Homework 7

Brian Cervantes Alvarez

May 29, 2024

ST 553 Statistical Methods

Problem 1

Part 1.1

A scenario where it would be more appropriate to model fertilizer as a random effect is when we have a large number of different fertilizer types, and each fertilizer type is randomly selected from a larger population of possible fertilizers. In this case, we are not interested in the effect of each specific fertilizer type, but rather in the variability of the effects across all fertilizer types. This approach allows for generalization to other fertilizers not included in the study.

Part 1.2

The random effects model is given,

$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{ijk},$$

where,

- y_{ijk} is the number of pods on the k^{th} sample in the j^{th} plot in the i^{th} treatment
- $\alpha_i \sim N(0, \sigma^2_\alpha)$ is the random effect of the fertilizer treatment
- $\beta_{j(i)} \sim N(0, \sigma^2_\beta)$ is the random effect of the plot nested within the treatment
- $\epsilon_{ijk} \sim N(0, \sigma^2)$ is the random error

All random effects (α_i, $\beta_{j(i)}$, and ϵ_{ijk}) are assumed to be mutually independent.

Part 1.3

To show that $\text{cov}(y_{ijk}, y_{ijk'}) = \sigma_\alpha^2 + \sigma_\beta^2$ for $k \neq k'$, we can use the given models,

$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{ijk}$$

and

$$y_{ijk'} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{ijk'}$$

Next, the covariance between y_{ijk} and $y_{ijk'}$ for $k \neq k'$ is,

$$\text{cov}(y_{ijk}, y_{ijk'}) = \text{cov}(\mu + \alpha_i + \beta_{j(i)} + \epsilon_{ijk}, \mu + \alpha_i + \beta_{j(i)} + \epsilon_{ijk'})$$

Since μ is a constant we have,

$$\text{cov}(y_{ijk}, y_{ijk'}) = \text{cov}(\alpha_i + \beta_{j(i)} + \epsilon_{ijk}, \alpha_i + \beta_{j(i)} + \epsilon_{ijk'})$$

Using the properties of covariance and the fact that ϵ_{ijk} and $\epsilon_{ijk'}$ are independent,

$$\text{cov}(y_{ijk}, y_{ijk'}) = \text{cov}(\alpha_i, \alpha_i) + \text{cov}(\beta_{j(i)}, \beta_{j(i)}) + \text{cov}(\epsilon_{ijk}, \epsilon_{ijk'}) = \sigma_\alpha^2 + \sigma_\beta^2 + 0$$

Therefore,

$$\text{cov}(y_{ijk}, y_{ijk'}) = \sigma_\alpha^2 + \sigma_\beta^2$$

Problem 2

Part 2.1

To show that MSA is equal to

$$\frac{1}{a-1}\sum_{i=1}^{a}\sum_{j=1}^{n}(\bar{y}_{i.}-\bar{y}_{..})^2$$

we can use the definition of MSA,

$$\text{MSA} = \frac{1}{a-1}\left(n\sum_{i=1}^{a}\bar{y}_{i.}^2 - an\bar{y}_{..}^2\right). \tag{1}$$

Given that

$$\bar{y}_{i.} = \frac{1}{n}\sum_{j=1}^{n} y_{ij}$$

and

$$n\bar{y}_{i.} = \sum_{j=1}^{n} y_{ij}$$

we have

$$\sum_{i=1}^{a}\bar{y}_{i.} = \sum_{i=1}^{a}\left(\frac{1}{n}\sum_{j=1}^{n} y_{ij}\right) = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij} = \frac{an}{n}\bar{y}_{..} = a\bar{y}_{..}$$

Therefore,

$$n\sum_{i=1}^{a}\bar{y}_{i.}^2 = \sum_{i=1}^{a} n\bar{y}_{i.}^2$$

Hence, the MSA remains,

$$\text{MSA} = \frac{1}{a-1}\left(n\sum_{i=1}^{a}\bar{y}_{i.}^2 - an\bar{y}_{..}^2\right)$$

Part 2.2

To show that the $E[MSA]$ is equal to $\sigma^2 + n\sigma_\alpha^2$, let's define the variance,

$$\mathrm{var}(\bar{y}_{i.}) = \frac{\sigma_\alpha^2}{n} + \frac{\sigma^2}{n}$$

where $\bar{y}_{..}$ is the grand mean, and has variance

$$\mathrm{var}(\bar{y}_{..}) = \frac{\sigma_\alpha^2}{an} + \frac{\sigma^2}{an}$$

