

# PubH 8452 Spring 2008

## Homework #2

Due Mar 5, 2008

1. Consider the following model that separate within-subject from between-subject variation:

$$Y_{ij} = u_i + \epsilon_{ij}, \quad i = 1, \dots, m; j = 1, \dots, n_i$$

where

$$\text{Var}(u_i) = \sigma_b^2 = \text{between-subject variance}$$

$$\text{Var}(\epsilon_{ij}) = \sigma_w^2 = \text{within-subject variance}$$

$$\text{Var}(Y_{ij}) = \sigma_t^2 = \sigma_b^2 + \sigma_w^2 = \text{total variance.}$$

- a) Define  $N = \sum_{i=1}^m n_i$ ,  $\bar{Y}_i = \sum_{j=1}^{n_i} Y_{ij}/n_i$  and  $\bar{Y}_.. = \sum_{i=1}^m \bar{Y}_i/m$ . Show that

$$\hat{\sigma}_w^2 = \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}{N - m}$$

is an unbiased estimator of  $\sigma_w^2$ .

- b) Assume for the rest of the problem that  $n_i = n$  for all  $i$ . Derive  $\text{var}(\bar{Y}_i)$  and compute the expected value of  $\sum_{i=1}^m (\bar{Y}_i - \bar{Y}_..)^2$ . Thereby derive an unbiased estimator of  $\sigma_b^2$ . Call this estimator  $\tilde{\sigma}_b^2$ .
- c) Following (a) and (b), derive an unbiased estimator of  $\sigma_t^2$ . Call this estimator  $\tilde{\sigma}_t^2$ .
- d) Are  $\hat{\sigma}_w^2$ ,  $\tilde{\sigma}_b^2$  and  $\tilde{\sigma}_t^2$  consistent as  $m \rightarrow \infty$ , holding  $n$  fixed? Are they consistent as  $n \rightarrow \infty$ , holding  $m$  fixed?
- e) Now consider the estimator

$$\hat{\sigma}_t^2 = \sum_{i=1}^m \sum_{j=1}^n (Y_{ij} - \bar{Y}) / (N - 1),$$

where  $\bar{Y} = \sum_{i=1}^m \sum_{j=1}^n Y_{ij}/N$ . By rewriting this estimator as a linear combination of  $\sum_{i=1}^m \sum_{j=1}^n (Y_{ij} - \bar{Y}_i)^2$  and  $\sum_{i=1}^n (\bar{Y}_i - \bar{Y})^2$ , derive the bias in  $\hat{\sigma}_t^2$ .

Will  $\hat{\sigma}_t^2$  be consistent as  $m \rightarrow \infty$ , holding  $n$  fixed? Is it consistent as  $n \rightarrow \infty$ , holding  $m$  fixed?

- f) What correlation structure for  $Y_{ij}$  does this model correspond to? Using your derivations above, what is a consistent estimator of the parameter in this correlation model?
2. Each year the U.S. Naval Postgraduate School sets aside a “Discovery Day” during which the general public is invited into their laboratories. Our data come from October 21 1995, when visitors could test their motor learning and hand-eye coordination in the Human Systems Integration Laboratory. The variable of interest, “contact time,” comes from a rotary pursuit tracking experiment. A rotating sheet contains a 3/4 inch target spot which moves. The sheet is either in the shape of a circle or a square. The subject is asked to maintain contact with the moving target spot using a metal wand. Each trial lasted 15 seconds and the total contact time was recorded. The target spot moves keeps a constant speed in a circular path on the circle sheet, but moves at varying speeds in an irregular path on the square sheet.

Each of 108 visitors completed the trial four times but for only one of the two shapes. Age and gender were also recorded. Visitors tended to come in family groups, but that information was not recorded.

The data set `tracking.dat` is available from the course website and the first few lines are shown here:

Sex	Age	Shape	Trial1	Trial2	Trial3	Trial4
M	31	Box	2.68	4.14	7.22	8.00
M	30	Box	7.09	8.55	8.79	9.68
M	30	Box	6.05	6.25	7.04	7.80
M	27	Box	4.35	6.50	5.17	6.50
M	30	Box	4.08	6.00	6.82	6.68

- a) Using centered age (defined as age minus average age) and centered age squared, gender, shape and trials as covariates and tracking time as response, fit random effect models with:
- Random intercept for such subject only.
  - Random intercept and slope (for trial) for each subject.
  - Nested Random intercepts for subjects within each gender.
  - Nested Random intercepts and slopes for subjects within each gender.
  - Random intercept for each subjects, but allow the variances of the random intercepts to differ between genders and/or shapes.

You can also choose additional models for the random effects that you think might be appropriate.

- b) Use appropriate test statistics and information criteria to choose a single random effect model which adequately describes the data. Provide a brief report summarizing your decisions and conclusions. Interpret your model of choice.
  - c) Based on the selected random effect model, simplify the model for the mean response. Provide a brief summary report to justify your decisions to include or exclude terms in the mean model.
  - d) Provide graphic evaluation of the goodness-of-fit of your final model.
3. Here we will investigate the impact of non-normality in the random effects, in particular, heterogeneity due to an unobserved variable.

Consider a single covariate  $\mathbf{X}_i = (1, 3, 5, 6, 10)^T$  for  $i = 1, \dots, m$ , and a response variable from the following model with random intercept:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + b_{i0} + b_{i1} X_{ij} + \epsilon_{ij}$$

where  $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ ,  $\beta_0 = \beta_1 = 1.0$ .

Instead of having normally distributed random effects,  $b_{i0}$  and  $b_{i1}$  are independent and are from a mixture distribution which is, with probability 0.5,  $\mathcal{N}(-2, 1)$  and with probability 0.5,  $\mathcal{N}(2, 1)$ .

- a) Simulate the random effects  $b_{i0}$  for  $m = 1000$  subjects. Plot its histogram.
- b) For  $\sigma^2 = 0.5, 5$  or  $30$ , simulate  $b_{i0}, b_{i1}$  and  $Y$ , fit the above model and plot the histogram of the empirical Bayes estimates of  $b_{i0}$  and  $b_{i1}$ . Explain the result and try to reach a more general conclusion.
- c) Generate 1000 datasets for  $\sigma^2 = 5$  or  $30$ , estimate the fixed effects  $(\beta_0, \beta_1)$  using the above linear mixed effect model, general linear model with independent work correlation, and GEE with exchangeable correlation. Calculate the confidence intervals using model-based (for LME and GLM) and empirical standard error estimates (GLM) of the fixed effects and their coverage percentage. Discuss your results.