May 2013 PhD Qualifying Examination
Department of Statistics
University of South Carolina
9:00AM–3:00PM

Instructions: This exam consists of six problems. You are to answer all six problems. Use separate sheets of paper for each problem. You are allowed to use the computers and the statistical software in the examination room. However, you are **not** allowed to use the Internet, except to examine help files of the statistical software and to examine data sets that are needed in some of the problems. Provide complete details in your solutions. You have **six hours** to complete this examination. Good luck.

1. An experiment was set up to determine the effects of three treatment factors on a fabric's resistance to abrasion:

- whether or not a surface chemical was applied to the fabric ($i = 1$: yes, $i = 2$: no)
- the type of filler used ($j = 1$: filler 1, $j = 2$: filler 2)
- the proportion of filler used ($k = 1$: 25%, $k = 2$: 50%, $k = 3$: 75%).

The response variable Y_{ijkm} is the loss of fabric in grams (`loss`) after being rubbed on an abrasive surface in a controlled manner for a fixed period of time. The data for this question are available at `http://www.stat.sc.edu/~hansont/abrasion.sas`.

(a) Obtain interaction plots for these data and discuss whether the various two-factor interactions might be necessary. Broadly discuss the impact of surface chemical (`surface`), filler type (`filler`), and proportion of filler (`prop`) on fabric loss based on these plots.

(b) Looking at the Type III tests from fitting the full three-way interaction model, check whether you can drop the three-way interaction and one or more of the two-way interactions that have Type III p-values larger than 0.05 with *one overall* F-test. State your fitted, final model.

(c) For your model in part (b), obtain a standard diagnostic panel and comment on modeling assumptions. Also obtain plots of the raw residuals $e_{ijkm} = Y_{ijkm} - \widehat{Y}_{ijkm}$ versus the indices of each of the three factors i, j, and k, and comment.

(d) Quantitatively describe the effects of the three factors on mean fabric loss. In particular, answer the following questions: Does surface chemical significantly reduce fabric loss? Are there significant differences between the two types of filler? Which filler results in less fabric loss?

2. Consider data that follow an exponential regression with no intercept

$$Y_i \overset{ind}{\sim} \exp(\beta x_i),$$

where the scalar parameter $\beta > 0$ is unknown and the x_i's > 0 are fixed and known for $i = 1, \ldots, n$. That is, $Y_1, \ldots, Y_n$ are independent random variables with density functions

$$f_{Y_i}(y) = \frac{1}{\beta x_i} \exp\left(-\frac{y}{\beta x_i}\right),$$

for $y > 0$. Note that $E(Y_i) = \beta x_i$.

(a) Derive the least squares estimator $\widetilde{\beta}$, i.e., minimize

$$Q(\beta) = \sum_{i=1}^{n} (Y_i - \beta x_i)^2.$$

What are the mean and variance of this estimator?

(b) Derive the maximum likelihood estimator $\widehat{\beta}$.

(c) What is the exact (i.e., finite-sample) sampling distribution of $\widehat{\beta}$?

(d) Which estimator, $\widetilde{\beta}$ or $\widehat{\beta}$, has smaller variance and why?

3. Let Z be a standard normal random variable, that is, $Z \sim \mathcal{N}(0,1)$, and let $X \sim \mathcal{N}(\mu, \sigma^2)$. Let $F_Z(z)$ and $F_X(x)$ denote the cumulative distribution functions (cdfs) of Z and X, respectively.

(a) Derive the distribution of $U = F_Z(Z)$. Provide a derivation; do not just state the answer.

(b) Show that

$$E[F_Z(X)] = F_Z\left(\frac{\mu}{\sqrt{1+\sigma^2}}\right).$$

(c) For this part only, suppose that $X_1, X_2, ..., X_n$ is an iid sample from $\mathcal{N}(\mu, \sigma^2)$, where σ^2 is known and equal to 1. Using the result in part (b) or otherwise, show that the uniformly minimum variance unbiased estimator (UMVUE) for $\tau(\mu) = P_\mu(X_1 < c)$ is

$$\widehat{\tau(\mu)} = F_Z\left[\sqrt{\frac{n}{n-1}}(c - \overline{X})\right],$$

where c is a known constant and $\overline{X} = n^{-1}\sum_{i=1}^{n} X_i$.

