Ph. D. Qualifying Exam
Saturday, August 21, 2004

Please submit solutions to at most **seven** problems. You have four hours. 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.

Applied Statistics

Problem 1. Consider the following fractional factorial experimental design:

	d_1						
Run	1	2	3	4=12	5=13	6=23	7=123
1	−	−	−	+	+	+	−
2	−	−	+	+	−	−	+
3	−	+	−	−	+	−	+
4	−	+	+	−	−	+	−
5	+	−	−	−	−	+	+
6	+	−	+	−	+	−	−
7	+	+	−	+	−	−	−
8	+	+	+	+	+	+	+
	d_2						
Run	−1	−2	−3	−4	−5	−6	−7
9	+	+	+	−	−	−	+
10	+	+	−	−	+	+	−
11	+	−	+	+	−	+	−
12	+	−	−	+	+	−	+
13	−	+	+	+	+	−	−
14	−	+	−	+	−	+	+
15	−	−	+	−	+	+	+
16	−	−	−	−	−	−	−

(a) What is the defining relation of design d_1? What is the resolution of design d_1? List the aliases of main effects 1 and 2 in this design.

(b) Design d_2 is obtained by "folding over" design d_1. It can be shown that d_2 has generators $4 = -12$, $5 = -13$, $6 = -23$, and $7 = 123$. What is the defining relation of design d_2? What is the resolution of design d_2? List the aliases of main effects 1 and 2 in design d_2.

(c) Denote by d the augmented design consisting of d_1 and d_2. The augmented design d has generators $5 = 234$, $6 = 134$, and $7 = 123$. What is the defining relation of design d? What is the resolution of design d? List the aliases of main effects 1 and 2 in design d.

Problem 2. In a complete factorial experiment, consider the following random-effects model:

$$
\begin{aligned}
Y_{ijk} &= \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}, \\
&\qquad i = 1, \ldots, a; \quad j = 1, \ldots, b; \quad k = 1, \ldots, r,
\end{aligned}
$$

where μ is the overall mean, the random effects α_i, β_j, and $(\alpha\beta)_{ij}$ are assumed to be independent and normally distributed with means of zero and variances σ^2_α, σ^2_β, and $\sigma^2_{\alpha\beta}$, respectively. The random errors $\{\epsilon_{ijk}\}$ are assumed to be independent and normally distributed with mean zero and variance σ^2. Furthermore, $\{\alpha_i\}$, $\{\beta_j\}$, $\{(\alpha\beta)_{ij}\}$, and $\{\epsilon_{ijk}\}$ are assumed to be independent of one another.

(a) What are the distributions of $\bar{Y}_{ij\cdot}$, $\bar{Y}_{i\cdot\cdot}$, and $\bar{Y}_{\cdot j\cdot}$? Find the covariance $\mathrm{Cov}(\bar{Y}_{ij\cdot}, \bar{Y}_{uv\cdot})$.

(b) Find unbiased estimates for the variances $\sigma^2_{\alpha\beta}$ and σ^2 based on the observations $\{Y_{ijk}, \quad i = 1, \ldots, a; \;\; j = 1, \ldots, b; \;\; k = 1, \ldots, , r\}$.

Problem 3. In a non-linear regression analysis, the true model has the form:

$$y_i = f(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}) + \epsilon_i$$

where $f(\cdot)$ is an unknown function and $\{\epsilon_i\}$ are iid random variables. Suppose that instead of the true model, a linear model with the form

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + \epsilon_i$$

is fitted to the data and the residuals $\{\hat{\epsilon}_i\}$ are calculated.

(a) What would a plot of $\hat{y}_i$ versus $\hat{\epsilon}_i$ look like in this problem?

(b) Give two ways to improve your fitting and state your justifications.

Problem 4. An experiment analyzes imperfection rates for two processes A and B used in fabricating wafers for computer chips. Let Y_1 and Y_2 be the numbers of imperfections for Processes A and B, respectively. Y_1 and Y_2 are independent Poisson random variables with means μ_A and μ_B, respectively.

(a) Suppose a log-linear model $\log(\mu) = \alpha + \beta x$ is fitted with $x = 1$ for treatment B and $x = 0$ for treatment A. Show that $\exp(\beta) = \mu_B/\mu_A$. Suggest a test for testing the null hypothesis $H_0: \;\; \mu_A = \mu_B$.

(b) Show that $Y = Y_1 + Y_2$ is Poisson with mean $\mu_A + \mu_B$. Conditionally, show that $Y_1|(Y = n)$ is binomial with $\pi = \mu_A/(\mu_A + \mu_B)$.

(c) Based on the result in (b), suggest a test for testing the null hypothesis $H_0: \;\; \mu_A = \mu_B$.

Problem 5. Let N be a positive integer-valued random variable and let $X_1, X_2, \ldots$ be random variables with finite expectations.

(a) Define $X_N(\omega) = X_{N(\omega)}(\omega)$. Show that X_N is a random variable.