Next, let's break down the expectation of the MSA,

$$E[\mathrm{MSA}] = \frac{1}{a-1} E\left(n\sum_{i=1}^{a} \bar{y}_{i.}^2 - an\bar{y}_{..}^2\right)$$

$$E\left(n\sum_{i=1}^{a} \bar{y}_{i.}^2\right) = n\sum_{i=1}^{a} E[\bar{y}_{i.}^2]$$

$$E[\bar{y}_{i.}^2] = \mathrm{var}(\bar{y}_{i.}) + [E(\bar{y}_{i.})]^2 = \sigma_\alpha^2 + \frac{\sigma^2}{n} + \mu^2$$

$$E[an\bar{y}_{..}^2] = an\left(\frac{\sigma_\alpha^2}{an} + \frac{\sigma^2}{an} + \mu^2\right) = \sigma_\alpha^2 + \frac{\sigma^2}{n} + a\mu^2$$

Thus,

$$E[\mathrm{MSA}] = \frac{1}{a-1}\left(n\sum_{i=1}^{a}(\sigma_\alpha^2 + \frac{\sigma^2}{n} + \mu^2) - a(\sigma_\alpha^2 + \frac{\sigma^2}{n} + \mu^2)\right)$$

We can simplify this down to,

$$E[\mathrm{MSA}] = \frac{1}{a-1}\left(a\sigma_\alpha^2 + \sigma^2 - \sigma_\alpha^2 - \frac{\sigma^2}{n}\right).$$

Therefore,

$$E[\mathrm{MSA}] = \sigma^2 + n\sigma_\alpha^2.$$

Problem 3

Type 3 Analysis of Variance Table (Table 1)

Source	DF	Sum of Squares	Mean Square	Expected Mean Square
A	3	594.105874	198.035291	Var(Residual) + 3.7115 Var(A*B) + 11.134 Var(A)
B	2	764.462739	382.231370	Var(Residual) + 3.7139 Var(A*B) + 14.856 Var(B)
A*B	6	46.946315	7.824386	Var(Residual) + 4.0085 Var(A*B)
Residual	39	36.225167	0.928850	Var(Residual)

Type 3 Analysis of Variance Table (Table 2)

Source	DF	Sum of Squares	Mean Square	Expected Mean Square
A	2	257.556875	128.778437	Var(Residual) + 2.9412 Var(A*B) + 5.8824 Var(A)
B	1	16.850700	16.850700	Var(Residual) + 2.88 Var(A*B) + 8.64 Var(B)
A*B	2	3.924522	1.962261	Var(Residual) + 2.9412 Var(A*B)
Residual	13	13.032500	1.002500	Var(Residual)

Now, to find the mean square for the denominator, we can do the following,

$$n_1(\text{Var (Residual)} + 4.0085\text{Var (A*B)}) + n_2(\text{Var (Residual)}) = \text{Var (Residual)} + 3.7139\text{Var (A*B)}$$

Then, this implies,

$$n_1 + n_2 = 1$$

$$4.0085n_1 = 3.7139$$

Solving for n_1 and n_2:

$$n_1 = \frac{3.7139}{4.0085} \approx 0.9265$$

$$n_2 = 1 - \frac{3.7139}{4.0085} \approx 0.0735$$

Thus, the mean square for the denominator for the full MSE is given by,

$$\text{MSE}_{\text{full}} = \frac{3.7139}{4.0085}\text{MS}_{\text{AB}} + \left(1 - \frac{3.7139}{4.0085}\right)\text{MSE}$$

Substituting the values from the table,

$$\text{MSE}_{\text{full}} = \frac{3.7139}{4.0085}(7.824386) + \left(1 - \frac{3.7139}{4.0085}\right)(0.928850) \approx 7.317607$$

The df are given by the Satterthwaite approximation,

$$\text{df} = \frac{\left(\sum_{i=1}^{4} g_i \text{MS}_i\right)^2}{\sum_{i=1}^{4}\left(\frac{g_i^2}{\nu_i}\right)\text{MS}_i^2}$$

Where,

$$g_1 = \frac{3.7139}{4.0085}$$

$$g_2 = 1 - \frac{3.7139}{4.0085}$$

Hence,

$$\text{df} = \frac{7.317607^2}{\left(\frac{(3.7139/4.0085)^2(7.824386)^2}{39}\right) + \left(\frac{(1-3.7139/4.0085)^2(0.928850)^2}{6}\right)} \approx \frac{53.54737}{8.758946} \approx 6.113449$$

