ST 518 Homework 10

Eric Warren

November 18, 2023

Contents

1 Problem 1

Define the symbols $Y, X, \beta, Z, \gamma, \epsilon, G, R$, and then use them to write out the general form for a linear, mixed effects model where all random effects are assumed mean zero, normally distributed random variables.

We know the following about each variable:

- Y is a N x 1 column vector, the outcome variable
- X is a N x p matrix of the p predictor variables
- β is a p x 1 column vector column vector of the fixed-effects regression coefficients
- Z is the N x q design matrix for the q random effects (the random complement to the fixed X)
- γ is the q x 1 vector of the random effects (the random complement to the fixed β)
- ϵ is the N x 1 column vector of the residuals, the part of Y that is not explained by the model
- G is the variance-covariance matrix of the random effects.
- R is the residual covariance structure known as $R = I\sigma_\epsilon^2$ where I is the identity matrix (diagonal matrix of 1s) and σ_ϵ^2 is the residual variance.

Our model for a linear mixed effects case is $Y = X\beta + Z\gamma + \epsilon$ where $\gamma_j \sim N(0, G)$ and $\epsilon_{ij} \sim N(0, R)$.

2 Problem 2

Consider the plant height example presented on slide 63/71 of the mixed model lecture notes. Consider a linear model with mixed effects for a two factor experiment where factor A is modeled with fixed effects and B random, nested in A: $Y_{ijk} = \mu + \alpha_i + B_{j(i)} + E_{ijk}$ where $B_{j(i)} \sim^{iid} N(0, \sigma^2_{B(A)})$ and $E_{ijk} \sim^{iid} N(0, \sigma^2)$.

2.1 Part A

Derive the variance of a treatment average: $Var(\bar{Y_{i..}})$

$Var(\bar{Y_{i..}}) = Var(\mu + \alpha_i + \bar{B_{.(i)}} + \bar{E_{i..}}) = \frac{\sigma^2_{B(\alpha)}}{b} + \frac{\sigma^2}{nb} = \frac{1}{nb}(n\sigma^2_{B(A)} + \sigma^2)$

2.2 Part B

Derive the variance of a contrast among treatment averages: $Var(\sum_i c_i \bar{Y_{i..}})$

$Var(\sum_i c_i \bar{Y_{i..}}) = \sum_i c_i^2 * Var(\mu + \alpha_i + \bar{B_{.(i)}} + \bar{E_{i..}}) = \sum_i c_i^2 (\frac{\sigma^2_{B(\alpha)}}{b} + \frac{\sigma^2}{nb}) = \sum_i c_i^2 (\frac{1}{nb}(n\sigma^2_{B(A)} + \sigma^2))$

2.3 Part C

Give an unbiased estimator of $Var(\bar{Y_{i..}})$ and show that it is unbiased.

$\hat{Var}(\bar{Y_{i..}}) = \frac{1}{nb}(n\hat{\sigma_{B(A)}}^2 + \hat{\sigma}^2) = \frac{1}{nb}(MS(B(A)))$. We will it is unbiased by saying that $E(\hat{Var}(\bar{Y_{i..}})) = Var(\bar{Y_{i..}})$. $E(\hat{Var}(\bar{Y_{i..}})) = E(\frac{1}{nb}(MS(B(A)))) = \frac{1}{nb}E(MS(B(A))) = \frac{1}{nb}(n\sigma^2_{B(A)} + \sigma^2) = Var(\bar{Y_{i..}})$. Therefore, $\hat{Var}(\bar{Y_{i..}})$ is an unbiased estimator of $Var(\bar{Y_{i..}})$.

2.4 Part D

Give an unbiased estimator of $Var(\sum_i c_i \bar{Y_{i..}})$ and show that it is unbiased.

$\hat{Var}(\sum_i c_i \bar{Y_{i..}}) = \sum_i c_i^2 * \frac{1}{nb}(n\hat{\sigma_{B(A)}}^2 + \hat{\sigma}^2) = \sum_i c_i^2 * \frac{1}{nb}(MS(B(A)))$. We will it is unbiased by saying that $E(\hat{Var}(\sum_i c_i \bar{Y_{i..}})) = Var(\sum_i c_i \bar{Y_{i..}})$. $E(\hat{Var}(\sum_i c_i \bar{Y_{i..}})) = E(\sum_i c_i^2 * \frac{1}{nb}(MS(B(A)))) = \sum_i c_i^2 * \frac{1}{nb}E(MS(B(A))) = \sum_i c_i^2 * \frac{1}{nb}(n\sigma^2_{B(A)} + \sigma^2) = Var(\sum_i c_i \bar{Y_{i..}})$. Therefore, $\hat{Var}(\sum_i c_i \bar{Y_{i..}})$ is an unbiased estimator of $Var(\sum_i c_i \bar{Y_{i..}})$.

2.5 Part E

Consider the matrix formulation of this model. Assume the covariance matrix of the error term ϵ is a diagonal matrix $R = I_{20} * \sigma^2$. Give the entire fixed and random effect design matrices, X and Z.

