hint: Don't bother simplifying the sum.)

(d) Use the Law of Total Probability again to derive the pmf of W.

(e) Show that the family of mass functions of W has MLR.

(f) Find a UMP test (based on X_1, X_2, X_3) of $H_0 : \theta \leq 1$ against $H_1 : \theta > 1$ with level $\frac{63}{64}$.