Table 2

Next, to find the mean square for the denominator, we can do the following,

$$n_1(\text{Var (Residual)} + 2.9412\text{Var (A*B)}) + n_2(\text{Var (Residual)}) = \text{Var (Residual)} + 2.88\text{Var (A*B)}$$

So, this implies that,

$$n_1 + n_2 = 1$$

$$2.9412n_1 = 2.88$$

Solving for n_1 and n_2 we get,

$$n_1 = \frac{2.88}{2.9412} \approx 0.979$$

$$n_2 = 1 - \frac{2.88}{2.9412} \approx 0.021$$

Thus, the mean square for the denominator full MSE is given by,

$$\text{MSE}_{\text{full}} = \frac{2.88}{2.9412}\text{MS}_{\text{AB}} + \left(1 - \frac{2.88}{2.9412}\right)\text{MSE}$$

Substituting the values from the table,

$$\text{MSE}_{\text{full}} = \frac{2.88}{2.9412}(1.9622) + \left(1 - \frac{2.88}{2.9412}\right)(1.0025) \approx 1.942$$

The df are given by the Satterthwaite approximation,

$$\text{df} = \frac{\left(\sum_{i=1}^{4} g_i \text{MS}_i\right)^2}{\sum_{i=1}^{4}\left(\frac{g_i^2}{\nu_i}\right)\text{MS}_i^2}$$

Where,

$$g_1 = \frac{2.88}{2.9412}$$

$$g_2 = 1 - \frac{2.88}{2.9412}$$

Therefore,

$$\text{df} = \frac{1.942^2}{\left(\frac{(2.88/2.94)^2(1.9622)^2}{13}\right) + \left(\frac{(1-2.88/2.9412)^2(1.0025)^2}{2}\right)} \approx \frac{3.7724}{1.8459} \approx 2.044$$

Problem 4

Part 4.1

This is a cross random effects model,

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

where,

- y_{ijk}: the cholesterol levels of the k^{th} blood sample from the j^{th} patient processed at the i^{th} lab
- μ is the overall mean cholesterol level
- α_i is the random effect of the i_{th} lab
- β_j is the random effect of the j_{th} patient
- $(\alpha\beta)_{ij}$ is the random interaction effect between the i_{th} lab and the j_{th} patient
- ϵ_{ijk}: the random error associated with the k^{th} blood sample from the j^{th} patient processed at the i^{th} lab

Assumptions:

- All random effects (α_i, β_j, $(\alpha\beta)_{ij}$, and ϵ_{ijk}) are assumed to be mutually independent.
- The random effects α_i, β_j, $(\alpha\beta)_{ij}$, and ϵ_{ijk} are independent.
- $\alpha_i \sim N(0, \sigma^2_\alpha)$, $\beta_j \sim N(0, \sigma^2_\beta)$, $(\alpha\beta)_{ij} \sim N(0, \sigma^2_{\alpha\beta})$, and $\epsilon_{ijk} \sim N(0, \sigma^2)$.

Part 4.2

Hypotheses:

- $H_0 : \sigma_\alpha^2 = 0$, no variability in cholesterol levels by lab
- $H_A : \sigma_\alpha^2 > 0$, variability in cholesterol levels by lab

The F-stat for the null hypothesis follows a $F_{2,10}$, which gives us,

$$F^* = \frac{\text{MSA}}{\text{MSAB}} \approx \frac{25.1775}{8.6075} \approx 2.925$$

The resulting p-value is approximately 0.09997.

Therefore, there is weak, and not convincing, evidence to suggest that there is no variability in the cholesterol levels by lab.

Part 4.3

Hypotheses:

- $H_0 : \sigma^2_{\alpha\beta} = 0$, no interaction between lab and patient
- $H_A : \sigma^2_{\alpha\beta} > 0$, interaction between lab and patient

The F-stat for the null hypothesis follows a $F_{10,18}$, which gives us,

$$F^* = \frac{\text{MSAB}}{\text{MSE}} \approx \frac{8.607500}{1.629444} \approx 5.28$$

We get the p-value of 0.0011.

In this case, there is strong and convincing evidence to suggest that the variability according to the lab is different for at least one patient.

Part 4.4

The covariance paramters are given in the following table:

Parameter	Point Estimate
σ_α^2	1.8308
σ_β^2	969.83
$\sigma_{\alpha\beta}^2$	3.4890
σ^2	1.6294

The SAS code that was used to produce this table

```
proc mixed data = Cholesterol method = REML;
class Lab Patient;
model chl =;
random Lab|Patient;
run;
```