

UNIREP ANOVA Expected Mean Squares

“It is hoped that the material here will be sufficiently illustrative to show what is involved generally, and also to enable the reader to decide whether he wants to be involved generally” (Crowder and Hand)

When your statistics professor says to the class, “You should know how to derive these results (but you won’t be tested on this)” think to yourself: “I could do that.” Be satisfied at that point. Or be satisfied that there is likely to be one idiot in the class (like me) who takes the declaration seriously.

In today’s post, I will show how to derive the expected mean squares of the univariate repeated measures ANOVA model. I’ll start with the sum of squares for between groups. I’ll also end there, as I would like to keep my hair from turning gray faster than it already is.

First, the preliminaries must be hashed out. Assume a linear model for observation Y_{hlj} of individual h ($1, \dots, r_l$) in group l ($1, \dots, q$) at time (or repeated measure) j ($1, \dots, n$) such that:

$$Y_{hlj} = \mu + \tau_l + \gamma_j + (\tau\gamma)_{lj} + b_{hl} + e_{hlj}$$

The parameters are:

- μ is the overall mean.
- τ_l is the deviation from μ associated with group l .
- γ_j is the deviation from μ associated with time j .
- $(\tau\gamma)_{lj}$ is the deviation associated with group l at time j (i.e., the interaction of time and group).
- b_{hl} is the random effect associated with unit h unit in group l .
- e_{hlj} represents the within-unit sources of variation.

Additional assumptions

- $b_{hl} \sim N(0, \sigma_b^2)$ and all independent.
- $e_{hlj} \sim N(0, \sigma_e^2)$ and all independent.
- b_{hl} and e_{hlj} are mutually independent.
- To force identifiability of the parameter estimates, we make the these constraints: $\sum_{l=1}^q \tau_l = 0$, $\sum_{j=1}^n \gamma_j = 0$, $\sum_{l=1}^q (\tau\gamma)_{lj} = 0 = \sum_{j=1}^n (\tau\gamma)_{lj}$.

The distributional assumptions imply a compound symmetric covariance structure for \mathbf{Y}_{hl} . That is, $\mathbf{Y}_{hl} = \sigma_b^2 \mathbf{J}_n + \sigma_e^2 \mathbf{I}_n$, where \mathbf{J}_n is a $n \times n$ matrix of ones and \mathbf{I}_n is a $n \times n$ identity matrix.

Define $\bar{Y}_{.l.} = \sum_{h=1}^{r_l} \sum_{j=1}^n Y_{hlj}$ as the sample mean across groups and $\bar{Y}_{...} = \sum_{l=1}^q \sum_{h=1}^{r_l} \sum_{j=1}^n Y_{hlj}$ as the overall sample mean. Now we can get started.

$$\begin{aligned}
E(MS_G) &= E\left(\frac{SS_G}{q-1}\right) = \frac{1}{q-1} E\left[\sum_{l=1}^q nr_l(\bar{Y}_{.l} - \bar{Y}_{...})^2\right] \\
&= \frac{n}{q-1} \sum_{l=1}^q r_l E[(\bar{Y}_{.l} - \bar{Y}_{...})^2] \\
&= \frac{n}{q-1} \sum_{l=1}^q r_l E\left[\left(\sum_{h=1}^{r_l} \sum_{j=1}^n (\mu + \tau_l + \gamma_j + (\tau\gamma)_{lj} + b_{hl} + e_{hlj})/r_l n \right. \right. \\
&\quad \left. \left. - \sum_{h=1}^q \sum_{h=1}^{r_l} \sum_{j=1}^n (\mu + \tau_l + \gamma_j + (\tau\gamma)_{lj} + b_{hl} + e_{hlj})/mn \right)^2\right] \\
&= \frac{n}{q-1} \sum_{l=1}^q r_l E\left[\left(\tau_l + \frac{1}{r_l n} \sum_{h=1}^{r_l} \sum_{j=1}^n b_{hl} + \frac{1}{r_l n} \sum_{h=1}^{r_l} \sum_{j=1}^n e_{hlj} \right. \right. \\
&\quad \left. \left. - \frac{1}{mn} \sum_{h=1}^q \sum_{h=1}^{r_l} \sum_{j=1}^n b_{hl} - \frac{1}{mn} \sum_{h=1}^q \sum_{h=1}^{r_l} \sum_{j=1}^n e_{hlj} \right)^2\right]
\end{aligned}$$

To see how μ , γ_j , and $(\tau\gamma)_{lj}$ fall out, remember the constraints! For example,

$$\sum_{h=1}^{r_l} \sum_{j=1}^n \frac{\tau_l}{r_l n} - \sum_{l=1}^q \sum_{h=1}^{r_l} \sum_{j=1}^n \frac{\tau_l}{mn} = \tau_l - \sum_{l=1}^q \frac{r_l \tau_l}{m} = \tau_l - 0 = \tau_l$$

Recall the following. $E[X^2] = E[X]^2 + V(X)$. $E(b_{hl}) = 0$. $E(e_{hlj}) = 0$. $E(\tau_l) = \tau_l$. $V(\tau_l) = 0$ (since it is a constant). b_{hl} is independent of e_{hlj} , so $Cov(b_{hl}, e_{hlj}) = 0$.

