Department of Statistics
University of Wisconsin-Madison
PhD Qualifying Exam Part II
February 1, 2007
1:00–4:00pm, Room 133 SMI

- There are a total of FOUR (4) problems in this exam. Please do a total of TWO (2) problems.
- Each problem must be done in a separate exam book.
- Please turn in TWO (2) exam books.
- Please write your code name and **NOT** your real name on each exam book.

1. **Definitions:** Let $(\Omega, \mathcal{F}, P)$ be a probability space and suppose that $\mathcal{D}$ is a sub-σ-algebra of $\mathcal{F}$. For an integrable random variable X, $E[X|\mathcal{D}]$ is the unique (up to changes on events of probability zero) $\mathcal{D}$-measurable random variable such that

$$\int_D X dP = \int_D E[X|\mathcal{D}] dP$$

for all $D \in \mathcal{D}$. If Y is a random variable and $\sigma(Y)$ is its σ-algebra, then $E[X|Y] \equiv E[X|\sigma(Y)]$.

Let X, Y, and Z be random variables. X and Y are *conditionally independent* given Z if and only if

$$E[f(X)g(Y)|Z] = E[f(X)|Z]E[g(Y)|Z]$$

for all bounded measurable f and g.

(a) Show that X and Y are conditionally independent given Z if and only if

$$E[f(X)|Y, Z] = E[f(X)|Z]$$

for all bounded measurable f.

(b) Let X, Y, and Z be independent, $\mathbb{R}$-valued random variables, and let $\psi : \mathbb{R}^2 \to \mathbb{R}$ and $\varphi : \mathbb{R}^2 \to \mathbb{R}$ be Borel measurable functions. Define $U = \psi(X, Z)$ and $V = \varphi(Y, Z)$. Show that U and V are conditionally independent given Z.

(c) Let X and Y be $\mathbb{R}$-valued random variables with $E[X] = 2$ and $E[Y] = 3$. Suppose that that $Z = X + 2Y$ is independent of Y. Show that

$$E[X|Y] = 8 - 2Y.$$

(d) Let X be exponentially distributed with parameter 1. Let $Y = [X]$, the integer part of X, and $Z = X - Y$. Compute $E[Z|Y]$.

2. Let $X_1, \ldots, X_n$ be independent $N(0, 2)$ random variables. Let $Y_1, \ldots, Y_n$ be independent random variables from $N(\mu_1, 2), \ldots, N(\mu_n, 2)$, respectively, for fixed $\mu_1, \ldots, \mu_n$. Let $P_{0,n}$ denote the joint distribution of $X_1, \ldots, X_n$ and $P_{1,n}$ denote the joint distribution of $Y_1, \ldots, Y_n$.

(a) Given $0 < \beta < 1$, let $\mathcal{B}$ be the collection of Borel sets B in $\mathbb{R}^n$ such that $P_{0,n}(B) = \beta$. Define

$$M_n = \inf\{P_{1,n}(B) \ : \ B \in \mathcal{B}\}.$$

Find M_n in terms of $\mu_1, \mu_2, \cdots, \mu_n$, and β.

(b) Suppose $\mu_1, \mu_2, \ldots$ are square summable, i.e., $\sum_{i=1}^{\infty} \mu_i^2 = C < \infty$. Show that M_n converges to a limit as $n \to \infty$, and identify the limit.

3. An experiment was conducted to evaluate the relative effectiveness of four different types of cancer treatments (Factor T) on four categories of patients (Factor C). Patients for the experiment were obtained from hospitals in four different locations (Factor L). At the end of the experiment, each patient's performance was scored on a scale of 0–12, with lower scores indicating better results. Data on each patient's smoking history (i.e., average number of cigarettes smoked per day) were also obtained.

Table 1: Data from cancer treatment experiment

Factors			Smoking		Factors			Smoking	
L	T	C	History	Score	L	T	C	History	Score
1	1	2	5	3	3	1	3	30	8
1	2	3	9	10	3	2	2	0	3
1	3	1	0	4	3	3	4	10	5
1	4	4	11	2	3	4	1	19	8
2	1	4	15	7	4	1	1	0	0
2	2	1	7	2	4	2	4	17	6
2	3	3	20	0	4	3	2	0	9
2	4	2	0	4	4	4	3	2	3

(a) Answer the following questions without using the information on smoking history. Use the 5% level of significance throughout.

i. Is there any statistically significant difference in performance among the four cancer treatments? If so, rank the treatments according to their performance.

ii. Are the mean responses of patients from different hospitals and patients from different categories the same? Test each hypothesis separately.

(b) Include smoking history as a covariate in your model.

i. Find the regression coefficient of smoking history.

ii. Find the residual sum of squares of the model.

iii. Find the estimated standard error of the regression coefficient of smoking history.

