Homework 3 - BIOS 6643

Dominic Adducci

2023-09-17

Question 1

Part A
Write the linear mixed effects model in matrix form for a model with categorical time (reference cell coding) and group (depressed vs not depressed). The marginal variance covariance is unstructured. You may use the subject level formulation. Write out the Y_i, X_i, β, any random effects, and R_i.

$$Y_{ij} = \beta_0 + b_{0i} + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \epsilon_{ij}$$

Where $R_i \sim N(0, R_i)$, and $b_0 \sim N(0, G_i)$. i can be any number between 1 and 52 (the number of subjects), and j can be any time point between 1 and 6. b_0 is the subject specific intercept for cortisol.

Putting this into matrix form:
$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i$$

Where Y_i is a (6x1) vector, X_i is a (1x7) vector, β is a (7x1) vector, Z_i is a (6x1) vector, b_i is a (1x6) vector, and ϵ_i is a (6x1) vector. $b_i \sim N(0, G_i)$ and $\epsilon \sim N(0, R_i)$.

Part B
How many parameters are there to estimate for fixed effects and what are they in your notation?

There are 7 parameters in total which need to be estimated for fixed effects of the subject level model. Each subject will have 7 parameters in their model.

- β_0: Fixed intercept, reference group is time 1 and women who are not depressed.
- β_1: Fixed effect for Time 2.
- β_2: Fixed effect for Time 3.
- β_3: Fixed effect for Time 4.
- β_4: Fixed effect for Time 5.
- β_5: Fixed effect for Time 6.
- β_6: Fixed effect for depressed group.

Part C
How many parameters are there to estimate in the marginal variance covariance (V_i) and what are they in your notation?

Fixed effects are not random, meaning $V(X\beta) = 0$. This leaves the variance of the random effect and the fixed effect, which returns the following marginal variance of Y.

$$V(Y_{ij}) = Z_i G_i Z_i^T + R_i$$

Because this is an unstructured the number of parameters is equal to $\frac{n(n-1)}{2}$, where in this case n is equal to 6. Therefore there are 21 parameters in the marginal variance-covariance matrix which need to be estimated. As more time points are added (n) the number of parameters which need to be estimated for the unstructured variance-covariance model increases rapidly.

Part D
Interpret 2 of the fixed effects (not including the intercept).

- β_1: Fixed effect for Time 2. This means the average difference in cortisol between time point 1 and time point 2, where 1 is the reference group.

- β_6: Fixed effect for depressed group. This means the average difference in cortisol between the control group and the depressed group, where the control group is the reference.

Question 2

Part A Write the linear mixed effects model in matrix form for a model with categorical time (reference cell coding) and group (depressed vs not depressed). The marginal variance covariance is compound symmetric. You may use the subject level formulation. Write out the Y_i, X_i, β, any random effects (including G_i), and R_i and V_i.

The marginal model will be as follows.

$$E(Y_{ij}) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \epsilon_{ij}$$

Where $Y \sim N(X_{ij}\beta, \sigma^2)$. i can be any number between 1 and 52 (the number of subjects), and j can be any time point between 1 and 6.

Putting this model into matrix form:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i$$

Where X_i is a (52x7) vector, β is a (7x1) vector, and ϵ is a (52x1) vector.

Part B How many parameters are there to estimate for fixed effects and what are they in your notation?

For fixed effects, there are again 7 parameters which need to be estimated for the subject level model..

- β_0: Fixed intercept, reference group is time 1 and women who are not depressed.
- β_1: Fixed effect for Time 2.
- β_2: Fixed effect for Time 3.
- β_3: Fixed effect for Time 4.
- β_4: Fixed effect for Time 5.
- β_5: Fixed effect for Time 6.
- β_6: Fixed effect for depressed group.

Part C How many parameters are there to estimate in the marginal variance covariance (V_i) and what are they in your notation?