4. Consider the density function

$$f_X(x|\theta) = \frac{1}{2}(1+\theta x) I(|x| < 1),$$

where $|\theta| < 1$, which describes the decay distribution of electrons from muon decay when $X = \cos(W)$, and W is the angle measured in an experiment. The parameter θ, which is related to polarization, is to be estimated using $X_1, X_2, ..., X_n$, an iid sample from this distribution.

(a) Find the method of moments (MOM) estimator $\widetilde{\theta}$, and also calculate $E_\theta(\widetilde{\theta})$ and $\mathrm{var}_\theta(\widetilde{\theta})$.

(b) Write out the likelihood function and also the score equation that would be solved to find the maximum likelihood estimator (MLE).

(c) Recall that, under certain regularity conditions, maximum likelihood estimators $\widehat{\theta}$ satisfy

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\widehat{\theta}}),$$

where $\sigma^2_{\widehat{\theta}} = I_1^{-1}(\theta)$ and

$$I_1(\theta) = E_\theta \left\{ \left[\frac{\partial \log f_X(X|\theta)}{\partial \theta} \right]^2 \right\}$$

is the Fisher information based on a single observation. Show that

$$\sigma^2_{\widehat{\theta}} = \frac{2\theta^3}{\log\left(\frac{1+\theta}{1-\theta}\right) - 2\theta}.$$

You may assume that the regularity conditions hold for this distribution.

(d) For small θ and large n, compare $\sigma^2_{\widehat{\theta}}$ with $n \times \mathrm{var}_\theta(\widetilde{\theta})$ using the fact that for $|\theta| < 1$,

$$\log\left(\frac{1+\theta}{1-\theta}\right) = 2\left(\theta + \frac{\theta^3}{3} + \frac{\theta^5}{5} + \cdots\right).$$

5. Thirty women were involved in a study to examine the effects of two drugs, A and B, on heart rates. The women were randomly divided into three groups of 10, with 10 receiving A, 10 receiving B; the remaining 10 were given a placebo P. Four measurements were taken on each individual: a baseline heart rate two minutes after injecting the drug and their heart rates after each of three subsequent five minute intervals. Let Y_{ijk} be the heart rate (beats per minute) of the ith woman ($i = 1, \ldots, 10$) assigned treatment j ($j = 1, 2, 3$) at time point k ($k = 1, 2, 3, 4$ for 2 minutes, 7 minutes, 12 minutes, and 17 minutes after injection). Note subject is nested within treatment. These data are available at `http://www.stat.sc.edu/~hansont/rate.sas`.

(a) Obtain a profile (spaghetti) plot for each of the three treatment groups with a LOESS smooth superimposed on top. Describe what you see in terms of heart rate patterns. Does there seem to be a difference among groups? Are the estimated mean functions (LOESS) approximately parallel? What does this imply about a possible time by treatment interaction?

(b) Consider a model that accounts for repeated measures over time, with factorial treatment structure:

$$Y_{ijk} = \mu + \rho_{i(j)} + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk}.$$

Here, $\rho_{i(j)} \overset{iid}{\sim} N(0, \sigma_\rho^2)$ independent of $\epsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2)$. Let $1 \leq s < t \leq 4$. Derive $\text{corr}(Y_{ijs}, Y_{ijt})$, the correlation between two measurements taken on the same individual at two different time points.

(c) Fit the model in part (b) in a statistical package. Report the ANOVA table for the fixed effects (i.e. the Type III tests). Is there a significant time by treatment interaction here? Does this surprise you given the spaghetti plots?

(d) Look at pairwise differences in the three treatments *at each time point* and discuss significant differences adjusting for multiple comparisons.

(e) Obtain a standard diagnostic panel for the conditional residuals. Are constant variance and normality reasonable for the ϵ_{ijk}?

(f) Obtain the fitted $\widehat{\rho}_{i(j)}$ and corresponding normal probability plot. Is normality among the subject effects reasonable?

6. Suppose that $X_1, X_2, ..., X_n$ is an iid sample from the distribution with density

$$f_X(x|\theta) = \frac{\theta}{x^2} I(x \geq \theta),$$

where $\theta > 0$.

(a) Show that there is an appropriate statistic $T = T(\mathbf{X})$ that has monotone likelihood ratio.

(b) Derive the uniformly most powerful (UMP) level α test for $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. **Note:** You must give an explicit expression for the rejection region R for a test of size α. The rejection region must be simplified as much as possible, and all critical values must be precisely identified.

(c) Write a $100(1 - \alpha)$ percent confidence set for θ.