(b) Suppose that N is bounded and let $\{N_n,\ n = 1, 2, \ldots\}$ be a sequence of positive integer-valued random variables such that $N_n \nearrow N$ almost surely. Show that

$$\lim_{n\to\infty} E(X_{N_n}) = E(X_N).$$

Problem 6. Let $X_1, X_2, \ldots$ be identically distributed random variables. Show that

(a) $\dfrac{X_n}{n} \to 0$ in probability.

(b) If the random variables have finite mean, then $\dfrac{X_n}{n} \to 0$ almost surely.

(c) Part (b) is false if the condition that the random variables have finite mean is omitted.

Problem 7. Let P_0, P_1 be two probability measures on $(\mathcal{X}, \mathcal{B})$. They are said to be singular with respect to one another if there is a set $A \in \mathcal{B}$ such that $P_0(A) = 1, P_1(A) = 0$, which will also mean that $P_0(A^c) = 0$, $P_1(A^c) = 1$.

(a) Suppose that there is a sequence of random variables $\{T_n\}$ on $(\mathcal{X}, \mathcal{B})$ such that

$$T_n \to 0 \ \ a.s. \text{ under } P_0\,, \text{ and}$$
$$T_n \to 1 \ \ a.s. \text{ under } P_1.$$

Show that P_0 and P_1 are singular with respect to one another.

(b) Show that the same conclusion is true if

$$T_n \to 0 \ \text{ in probability under } P_0\,, \text{ and}$$
$$T_n \to 1 \ \text{ in probability under } P_1.$$

Problem 8. Consider the linear model $Y = X\beta + e$, where $e \sim N(0, \sigma^2 I)$. Let M be the perpendicular projection operator onto $C(X)$, the range or column space of X. Define $\hat{e} = (I - M)Y$.

(a) Find $\text{Cov}(\hat{e}, MY)$

(b) Find $E(\hat{e}^T \hat{e})$

(c) Consider the following particular regression model:

$$Y_1 = \beta_0 + \beta_1 + e_1\,, \quad Y_2 = \beta_0 + \beta_1 + e_2\,, \quad Y_3 = \beta_0 + \beta_2 + e_3\,, \quad Y_4 = \beta_0 + \beta_2 + e_4\,.$$

Show that the parameter $\beta_1 - \beta_2$ is estimable.

Problem 9. Let $X_1, X_2, \ldots, X_n$ be a random sample from the density (pdf)

$$f(x \mid \theta) = \begin{cases} \theta/x^2 & \text{for } x \geq \theta \\ 0 & \text{for } x < \theta \end{cases}$$

where $\theta > 0$.

(a) Find the MLE of θ.

(b) Suppose θ has a prior distribution with density $\pi(\theta) = 1/\theta^2$ for $\theta \geq 1$ (and $\pi(\theta) = 0$ for $\theta < 1$). Find the posterior mean of θ. Assume $n \geq 2$.

Problem 10. Let $X_1, X_2, \ldots, X_n$ be iid $N(\theta, \theta^2)$ where $\theta > 0$.

(a) What is the Cramér-Rao lower bound for the variance of an unbiased estimator of θ? For an unbiased estimator of θ^2?

(b) Use your answers to part (a) to show that $\bar{x}$ and s^2 (the sample mean and variance) are **not** asymptotically efficient estimators of θ and θ^2, respectively.

Hint: Depending on how you do the problem, you may find the following facts to be useful. If $Y \sim N(\mu, \sigma^2)$, then $E(Y - \mu)^4 = 3\sigma^4$. If $U \sim \chi^2_k$, then $\text{Var}(U) = 2k$.

Problem 11. The results of a **cohort** study examining the relationship between gender and coronary heart disease (CHD) is summarized in the following table. In this study, 4,541 individuals were followed 10 years and it was determined which of these individuals actually experienced CHD. For example, 236 out of 2,009 males developed the disease, 1,773 males did not develop the disease, etc.

		CHD		
		Yes	No	Total
Male	Yes	236	1,773	2,009
	No	142	2,390	2,532
	Total	378	4,163	4,541

Provide definitions and results for each of the following statistics. (**You do not have to actually do the calculations, just give the correct definitions and put the right numbers into a formula.)**

(a) The relative risk of CHD for males.

(b) The odds ratio of CHD for males.

(c) Recall that there are several types (definitions) of attributable risk. Define any one and provide the correct estimator .

Problem 12. The following data are the results of seven placebo-controlled randomized studies of the effect of aspirin in preventing death after myocardial infarction. Each row of the table provides data from a single clinical trial. For example, the MRC-1 study randomized 615 patients to receive aspirin treatment and 624 patients to receive placebo treatment. In the MRC-1 study, 49 deaths occurred in the patients receiving aspirin treatment and 67 deaths occurred among the patients receiving placebo.

	Aspirin		Placebo	
	Status at Study End		Status at Study End	
Study	Alive	Dead	Alive	Dead
MRC-1	566	49	557	67
CDP	714	44	707	64
MRC-2	730	102	724	126
GASP	285	32	271	38
PARIS	725	85	354	52
AMIS	2021	246	2038	219
ISIS-2	7017	1570	6880	1720

Explain the steps you would go through to produce a meta-analysis of these results. (**Again, you do not have to actually do any calculations.)**