For a compound symmetric variance-covariance structure there are only two parameters which need to be estimated as variances are all equal to each other, and covariances are all equal to each other. The variance is the within subject correlated errors, which can be represented by G_{ii} where $G_{11} = G_{22} = G_{33}...$, added to the variance from the error, R_i. Therefore, the variance is:

$$V_i = G_{ii} + R_i$$

Question 3

Part A Write the linear mixed effects model in matrix form for a model with categorical time (reference cell coding) and group (depressed vs not depressed). There is a random intercept for subject. You may use the subject level formulation.

$$Y_{ij} = \beta_0 + b_{0i} + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \epsilon_{ij}$$

Putting this into matrix form:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i$$

Where Y_i is a (6x1) matrix, X_i is a (1x7) vector, β is a (7x1) matrix, Z_i is a (6x1) matrix ,b_i is a scalar, and ϵ_i is a (6x1) vector. $b_i \sim N(0, G_i)$ and $\epsilon \sim N(0, R_i)$.

Part B How many parameters are there to estimate for fixed effects and what are they in your notation?

There are 7 parameters in total which need to be estimated for fixed effects of the subject level model. Each subject will have 7 parameters in their model.

- β_0: Fixed intercept, reference group is time 1 and women who are not depressed.
- β_1: Fixed effect for Time 2.
- β_2: Fixed effect for Time 3.
- β_3: Fixed effect for Time 4.
- β_4: Fixed effect for Time 5.
- β_5: Fixed effect for Time 6.
- β_6: Fixed effect for depressed group.

Part C How many parameters are there to estimate in the marginal variance-covariance (V_i) and what are they in your notation?

Fixed effects are not random, meaning $V(X\beta) = 0$. This leaves the variance of the random effect and the fixed effect, which returns the following marginal variance of Y.

$$V(Y_{ij}) = Z_i G_i Z_i^T + R_i$$

Where $V(Y_{ij}) = Z_i G_i Z_i^T$ is the subject specific random intercept variance. R_i is the variance of the error.

Part D Interpret the random effects.

The random effect accounts for the difference in baseline cortisol for each individual subject.

Part E Show that this model is equivalent to the compound symmetry marginal model above.

The expected value for both the random intercept and compound symmetry marginal models is the fixed intercepts. The variance are equivalent because in the compound symmetry marginal model the variance is assumed constant for the subject, and in the random intercept model the variance for each subject is fixed as G_i. Both variances are then summed with the error of the variance, R_i. The marginal model can be written as follows:

$$Y_i \sim N(X_i\beta, Z_i G_i Z_i^T)$$

Question 4

Part A Write the linear mixed effects model in matrix form for a model with an interaction between categorical time and group with a compound symmetry marginal variance. You may use the subject level formulation.

There are many coefficients in this model, so only the matrix form will be shown.

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\epsilon}_i$$

Where Y_i is a (6x1) vector, X_i is a (1x12) vector, β is a (12x1) vector, and ϵ_i is a (6x1) vector. $\epsilon \sim N(0, R_i)$.

Part B How many parameters are there to estimate for fixed effects and what are they in your notation?

The parameters for the fixed effects are in the $X_i\beta$ term. There are 12 which need to be estimated.

Part C How many parameters are there to estimate in the marginal variance-covariance (V_i), and what are they in your notation?

Due to the compound symmetry there are only two parameters which need to be estimated for marginal variance-covariance.

$$V_i = g_{ii} J_n + R_i$$

Where J_n is a (nx1) vector, g_{11} is the within subject variance, and R_i is the varience from error.

Question 5

Let $\mathbf{Y} = (Y_1, ..., Y_n)^T$ be the vector of independent measurements. $Y_i \sim N(\mu, \sigma^2)$. An MLE based estimator for $\hat{\sigma}^2 = \sum_i (Y_i - \bar{Y})^2 / n$. Show the estimator is biased (i.e. find $E(\hat{\sigma}^2)$).

$$E(\hat{\sigma}^2) = E\left(\sum_i (Y_i - \bar{Y})^2 / n\right)$$

$$= E\left(\frac{1}{n} \sum_i (Y_i^2 - 2Y_i\bar{Y} + \bar{Y}^2)\right)$$

$$= E\left(\frac{1}{n}\left(\sum_i Y_i^2 - 2\sum_i Y_i\bar{Y} + \sum_i \bar{Y}^2\right)\right)$$

