

First Year Examination
Department of Statistics, University of Florida
August 19, 2010, 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. Questions 1 through 5 are the “theory” questions and questions 6 through 10 are the “applied” questions. You must do exactly four of the theory questions and exactly four of the applied questions
4. **Write your answers to the theory questions on the blank paper provided. Write only on one side of the paper, and start each question on a new page.**
5. **Write your answers to the applied questions on the exam itself. If you use extra sheets, write only on one side of the paper, and start each question on a new page.**
6. While the 10 questions are equally weighted, some questions 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:

- cdf = cumulative distribution function
- iid = independent and identically distributed
- mgf = moment generating function
- ML = maximum likelihood
- MSE = mean squared error
- pdf = probability density function
- pmf = probability mass function
- $\mathbb{R}^+ = (0, \infty)$

You may use the following facts/formulas without proof:

Iterated Expectation Formula: $E(X) = E[E(X|Y)]$.

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

1. Three balls are distributed randomly into three urns. Define the random variable X to be the number of empty urns, and define the random variable Y to be the number of balls in the third urn.

- Find the joint pmf of X and Y .
- Are X and Y independent? (A yes/no answer is not sufficient here.)
- Find the conditional pmf of Y given $X = 0$
- Find the conditional expectation and variance of X given $Y = 0$.

2. Let (X, Y) be a continuous bivariate random vector with joint pdf given by

$$f(x, y) = \begin{cases} \frac{15}{16}(x^2 + y^3) & -1 < x < y < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

- Find $P(X + Y < 0)$.
- Find the marginal pdf of X .
- Find the conditional pdf of X given that $Y = y$.
- Find $P(Y > 1/4 \mid X > 1/2)$. (Just set up the integral(s) - do not compute them.)
- Find another joint pdf that has the same marginals as $f(x, y)$.

3. Suppose that $Y \sim N(\theta, 1)$, and that the prior density for θ , call it $\pi(\theta)$, is $N(\mu, \tau^2)$.

- Find the posterior density of θ given y , call it $\pi(\theta|y)$.
- Find the Bayes estimator of θ , call it $\tilde{\theta}(Y)$, and show that it can be written as a weighted average of the ML estimator of θ , which is $\hat{\theta}(Y) = Y$, and the prior mean of θ .
- Find the MSE of $\tilde{\theta}(Y)$ and the MSE of $\hat{\theta}(Y)$ (as estimators of θ).
- Does either of these estimators dominate the other in terms of MSE? In other words, is one of the MSE functions uniformly below the other?

Now consider an alternative prior that is a mixture of k normal densities; that is,

$$\pi^*(\theta) = \sum_{i=1}^k \frac{p_i}{\sqrt{2\pi\tau_i^2}} \exp\left\{-\frac{1}{2\tau_i^2}(\theta - \mu_i)^2\right\},$$

where each $p_i \in (0, 1)$, $\sum_{i=1}^k p_i = 1$, each $\mu_i \in \mathbb{R}$ and each $\tau_i^2 \in \mathbb{R}^+$. Note that $\pi^*(\theta) = \sum_{i=1}^k p_i \pi_i(\theta)$, where $\pi_i(\theta)$ is a $N(\mu_i, \tau_i^2)$ density.

- Find the exact posterior density of θ given y , call it $\pi^*(\theta|y)$. (Hint: You may use the fact that the marginal density of y under $\pi(\theta)$ is $N(\mu, \tau^2 + 1)$.)
- Is $\pi^*(\theta)$ a conjugate prior for the normal mean? (A yes/no answer is not sufficient here.)

4. Suppose that $n \geq 2$ and that X_1, \dots, X_n are iid Bernoulli(p).
- Derive the mgf of X_1 and use it to calculate the expectation and variance of X_1 .
 - Find the ML estimator of p . Is it unbiased?
 - Find the best unbiased estimator of p .
 - Find the ML estimator of $p(1 - p)$. Is it unbiased?
 - Find the Cramér-Rao Lower Bound for the variance of an unbiased estimator of $p(1 - p)$.
 - The sample variance, $\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$, is an unbiased estimator of $p(1 - p)$. Is it best unbiased?