We know that X deals with the α_i and Z deals with the $B_{j(i)}$. Therefore, we just got to match terms where X is a 20x6 matrix with the columns being $\mu, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$ and Z is a 20x10 matrix with the

columns being $B_{1(1)}, B_{2(1)}, B_{1(2)}, B_{2(2)}, B_{3(1)}, B_{3(2)}, B_{4(1)}, B_{4(2)}, B_{5(1)}, B_{5(2)}$. Therefore, we know that $X =$

$$\begin{pmatrix}
1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}$$

and we know that $Z =$

$$\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}$$

Give the covariance matrix, G of the vector of random effects, γ.

We know that G is also an ab x ab diagonal matrix of $I * \sigma^2_{B(A)}$. In this case $a = 5$ and $b = 2$ so we have a 10 x 10 diagonal matrix of $I * \sigma^2_{B(A)}$ in this case. Thus our matrix $G =$

$$\begin{pmatrix}
\sigma^2_{B(A)} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \sigma^2_{B(A)} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \sigma^2_{B(A)} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \sigma^2_{B(A)} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \sigma^2_{B(A)} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \sigma^2_{B(A)} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{B(A)} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{B(A)} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{B(A)} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{B(A)}
\end{pmatrix}$$

3 Problem 3

An experiment was conducted to compare four labs in their assessment of amylase. Note that inference is to be limited to these four labs ("narrow scope") as opposed to the population of all labs. Four technicians were sampled from each lab, and each technician was assigned n = 10 specimens to measure. A nested model was fit with mean squares given below, along with lab means.

3.1 Part A

Write out a linear mixed effects model appropriate to the experiment.

Since we are seeing that there are a finite number of levels that would be used again in the same experience we know labs is fixed. Other than being told in the directions that it is a mixed model, we can also see technicians could involve new levels if resampled from the lab population and is selected conceptually from an infinite population (we do not know how many technicians there are). Thus, technicians is a random

effect. On top of that, technicians come from specific labs so they are nested within the labs. We can now write our model as $Y_{ijk} = \mu + \alpha_i + B_{j(i)} + E_{ijk}$ where $i = 1, 2, 3, 4$, $j = 1, 2, 3, 4$, and $k = 1, 2, ..., 10$.

3.2 Part B

Is the observed difference between the means for labs 2 and 3 significant at level $\alpha = 0.05$? Explain.

We know that $\bar{Y_{2..}} = 3.6$ and $\bar{Y_{3..}} = 4.6$. Thus, $\bar{Y_{3..}} - \bar{Y_{2..}} = 4.6 - 3.6 = 1.0$. Now we need to find

$$
\begin{aligned}
& SE(\bar{Y_{3..}} - \bar{Y_{2..}}) \\
&= \sqrt{Var(\bar{Y_{3..}} - \bar{Y_{2..}})} \\
&= \sqrt{Var(\frac{1}{nb}\sum_{j}^{4}\sum_{k}^{10}\mu + \alpha_3 + B_{j(3)} + E_{3jk} - (\frac{1}{nb}\sum_{j}^{4}\sum_{k}^{10}\mu + \alpha_2 + B_{j(2)} + E_{2jk}))} \\
&= \sqrt{Var(\mu + \alpha_3 + \bar{B_{.(3)}} + \bar{E_{3..}} - (\mu + \alpha_2 + \bar{B_{.(2)}} + \bar{E_{2..}}))} \\
&= \sqrt{Var(\alpha_3 - \alpha_2 + \bar{B_{.(3)}} - \bar{B_{.(2)}} + \bar{E_{3..}} - \bar{E_{2..}})} \\
&= \sqrt{\frac{\sigma^2_{B(\alpha)}}{b} + \frac{\sigma^2_{B(\alpha)}}{b} + \frac{\sigma^2}{nb} + \frac{\sigma^2}{nb}} \\
&= \sqrt{\frac{2}{nb}(n\sigma_{B(\alpha)} + \sigma)}
\end{aligned}
$$

Now, we need $SE(\bar{Y_{3..}} - \bar{Y_{2..}}) = Var(\bar{Y_{3..}} - \bar{Y_{2..}}) = \sqrt{\frac{2}{nb}(n\hat{\sigma_{B(\alpha)}} + \hat{\sigma})} = \sqrt{\frac{2}{nb}MS(B(\alpha))}$

Knowing that $SE(\bar{Y_{3..}} - \bar{Y_{2..}}) = \sqrt{\frac{2}{nb}MS(B(\alpha))}$, we can make a $1 - \alpha$ or $1 - .05 = .95$ or 95% confidence interval instead of a normal hypothesis test. If 0 is not in the confidence interval then this is the same as rejecting a hypothesis test. Note that $MS(B(\alpha)) = 1.63$ which comes from the MS(technicians) part of our output in the directions since that is $MS(B(\alpha))$. Our confidence interval is $\bar{Y_{3..}} - \bar{Y_{2..}} \pm t_{1-\frac{\alpha}{2},a(b-1)} * SE(\bar{Y_{3..}} - \bar{Y_{2..}})$ ===> $1.0 \pm t_{1-\frac{.05}{2},4(4-1)}\sqrt{\frac{2}{nb}MS(B\alpha)}$ ===> $1.0 \pm t_{.975,12}\sqrt{\frac{2}{10*4}1.63}$. Now, $t_{.975,12} = 2.178813$ which comes from `qt(.975, 12)` in R so our confidence interval is now $1.0 \pm t_{.975,12}\sqrt{0.0815}$ ===> $1.0 \pm 2.178813 * 0.285482$ ===> 1.0 ± 0.6220119 which gives us bounds of (0.3779881, 1.6220119). As we can see 0 is **NOT** contained in the interval, so we do have statistically significant to say that observed means between labs 2 and 3 are significantly different.