We can use the fact that $\sum_i Y_i^2 = nY_i^2$, $\sum_i Y_i = n\bar{Y}$, and $\sum_i \bar{Y}^2 = n\bar{Y}^2$. We can simplify the middle term with the following operation:

$$2\sum_i Y_i\bar{Y} = 2\bar{Y}\sum_i Y_i = 2n\bar{Y}^2$$

$$E(\hat{\sigma}^2) = \frac{1}{n}E\left(nY_i^2 - 2n\bar{Y}^2 + n\bar{Y}^2\right)$$

$$= \frac{1}{n}E\left(nY_i^2 - n\bar{Y}^2\right)$$

$$= \frac{1}{n}\left(E(nY_i^2) - E(n\bar{Y}^2)\right)$$

$$= \frac{1}{n}\left(nE(Y_i^2) - nE(\bar{Y}^2)\right)$$

$$= E(Y_i^2) - E(\bar{Y}^2)$$

We can then use the definition of variance, $\sigma^2 = E(Y^2) - E(Y)^2$, to rewrite both terms.

$$E(\hat{\sigma}^2) = E(Y_i^2) - E(\bar{Y}^2)$$
$$= \sigma_Y^2 + E(Y)^2 - \sigma_{\bar{Y}}^2 - E(Y)^2$$
$$= \sigma_Y^2 - \sigma_{\bar{Y}}^2$$

From here we know the variance of $\bar{Y}$ can be expressed as follows:

$$\sigma_{\bar{Y}}^2 = V\left(\frac{1}{n}\sum_i Y_i\right)$$

$$= \frac{1}{n^2}V\left(\sum_i Y_i\right)$$

$$= \frac{1}{n}\sigma_Y^2$$

Plugging this expression in:

$$E(\hat{\sigma}^2) = \sigma_Y^2 - \sigma_{\bar{Y}}^2$$

$$= \sigma_Y^2 - \frac{1}{n}\sigma_Y^2$$

$$= \frac{n-1}{n}\sigma_Y^2$$

Thus, the estimator is biased.

Question 6

Now let $\mathbb{1}_n$ be the n-dimension vector of 1's. The distribution of Y can be written as $N(\mu\mathbb{1}_n, \sigma^2 I_n)$. Let A be any nx(n-1) matrix with n-1 columns linearly independent and orthogonal to the vector $\mathbb{1}_n$. Show the distribution of $U = A^T Y$ (the so called error contrasts) is $N(0, \sigma^2 A^T A)$. Then show the MLE for σ^2 is $\hat{\sigma}^2 = Y^T A(A^T A)^{-1}A^T Y/(n-1) = \Sigma_i(Y_i - \bar{Y})^2/(n-1)$. This is a REML estimator because the likelihood is restricted based on the (n-1) error contrasts.

Finding the expected value of U:

$$E(U) = E(A^T Y) = A^T E(Y) = A^T \mu\mathbb{1}_n$$

After transposing the A matrix each μ that is multiplied by 1 is also multiplied by a -1 in the same row. This means every μ term goes to 0, resulting in $E(U) = 0$.

The variance can be found with the following operation, where A^T is a constant:

$$V(U) = V(A^T Y) = A^T V(Y)A = A^T \sigma A = \sigma A^T A$$

σ is a scalar, so it can be moved in front of the A^T. The identity matrix drops out because $A^T I_n = A^T$.

