APPLIED MATHEMATICS 166   SAMPLE EXAM PROBLEMS

May 2, 2003

1. The article "Prediction, linear regression and minimum sum of relative errors" by S.C. Narula and J.F. Wellington in *Technometrics*, Vol 19, 1977, studies the predictive value of several variables for determining the selling price of a home. Independent variables such as number of rooms, number of bathrooms, lot size, number of bedrooms, and living space are considered. A simple linear regression of PROPERTY VALUE vs LIVING AREA was performed. Both variables are in units of thousands—"dollars" for PROPERTY VALUE and "square feet" for LIVING AREA. Partial results are:

$$X^t X = \begin{pmatrix} 24.00000000 & 33.20600000 \\ 33.20600000 & 47.69869600 \end{pmatrix},$$

and

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F-VALUE	$\Pr > F$
MODEL	1	418.10361719	418.10361719	22.38	0.0001
ERROR	22	410.94263281	18.67921058		
CORRECTED TOTAL	23	829.04625000			

Coefficients:

	Value	Std. Error	t value	$\Pr(> |t|)$
(Intercept)		4.59870341	2.88	
LIVING_AREA	15.43300413	3.26202863		

a) What is the least squares estimate of β_0?

b) Test the null hypothesis $H_0 : \beta_1 = 0$ against the *one-sided* alternative $H_1 : \beta_1 > 0$. Test the hyhothesis at level $\alpha = 0.01$ and explain your choice of test criterion carefully.

c) Construct a 95% confidence interval for β_1.

d) Find an 80% confidence interval for the error-variance σ^2.

2. The following model was proposed for testing whether there was evidence of salary discrimination against women faculty members in a state university system:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_2^2 + \epsilon$$

where $Y = $ annual salary (in thousands of dollars)

$$x_1 = \begin{cases} 1, & \text{if female} \\ 0, & \text{if male} \end{cases}$$

and $x_2 = $ amount of experience (in years).

When this model was fit to data obtained from the records of 200 faculty members, SSE $= 783.90$. The recuced model $Y = \gamma_0 + \gamma_1 x_2 + \gamma_2 x_2^2 + \epsilon$ was also fit and produced a value of SSE $= 795.23$. Do the data provide sufficient evidence to support the claim that the mean salary depends on the gender of the faculty member? Use $\alpha = 0.05$.

3. Consider the following regression model (which has no constant term β_0):

$$Y_i = \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$$

for $i = 1, \ldots, n$. Assume that the ϵ_i are uncorrelated and $\mathrm{Var}(\epsilon_i) = \sigma^2$.

a) Derive the least squares estimates of β_1 and β_2.

b) Assuming that the ϵ_i are independent normal random variables, state the likelihood function and obtain the maximum likelihood estimators of β_1 and β_2. How do these latter estimators relate to the least-squares estimates of part (a)?

4. a) In an experiment to investigate the performance of four different brands of spark plugs intended for use on a 125-cc two-stroke motorcycle, five plugs of each brand were tested and the number of miles (at a constant speed) until failure was observed. The partial ANOVA table for the data appears below. Fill in the missing entries, state the relevant hypotheses, and carry out a test.

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F-VALUE	Pr > F
MODEL					
ERROR			14,713.69		
CORRECTED TOTAL		310,500.76			

4. b) In a study to assess the effects of malaria infection on mosquito hosts ("Plasmodium Cynomolgi: Effects of Malaria Infection on Laboratory Flight Performance of Anopheles Stephensi Mosquitos," *Experimental Parasitology*, 1977, pp. 397–404) mosquitos were fed on either infective or noninfective rhesus monkeys. Subsequently, the distance they flew during a 24-hour period was measured using a flight mill. The mosquitos were divided into four groups of eight mosquitos each: infective rhesus and sporozites present (IRS); infective rhesis and oocysts present (IRD); infective rhesus and no infection developed (IRN); and noninfective (C). The summary data are: $\bar{Y}_{1.} = 4.39$ (IRS), $\bar{Y}_{2.} = 4.52$ (IRD), $\bar{Y}_{3.} = 5.49$ (IRN), $\bar{Y}_{4.} = 6.36$ (C), $\bar{Y}_{..} = 5.19$, and $\sum\sum Y_{ij}^2 = 911.91$. Carry out an ANOVA F-test at level $\alpha = 0.05$ to decide whether there are any differences between true average flight times for the four treatments.