M. S. Comprehensive Exam
Friday, August 25, 2006

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. In a regression setting, there are $p-1$ explanatory variables $X_1, X_2, \ldots, X_{p-1}$ and n observations. Let $\mathrm{SSE}(X_{i_1}, X_{i_2}, \ldots, X_{i_k}|R)$ be the sum of squares error when regressing some response function R on $k \leq p-1$ variables and the intercept. For example, the error when regressing Y on the intercept and X_1 is $\mathrm{SSE}(X_1|Y)$. For each of the following hypotheses, give the full and reduced model; the test statistic using SSE notation; and the distribution of the test statistic.

(a) $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$, H_a : not H_0

(b) $H_0 : \beta_1 = \beta_2$, $H_a : \beta_1 \neq \beta_2$

(c) $H_0 : \beta_1 = 2, \beta_2 = -3$, H_a : not H_0

Problem 2. Suppose it is desired to conduct a fractional factorial experiment involving five factors (A, B, C, D, and E) each set at two levels. An eight run, one-fourth fraction is selected, that is, a 2^{5-2} design with generators $I = ABD$ and $I = ACE$.

(a) Construct the appropriate design for the above situation indicating how each of the of the factors should be set for each of the eight runs. (Use a '–' for a low setting and a '+' for a high setting.)

(b) Show how each main effect and two-factor interaction is confounded.

(c) Give the resolution of this design and provide a brief explanation on how it is determined.

Problem 3. 100 patients were randomly assigned to 5 methods of treating breast cancer. A numerical score measuring the recovery of each patient was calculated. The table below summarizes the recovery scores for each method:

Group	Method	n	mean	s.d.
1	chemotherapy	20	30.57	7.06
2	surgery and radiation	20	31.17	7.49
3	radiation and chemotherapy	20	26.47	6.82
4	hormonal therapy and radiation	20	29.48	7.63
5	surgery and chemotherapy	20	30.82	7.30

We assume the model $Y_{ij} = \tau_i + \epsilon_{ij}$, $i = 1, \ldots, 5$, $j = 1, \ldots, 20$, where $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$. Here Y_{ij} is the observed response for the jth subject in the ith group, τ_i is the fixed effect for the ith group, and ϵ_{ij} is the error in the measurement for the jth subject in the ith group.

(a) Identify the weights for a linear contrast comparing the methods with chemotherapy to those methods without chemotherapy.

(b) Estimate the contrast and compute a 95 percent confidence interval for the contrast. Would you reject the null hypothesis that the contrast is zero?

(c) Explain what rejecting the hypothesis that the contrast is zero tells us about the group means. Explain what a failure to reject that hypothesis (i.e., accepting the hypothesis that the contrast is zero) tells us about the group means.

(d) A researcher notices that Group 2 has the highest mean and Group 3 has the lowest mean. To explore he carries out a two-sample test and finds the group 2 population mean is significantly higher than the group 3 population mean at the usual .05 error rate. Explain why the researcher's examining the data first before carrying out a test causes an interpretation problem.

Problem 4. Let X be distributed as a Poisson random variable with mean λ.

(a) Let g be any function such that $Eg(X)$ is finite. Show that

$$E(\lambda g(X)) = E(Xg(X-1)).$$

(b) Use the result of part (a) to find $E(X^3)$.

Problem 5. Let X be an exponential random variable with mean 1. Find the expected value of the distance between X and the nearest integer to X.

Problem 6. Ninety-six percent of all babies survive delivery. However, 20 percent of all births involve Cesarean (C) sections, and when a C section is performed the baby survives 90 percent of the time. If a randomly chosen pregnant woman does **not** have a C section, what is the probability that her baby survives ?

Problem 7. Let $X_1, \ldots, X_n$ be i.i.d. random variables with common expected value μ and $\sigma^2 = 1$. Find n such that we can be 90 percent certain that $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ is an accurate estimate of μ, within an error margin of plus/minus 0.25 units. State the approximation result you are using.

Problem 8. Let $X_1, \ldots, X_n$ be independent trials with probability of success $p \in (0, 1/2]$.

(a) Compute the maximum likelihood estimator of p.

Define

$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i \quad \text{and} \quad \widehat{p} \; = \; \begin{cases} \bar{X} & \text{if } \bar{X} < 1/2, \\ 1/2 & \text{if } \bar{X} \geq 1/2. \end{cases}$$

(b) Is $\widehat{p}$ biased?

(c) Show that the mean squared error of $\widehat{p}$ is smaller than the mean squared error of $\bar{X}$.

(d) Is $\widehat{p}$ consistent?

Problem 9.

(a) Let $X_1, X_2, \ldots, X_n$ be iid with density

$$f(x \,|\, \theta) = \begin{cases} e^{-(x-\theta)} & \text{for } x \geq \theta \\ 0 & \text{otherwise.} \end{cases}$$

Find the method of moments (MOM) estimator for θ.

(b) Let $X_1, X_2, \ldots, X_n$ be iid with density

$$f(x \,|\, \theta, \psi) = \begin{cases} \frac{1}{\psi} e^{-(x-\theta)/\psi} & \text{for } x \geq \theta \\ 0 & \text{otherwise} \end{cases}$$

where $\psi > 0$. Find the method of moments (MOM) estimator for the pair (θ, ψ).

Hint for both parts: You may use without proof the fact that the exponential distribution with density $\frac{1}{\beta}e^{-x/\beta}$ for $x \geq 0$ has mean β and variance β^2.