3.3 Part C

Estimate the correlation between two measurements made by the same technician.

First, by identifying the situation, we know that the same technician must be at the same lab. Thus, the formula we use to find this correlation is $Corr(Y_{ijk1}, Y_{ijk2}) = \frac{Cov(Y_{ijk1}, Y_{ijk2})}{\sigma^2_A + \sigma^2_{B(A)} + \sigma^2} = \frac{\sigma^2_A + \sigma^2_{B(A)})}{\sigma^2_A + \sigma^2_{B(A)} + \sigma^2}$. Now we just need to find the values.

- $\sigma^2 = MS(E) = 1.02$
- $\sigma^2_{B(A)} = \frac{MS(B(A)) - MS(E)}{n} = \frac{1.63 - 1.02}{10} = 0.061$
- $\sigma^2_A = \frac{MS(A) - MS(B(A))}{nb} = \frac{9.07 - 1.63}{10*4} = \frac{7.44}{40} = 0.186$

Now we can plug this back into our formula of $Corr(Y_{ijk1}, Y_{ijk2}) = \frac{Cov(Y_{ijk1},Y_{ijk2})}{\sigma_A^2+\sigma_{B(A)}^2+\sigma^2} = \frac{\sigma_A^2+\sigma_{B(A)}^2)}{\sigma_A^2+\sigma_{B(A)}^2+\sigma^2} = \frac{.186+.061}{.186+.061+1.02} = \frac{0.247}{1.267} = 0.1949487$. Therefore, the estimated correlation between two measurements made by the same technician is about 0.1949487.

3.4 Part D

Suppose Labs 1 and 2 use one technique while Labs 3 and 4 use another. Test the hypothesis that the average of the Lab 1 and Lab 2 means is equal to the average of the Lab 3 and Lab 4 means. Report an 95% appropriate confidence interval to evaluate the effect of the technique.

First let us get the appropriate values.

- $\bar{Y_{1..}} = 3.8$
- $\bar{Y_{2..}} = 3.6$
- $\bar{Y_{3..}} = 4.6$
- $\bar{Y_{4..}} = 4.4$

Therefore what we are testing which we will call $\hat{\theta} = \frac{1}{2}(\bar{Y_{3..}}+\bar{Y_{4..}})-\frac{1}{2}(\bar{Y_{1..}}+\bar{Y_{2..}}) = \frac{1}{2}(4.6+4.4)-\frac{1}{2}(3.8+3.6) = 4.5 - 3.7 = 0.8$. Now we are testing if $H_0 : \theta = 0$ and $H_A : \theta \neq 0$ by seeing if we make a 95% confidence interval contains the value of 0 or not. If it does not, then we have statistically significant evidence to say that the average of the Lab 1 and Lab 2 means is **NOT** equal to the average of the Lab 3 and Lab 4 means.

Now we need to find our confidence interval. We know with help from **Problem 2**, **Problem 3 Part B**, and looking at past modules, we know that our standard error is $\sqrt{\sum_{i=1}^{4} c_i^2 * \frac{1}{nb} * MS(B(A))}$. This is the same as $n * b = 10 * 4 = 40$ for each lab. We can rewrite this to get our standard error to be

$$\sqrt{\sum_{i=1}^{4} c_i^2 * \frac{1}{nb} * MS(B(A))}$$

$$\sqrt{\sum_{i=1}^{4} c_i^2 \frac{1}{40} * MS(B(A))}$$

$$\sqrt{((-\frac{1}{2})^2 + (-\frac{1}{2})^2 + (\frac{1}{2})^2 + (\frac{1}{2})^2) * \frac{1}{40} * 1.63}$$

$$\sqrt{(\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4}) * 0.04075}$$

$$\sqrt{1 * 0.04075}$$

$$\sqrt{0.04075} ===> 0.2018663$$

.

Now our confidence interval is $\hat{\theta} \pm t_{.975,4(4-1)} * SE(\hat{\theta}) ===> 0.8 \pm t_{.975,12} * 0.2018663 ===> 0.8 \pm 2.178813 * 0.2018663 ===> 0.8 \pm 0.4398289$, which gives us bounds of (0.3601711, 1.2398289) for our 95% confidence interval. As we can see 0 is **NOT** contained in the interval, so we do have statistically significant to say that the average of the Lab 1 and Lab 2 means is **NOT** equal to the average of the Lab 3 and Lab 4 means. Actually since all values in the interval are positive, we have reason to believe that the average of the Lab 3 and Lab 4 means are greater than the average of the Lab 1 and Lab 2 means.