Finding the estimate of $\hat{\sigma}^2$ is the same process as finding the variance estimate for the multivariate regression, where the log likelihood derivative is taken with respect to σ. (Not entirely sure on this one).

$$logL(U) = \frac{n}{2}log(2\pi) - nlog(\sigma) - \frac{1}{2\sigma^2}\left[(U-0)^T(U-0)\right]$$

$$logL(U) = \frac{n}{2}log(2\pi) - nlog(\sigma) - \frac{1}{2\sigma^2}\left[U^T U\right]$$

Taking the derivative with respect to σ and setting equal to 0:

$$-\frac{n}{2\sigma} - \frac{1}{2(\sigma^2)^2}[U^T U] = 0 \rightarrow \frac{n}{2\sigma} = \frac{1}{2(\sigma^2)^2}[U^T U]$$

$$\hat{\sigma}^2 = \frac{U^T U}{n} = \frac{(A^T Y)^T A^T Y}{n}$$

Doesn't match, not sure which transformation to do from here (assuming this is an intermediate step).

Question 7

Consider the multiple linear regression framework $Y_i = \beta_0 + \beta_1 X_{i1} + ... + \beta_{p-1} X_{i,p-1} + E_i$. Show that the MLE for σ^2 in a multiple regression is: $(Y - HY)^T(Y - HY)/n$ and that it is biased downward by a factor $(n-p)/n$.

Starting with the likelihood function for a multivariate linear regression:

$$L(Y) = \prod \frac{1}{(2\pi)^{p/2}\sigma} exp\left[\frac{-1}{2}(Y - X\hat{\beta})^T\Sigma^{-1}(Y - X\hat{\beta})\right]$$

Where $X\hat{\beta}$ is the expected value for the regression, and $\Sigma^{-1} = \sigma^{-2}I$. The identity matrix drops out of the expression when plugged in because any matrix multiplied by the identity matrix is itself. The expression $X\hat{\beta}$ can be rewritten as HY, because $\hat{\beta}$ can be expressed as the least squares estimate, $(X^TX)^{-1}X^TY$, and the hat matrix is $H = X(X^TX)^{-1}X^T$.

$$L(Y) = \prod \frac{1}{(2\pi)^{p/2}\sigma} exp\left[\frac{-1}{2\sigma^2}(Y - HY)^T(Y - HY)\right]$$

Taking the log of this to get the log-likelihood:

$$logL(Y) = -n\frac{p}{2}log(2\pi) - nlog(\sigma) - \frac{1}{2\sigma^2}(Y - HY)^T(Y - HY)$$

Taking the derivative of this with respect to σ, and setting equal to 0 to find the maximum:

$$\frac{\partial}{\partial\sigma}logL(Y) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}(Y - HY)^T(Y - HY) = 0$$

Rearranging this to find σ^2:

$$-\frac{n}{\sigma} + \frac{1}{\sigma^3}(Y - HY)^T(Y - HY) = 0$$
$$\frac{n}{\sigma} = \frac{1}{\sigma^3}(Y - HY)^T(Y - HY)$$
$$\sigma^2 = \frac{1}{n}(Y - HY)^T(Y - HY)$$

Which matches the given expression. The maximum is confirmed by finding the second derivative and determining if it is negative. The bias for the variance of a single variate linear regression is a factor of $\frac{n-1}{n}$. Extending this to the multivariate model each covariate reduces the bias, $\frac{n-p}{n}$.

Question 8

Let A be an $n \times (n - p)$ with $n - p$ linearly independent columns orthogonal to the columns of the design matrix X. As above, let $U = A^TY$ be the so called error contrasts and show that $U \sim N(0, \sigma^2 A^T A)$.

This appears to be the first part of question 6?

Finding the expected value of U:

$$E(U) = E(A^TY) = A^TE(Y) = A^T\mu\not{1}_n$$

After transposing the A matrix each μ that is multiplied by 1 is also multiplied by a -1 in the same row. This means every μ term goes to 0, resulting in $E(U) = 0$.

The variance can be found with the following operation, where A^T is a constant:

$$V(U) = V(A^TY) = A^TV(Y)A = A^T\sigma A = \sigma A^T A$$

σ is a scalar, so it can be moved in front of the A^T. The identity matrix drops out because $A^T I_n = A^T$.

Question 9

Show that the MLE likelihood estimator of σ^2 using the likelihood of U is $Y^T(I - H)Y/(n - p)$. This is the REML estimator of the variance term.