BIOS 667, Sample Final Exam

The notation established in class will be used: $Y_{ij}, b_i, \nu_{ij}, \mu_{ij}, x_{ij}, Y_i, X_i, n_i, G, R_i, \Sigma_i, Y, X, K$, etc.

1. Consider the model M_1:

$$\nu_{ij} = \mathrm{E}[Y_{ij}|b_{i1}] = \beta_1 + b_{i1} + \beta_2 x_{ij}.$$

   Using data from 5000 independent subjects, the model parameters were estimated by REML (not full ML) and the maximized -2 log-likelihood was 200. Next, a model differing from M_1 only by not having b_{i1} in it was estimated by REML and the maximized -2 log-likelihood was 203.84.

   In the context of model M_1, and given the above information, is it possible to test the null hypothesis $g_{11} = 0$ against the alternative hypothesis $g_{11} > 0$? If yes, give the details (and compute a p-value). If no, explain why.

2. Consider the model M_2:

$$\nu_{ij} = \mathrm{E}[Y_{ij}|b_{i1}] = \beta_1 + b_{i1} + \beta_2 x_{ij}.$$

   Using data from 3000 independent subjects, the model parameters were estimated by REML (not full ML) and the maximized -2 log-likelihood was 300. Next, a model differing from M_2 only by not having the term $\beta_2 x_{ij}$ in it was estimated by REML and the maximized -2 log-likelihood was 305.

   In the context of model M_2, and given the above information, is it possible to test the null hypothesis $\beta_2 = 0$ against the alternative hypothesis $\beta_2 \neq 0$? If yes, give the details (including how to compute a p-value). If no, explain why.

3. Suppose that each of K independent subjects has 4 observations, and the model is a linear mixed model with no time trend,

$$\nu_{ij} = \mathrm{E}[Y_{ij}|b_i] = b_i + \beta \qquad j = 1, 2, 3, 4,$$

   where $b_1, \cdots, b_K$ are random variables distributed as iid $N(0, 4)$, and Y_i given b_i is multivariate normal with mean vector ν_i and covariance matrix $R_i = I_{4\times 4}$. Define $\bar{Y}_i = (1/4)(Y_{i1} + Y_{i2} + Y_{i3} + Y_{i4})$ and $\bar{Y} = (1/K)\sum_{i=1}^{K} \bar{Y}_i$

   (a) Find the best linear unbiased predictor (BLUP) of b_i based on *only* Y_{i1}. Compute (the numerical value of) its prediction mean squared error.

   (b) Find the best linear unbiased predictor (BLUP) of b_i based on *only* $\bar{Y}_i$. Compute (the numerical value of) its prediction mean squared error.

4. Suppose that in a longitudinal study there are 2 observations for each subject. Define $\nu_{ij} = \mathrm{E}[Y_{ij}|b_i]$ where b_i is a subject-specific random effect. The random effects have the

following discrete distribution: $P(b_i = -1) = P(b_i = 1) = 0.5$. The responses Y_{i1} and Y_{i2} are conditionally independent given b_i, and $\text{var}(Y_{ij}|b_i) = 2\nu_{ij}$. Consider the model

$$\log \nu_{ij} = b_i + 2 + 2x_{ij} \qquad j = 1, 2,$$

where $x_{i1} = 0, x_{i2} = 1, i = 1, \ldots, K$. The covariate x_{ij} is a treatment indicator in a cross-over study. In the following, compute numerical values and simplify the answers as much as possible.

(a) Compute the marginal mean (2×1) vector, $\mu_i = \text{E}[Y_i]$.

(b) Compute the (2×2) marginal covariance matrix, $\text{cov}(Y_i)$. Compute $\text{corr}(Y_{i1}, Y_{i2})$.

(c) Compute the intra-class correlation (ICC) for Y_{i1}.

(d) Compute α_1 and α_2 in

$$\log \text{E}[Y_{ij}] = \log \mu_{ij} = \alpha_1 + \alpha_2 x_{ij}.$$

Is there an attenuation effect?

Formulae

ICC = between / total

$\Sigma_i = Z_i G Z_i^\top + R_i$

If A is a predictor of B, the prediction MSE is $\text{E}[(A - B)^2]$

Under H_0 (i.e. when H_0 is true),

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \lambda_0^2) \text{ as } n \to \infty,$$

under H_1,

$$\sqrt{n}(\hat{\theta} - \theta_1) \xrightarrow{d} N(0, \lambda_1^2) \text{ as } n \to \infty.$$

$$\sqrt{n} \quad = \quad \frac{Z_{1-\alpha/2}\lambda_0 + Z_{1-\gamma}\lambda_1}{\theta_1 - \theta_0}$$