5. 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:

Test I: Reject H_0 if $X_1 > .95$

Test II: Reject H_0 if $X_1 + X_2 > C$

where $C > 0$.

- Show that $X_1 - \theta$ is Uniform($0, 1$).
- Derive the power function of Test I; that is, find $\beta_1(\theta) = P_\theta(X_1 > .95)$ for $\theta \geq 0$.
- Let U_1, U_2 be iid Uniform($0, 1$). Find the cdf of $U_1 + U_2$.
- Derive the power function of Test II; that is, find $\beta_2(\theta) = P_\theta(X_1 + X_2 > C)$ for $\theta \geq 0$.
- What value of C should we use if we want the two tests to have the same size? With this value of C , is Test II more powerful than Test I?
- Explain how we could increase the power of Test II without affecting its size.

6. An experiment is conducted to compare the effects of 3 brands of basketball shoes on jumping ability among high school basketball players (these are the only 3 brands of interest to the experimenter). A random sample of 18 high school basketball players is selected from public high schools in a large city. Each of the players jumps in each of the brands (in an order such that each brand is measured 1st, 2nd, and 3rd on 6 jumpers). The researcher does not believe there will be an order effect, so she does not include that in the model. The model is made up of the following components:

- Y_{ij} = Jump height (inches) when player j wears brand i ($i=1,2,3; j=1, \dots, 18$)
- μ = Overall mean among all high school players among these 3 brands
- α_i = Effect of brand i (fixed)
- β_j = Effect of player j (random)
- ε_{ij} = Random Error Term

Model: $Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$

where: $\sum_{i=1}^3 \alpha_i = 0$ $\beta_j \sim NID(0, \sigma_\beta^2)$ $\varepsilon_{ij} \sim NID(0, \sigma_\varepsilon^2)$ $COV(\beta_j, \varepsilon_{ij'}) = 0 \forall i, j, j'$

Derive the following terms:

- $V(Y_{ij})$
- $V(Y_{i\bullet})$ where $Y_{i\bullet} = \sum_j Y_{ij}$
- $V(\bar{Y}_{i\bullet})$ where $\bar{Y}_{i\bullet} = \frac{1}{18} \sum_j Y_{ij}$
- $COV(\bar{Y}_{i\bullet}, \bar{Y}_{i'\bullet})$ $i \neq i'$
- $V(\bar{Y}_{i\bullet} - \bar{Y}_{i'\bullet})$ $i \neq i'$

7. An experiment is conducted to measure the effects of 3 types of paint and 2 types of wood on the brightness of the paint (measured by a brightness meter). The experimenter conducts the experiment in 2 replicates of each combination of paint and wood, and runs the analysis as a multiple regression with two dummy variables for paint type, and 1 dummy variable for wood type, and an intercept term.

We fit the following sequence of models (assume the Total sum of squares = 1000):

$$\text{Where: } X_1 = \begin{cases} 1 & \text{Paint Type 1} \\ 0 & \text{otherwise} \end{cases} \quad X_2 = \begin{cases} 1 & \text{Paint Type 2} \\ 0 & \text{otherwise} \end{cases} \quad X_3 = \begin{cases} 1 & \text{Wood Type 1} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Model 1: } Y = \alpha_0 + \varepsilon \quad \varepsilon \sim NID(0, \sigma^2)$$

$$\text{Model 2: } Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \quad \varepsilon \sim NID(0, \sigma^2)$$

$$\text{Model 3: } Y = \delta_0 + \delta_1 X_1 + \delta_2 X_2 + \delta_3 X_3 + \varepsilon \quad \varepsilon \sim NID(0, \sigma^2)$$

$$\text{Model 4: } Y = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 + \gamma_4 X_1 X_3 + \gamma_5 X_2 X_3 + \varepsilon \quad \varepsilon \sim NID(0, \sigma^2)$$

$$R_1^2 = 0.00 \quad R_2^2 = 0.65 \quad R_3^2 = 0.72 \quad R_4^2 = 0.75$$

Complete the following parts (for tests give null and alternative hypotheses in terms of model parameters, the test statistic, and rejection region based on $\alpha = 0.05$ significance level):

