M. S. Comprehensive Exam
Thursday, August 21, 2008

You have three hours. Do as many problems as you can. No one is expected to answer all the problems correctly. Partial credit will be given. All problems are worth an equal amount of credit.

Put your solution to each problem on a separate sheet of paper.

Problem 1. Answer the following questions:

(a) We have a linear regression problem with n observations

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i, i = 1, ..., n$$

which can be written in matrix form as $Y = X\beta + \epsilon$. Assuming that X, Y are given matrices of real numbers and ϵ_i are i.i.d. from $\mathcal{N}(0, \sigma^2)$, find the least squares estimators of β and derive their distribution.

(b) What is a graphical tool for checking the normality of the residuals?

(c) To find outliers, jackknife residuals (named studentized residuals in SAS) are computed for each observation and compared to a cutoff. What type of outliers can**not** be detected using this method?

Problem 2. The following regression situation is observed:

$$Y_{ij} = \beta_0 + \beta_1 X_i + \varepsilon_{ij}, \quad i = 1, 2, \ldots, l, \quad j = 1, 2, \ldots n_i.$$

The errors are independent normal $(0, \sigma^2)$ random variables. We wish to see if this model adequately fits the data by performing a lack of fit test. The collected data is shown in the table below. The lack of fit test is performed using the sum of squares pure error

X Level	Responses
X_1	Y_{11}, Y_{12}, Y_{13}
X_2	Y_{21}
X_3	Y_{31}, Y_{32}
X_4	Y_{41}, Y_{42}, Y_{43}
X_5	$Y_{51}, Y_{52}, Y_{53}, Y_{54}$

(SSPE) and the sum of squares lack of fit (SSLOF).

(a) Prove that the sum of squared errors (SSE) from fitting the model to the data is equal to $SSPE + SSLOF$. For simplicity, use notation $Y_{i\cdot} = \sum_j Y_{ij}$, $\overline{Y}_{i\cdot} = n_i^{-1} \sum_j Y_{ij}$ and $\hat{Y}_{ij}$ as the fitted value.

(b) If $SSPE = 800$ and $SSLOF = 6000$, perform the lack of fit test at significance level $\alpha = 0.05$. State the hypotheses, the test statistic and its numeric value, distribution of the test statistic, degrees of freedom of the test statistic, and the conclusion.

Problem 3. Consider a 2^{6-3} fractional factorial design in Factors A, B, C, D, E, and F. An alternate fraction is used (not the principal fraction) with the extra factor assignments as $D = -AB$, $E = AC$, $F = -BC$.

(a) Specify the defining relation.

(b) Identify all main effect and two-factor interaction aliases. Show all work.

(c) Suppose you desired to run a second design. This second design is fold-over of the above design produced switching all of its signs. What would be the generators of this second design?

(d) Suppose you create a third design which is produced by placing the first and second designs above together (you may ignore any kind of blocking). What would be the defining relation of this third design? Also comment on the resolution of this third design.

Problem 4. In any given year a male automobile policyholder will make a claim with probability 0.3, and a female policyholder will make a claim with probability 0.2. The fraction of the policyholders that are male is 0.6. A policyholder is randomly chosen.

(a) What is the probability that the policyholder will make a claim in year 1?

(b) If the policyholder made a claim in the year 1, what is the probability that the policyholder is a male?

(c) What is the probability that the policyholder will make a claim in year 2, given that he/she makes a claim in year 1?

Problem 5. Let X be a non-negative, continuous random variable.

(a) Show that

$$E(X) = \int_0^\infty P(X > t)dt.$$

Hint: Show that $\int_0^\infty P(X > t)dt = \int_0^\infty xf(x)dx$, by changing order of integration after observing that any positive x can be written as $x = \int_0^x dt$.

(b) Use the result stated above in (a) to show that for any non-negative and continuous random variable Y we have

$$E(Y^n) = \int_0^\infty n y^{n-1} P(Y > y)dy.$$

Hint: Apply the result in (a) to $X = Y^n$ and make the change of variables $t = y^n$.

Problem 6. If $Y_n = \min_{1\le i\le n} X_i$, where $X_1, X_2, \ldots$ are i.i.d. r.v.'s with density $f(x) = \beta x^{\beta-1}$, $0 < x < 1$:

(a) What is the distribution of Y_n?

(b) Find a value of β so that $\sqrt{n}\, Y_n$ converges in distribution. What is the limiting distribution?

Problem 7. Let $X_1, \ldots, X_n$ be a random sample from the density

$$f(x \mid \mu, \sigma) = \frac{1}{\sigma} e^{(x-\mu)/\sigma}, \quad -\infty < x < \mu, \ \sigma > 0\,.$$

Find a two-dimensional sufficient statistic for (μ, σ).

Problem 8. Let $X_1, \ldots, X_n$ be i.i.d. from the density

$$f(x|\theta) = \frac{x}{\theta^2} \exp\left(-\frac{x^2}{2\theta^2}\right), \quad x \ge 0$$

with parameter $\theta > 0$.

(a) Find the method of moments estimator of θ.

(b) Find the maximum likelihood estimator (mle) of θ.

(c) Find the asymptotic variance of the mle.

Problem 9. Let $X_1, X_2, \ldots, X_n$ be a random sample of gamma $(\alpha, 1)$ random variables:

$$f(x_i) = \frac{1}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-x_i}, \quad \alpha > 0, \ \ x_i > 0, \ \ i = 1, 2, \ldots, n.$$

Find the maximum likelihood estimate for $\Gamma'(\alpha)/\Gamma(\alpha)$. (Recall that $\lim_{\alpha\to 0}\Gamma(\alpha) = \lim_{\alpha\to\infty}\Gamma(\alpha) = \infty$.)