(c) Comment on the design of this experiment.

4. *Note: Please pay attention to the specific model that each part in this problem refers to. Some hints on matrix inversion are provided at the end.*

Consider a design on the plane given in Figure 1. This design contains nine points (x_{i1}, x_{i2}) with $x_{i1}, x_{i2} \in \{-1, 0, 1\}$ as marked by open circles. The big circle corresponds to a circle of radius 1 around the origin.

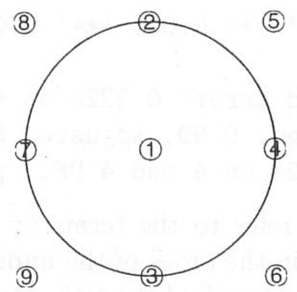

Figure 1: Layout of the design.

Consider the following model:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{12} x_{i1} x_{i2} + \beta_{22} x_{i2}^2 + \epsilon_i, \tag{1}$$

where ϵ_i, $i = 1, \ldots, 9$, are independent and normally distributed with mean 0 and variance σ^2. Suppose model (1) is the *true* model but the data are analyzed under the incorrect assumption:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + +\beta_{22} x_{i2}^2 + \epsilon_i. \tag{2}$$

(a) Consider the following R output corresponding to the fit of model (2).

```
Residuals:
       1        2        3        4        5       6
-0.13167  0.28000 -0.14833 -0.01267 -0.07500 0.08767
       7        8        9
 0.14433 -0.20500  0.06067
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.88167    0.16573   5.320 0.006005 **
         x1 -1.05050    0.09077 -11.573 0.000318 ***
         x2  1.20183    0.09077  13.240 0.000188 ***
        x11  1.07750    0.15722   6.853 0.002373 **
        x22  1.00350    0.15722   6.383 0.003092 **
---
Signif.codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2223 on 4 degrees of freedom
Multiple R-Squared: 0.99, Adjusted R-squared: 0.98
F-statistic: 99.24 on 4 and 4 DF,  p-value: 0.0002966
```

Here x11 and x22 refer to the terms x_1^2 and x_2^2, respectively, and the residuals are listed in the order of the numbers shown in Figure 1. Based on this R output, can you find the estimates for the regression coefficients of the true model (1)? Explain why or why not. Can you estimate β_{12} of the true model (1) using this output. If yes, do so.

(b) Suppose you are told that indeed the true model is given by (1). Now, consider testing $\beta_{11} = \beta_{22}$ within model (1). Write down explicitly what the hypotheses are and construct a test statistic and give its distribution. You need to fill in the numerical values of each of the quantities (or explain how you would obtain them using the above R output) contributing to the test statistics, up to matrix multiplications and inverses, to get full credit.

(c) Consider predicting the result of a future experiment:

$$Y^* = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \beta_{11} x_1^{*2} + \beta_{12} x_1^* x_2^* + \beta_{22} x_2^{*2} + \epsilon^*,$$

independent of $Y_1, \cdots, Y_9$, where $x_1^* = \cos\theta$, $x_2^* = \sin\theta$ for some $\theta \in [0, 2\pi]$. A candidate predictor is $\hat{Y}^* = \hat{\beta}_0 + \hat{\beta}_1 x_1^* + \hat{\beta}_2 x_2 * + \hat{\beta}_{11} x_1^{*2} + \hat{\beta}_{12} x_1^* x_2^* + \hat{\beta}_{22} x_2^{*2}$ where $\hat{\beta}_0, \hat{\beta}_1, \cdots$, are least squares estimators under the model (1). Find an expression for $E\{(\hat{Y}^* - Y^*)^2\}$ in terms of σ^2 and θ. Find upper and lower bounds for this expression.

(d) Suppose that under the incorrect model (2), the predictor of Y^* is $\tilde{Y}^*$. Find an expression for $E\{(Y^* - \tilde{Y}^*)^2\}$.

(e) Show that a necessary and sufficient condition for $\tilde{Y}^*$ to have smaller mean squared prediction error than $\hat{Y}^*$ is $4\beta_{12}^2 < \sigma^2$, except when (x_1^*, x_2^*) is one of $(1,0)$, $(0,1)$, $(-1,0)$ or $(0,-1)$. Explain these exceptions.

Hints: You may find the following formula on matrix inversion in block form useful:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}.$$

For example,

$$\begin{bmatrix} 6 & 0 & 4 \\ 0 & 4 & 0 \\ 4 & 0 & 6 \end{bmatrix}^{-1} = \begin{bmatrix} 3/10 & 0 & -2/10 \\ 0 & 1/4 & 0 \\ -2/10 & 0 & 3/10 \end{bmatrix}.$$