- For each model, Give the matrix $\mathbf{X}'\mathbf{X}$
- Test whether there is a main effect for paint type (ignoring wood type and interaction)
- Test whether there is a main effect for wood type (controlling for paint type but ignoring interaction).
- Test whether there is an interaction between paint and wood types (after controlling for main effects).
- Give the analysis of variance table with degrees of freedom and sums of squares for paint, wood, interaction, and error (using the sequential sums of squares).

8. An experiment is conducted to compare the average breaking strengths of 4 types of steel bars. The following table gives the mean, standard deviations and sample sizes (number of replicates) for measurements made on each of the 4 steel types (these are the only types of steel of interest).

Type	Mean	SD	# reps
1	26	3	5
2	22	4	6
3	30	4	5
4	23	4	6

Complete the following parts:

- Compute the between types (treatment) sum of squares and degrees of freedom
- Compute the within types (error) sum of squares and degrees of freedom
- Give the Analysis of Variance table
- Test whether the population means differ among the 4 steel types ($\alpha = 0.05$ significance level).
- Compute the Bonferroni Minimum significant difference for each of the pairs of means with an experimentwise error rate of 0.05.

9. Use the following output to obtain the quantities given below:

X		
1	0	4
1	4	4
1	8	4
1	0	8
1	4	8
1	8	8

Y
14
16
21
17
20
24

(X'X) ⁻¹		
1.9167	-0.0625	-0.2500
-0.0625	0.0156	0.0000
-0.2500	0.0000	0.0417

X'Y
112
504
692

Beta-hat
10.1667
0.8750
0.8333

P

0.5833	0.3333	0.0833	0.2500	0.0000	-0.2500
0.3333	0.3333	0.3333	0.0000	0.0000	0.0000
0.0833	0.3333	0.5833	-0.2500	0.0000	0.2500
0.2500	0.0000	-0.2500	0.5833	0.3333	0.0833
0.0000	0.0000	0.0000	0.3333	0.3333	0.3333
-0.2500	0.0000	0.2500	0.0833	0.3333	0.5833

Y'Y	2158.00
Y'PY	2156.33
Y'(I-P)Y	1.67
Y'(J/n)Y	2090.67
Y'(P-J/n)Y	65.67

Total Corrected: Sum Of Squares _____ Degrees of Freedom _____

Regression: Sum Of Squares _____ Degrees of Freedom _____

Residual: Sum Of Squares _____ Degrees of Freedom _____

S^2 _____ $s\{\hat{\beta}_1\}$ _____

Testing $H_0: \beta_1 = \beta_2 = 0$ F-stat _____ Num df _____ Den df _____

Rejection Region ($\alpha = 0.05$) _____

Predicted Value for Y_2 : Based on: (SHOW CALCULATIONS for each method)

$X\hat{\beta}$ _____

PY _____

$s\{\hat{Y}_2\} =$ _____ $s\{e_2\} =$ _____

10. Consider the simple linear regression model in scalar form:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad \varepsilon_i \sim iid(E(\varepsilon_i) = 0, V(\varepsilon_i) = \sigma^2) \quad i = 1, \dots, n$$

Complete the following parts (show all work):

- a) Derive the least squares estimators of β_1, β_0
- b) Derive the mean and variance of the least squares estimator of β_1
- c) Derive the covariance of the least squares estimators of β_1, β_0