$$\begin{aligned}
E(MS_G) &= \frac{n}{q-1} \sum_{l=1}^q r_l \left\{ \left[E\left(\tau_l + \frac{1}{r_l n} \sum_{h=1}^{r_l} \sum_{j=1}^n b_{hl} + \frac{1}{r_l n} \sum_{h=1}^{r_l} \sum_{j=1}^n e_{hlj} - \frac{1}{mn} \sum_{l=1}^q \sum_{h=1}^{r_l} \sum_{j=1}^n b_{hl} - \frac{1}{mn} \sum_{l=1}^q \sum_{h=1}^{r_l} \sum_{j=1}^n e_{hlj}\right) \right]^2 \right. \\
&\quad \left. + V\left(\tau_l + \frac{1}{r_l n} \sum_{h=1}^{r_l} \sum_{j=1}^n b_{hl} + \frac{1}{r_l n} \sum_{h=1}^{r_l} \sum_{j=1}^n e_{hlj} - \frac{1}{mn} \sum_{l=1}^q \sum_{h=1}^{r_l} \sum_{j=1}^n b_{hl} - \frac{1}{mn} \sum_{l=1}^q \sum_{h=1}^{r_l} \sum_{j=1}^n e_{hlj}\right) \right\} \\
&= \frac{n}{q-1} \sum_{l=1}^q r_l \left\{ \tau_l^2 + V\left(\frac{1}{r_l n} \sum_{h=1}^{r_l} \sum_{j=1}^n b_{hl} - \frac{1}{mn} \sum_{l=1}^q \sum_{h=1}^{r_l} \sum_{j=1}^n b_{hl}\right) \right. \\
&\quad \left. + V\left(\frac{1}{r_l n} \sum_{h=1}^{r_l} \sum_{j=1}^n e_{hlj} - \frac{1}{mn} \sum_{l=1}^q \sum_{h=1}^{r_l} \sum_{j=1}^n e_{hlj}\right) \right\}
\end{aligned}$$

Let's take the variance terms one at a time.

$$\begin{aligned}
& V \left(\frac{1}{r_l n} \sum_{h=1}^{r_l} \sum_{j=1}^n b_{hl} - \frac{1}{mn} \sum_{l=1}^q \sum_{h=1}^{r_l} \sum_{j=1}^n b_{hl} \right) \\
&= V \left(\frac{1}{r_l n} \sum_{h=1}^{r_l} \sum_{j=1}^n b_{hl} \right) + V \left(\frac{1}{mn} \sum_{l=1}^q \sum_{h=1}^{r_l} \sum_{j=1}^n b_{hl} \right) - 2Cov \left(\frac{1}{r_l n} \sum_{h=1}^{r_l} \sum_{j=1}^n b_{hl}, \frac{1}{mn} \sum_{l=1}^q \sum_{h=1}^{r_l} \sum_{j=1}^n b_{hl} \right) \\
&= \frac{1}{r_l^2} \sum_{h=1}^{r_l} V(b_{hl}) + \frac{1}{m^2} \sum_{l=1}^q \sum_{h=1}^{r_l} V(b_{hl}) - \frac{2}{r_l m} Cov \left(\sum_{h=1}^{r_l} b_{hl}, \sum_{l=1}^q \sum_{h=1}^{r_l} b_{hl} \right) \\
&= \frac{1}{r_l^2} \sum_{h=1}^{r_l} V(b_{hl}) + \frac{1}{m^2} \sum_{l=1}^q \sum_{h=1}^{r_l} V(b_{hl}) - \frac{2}{r_l m} V \left(\sum_{h=1}^{r_l} b_{hl} \right) \\
&= \frac{1}{r_l^2} \sum_{h=1}^{r_l} V(b_{hl}) + \frac{1}{m^2} \sum_{l=1}^q \sum_{h=1}^{r_l} V(b_{hl}) - \frac{2}{r_l m} \sum_{h=1}^{r_l} V(b_{hl}) \\
&= \frac{1}{r_l} \sigma_b^2 + \frac{1}{m} \sigma_b^2 - 2 \frac{1}{m} \sigma_b^2 \\
&= \sigma_b^2 \left(\frac{1}{r_l} - \frac{1}{m} \right)
\end{aligned}$$

The fourth line in the above section follows from the assumption that the units are independent, thus $Cov(b_{hl}, b_{hl}) = 0$ where the indexes are not equal.

Following a similar process, we can show:

$$V \left(\frac{1}{r_l n} \sum_{h=1}^{r_l} \sum_{j=1}^n e_{hlj} - \frac{1}{mn} \sum_{l=1}^q \sum_{h=1}^{r_l} \sum_{j=1}^n e_{hlj} \right) = \sigma_e^2 \left(\frac{1}{r_l n} - \frac{1}{mn} \right)$$

Now, let's go back to where we left off.

$$\begin{aligned}
E(MS_G) &= \frac{n}{q-1} \sum_{l=1}^q r_l \left\{ \tau_l^2 + V \left(\frac{1}{r_l n} \sum_{h=1}^{r_l} \sum_{j=1}^n b_{hl} - \frac{1}{mn} \sum_{l=1}^q \sum_{h=1}^{r_l} \sum_{j=1}^n b_{hl} \right) \right. \\
&\quad \left. + V \left(\frac{1}{r_l n} \sum_{h=1}^{r_l} \sum_{j=1}^n e_{hlj} - \frac{1}{mn} \sum_{l=1}^q \sum_{h=1}^{r_l} \sum_{j=1}^n e_{hlj} \right) \right\} \\
&= \frac{n}{q-1} \sum_{l=1}^q r_l \tau_l^2 + \frac{n \sigma_b^2}{q-1} \sum_{l=1}^q r_l \left(\frac{1}{r_l} - \frac{1}{m} \right) + \frac{n \sigma_e^2}{q-1} \sum_{l=1}^q r_l \left(\frac{1}{r_l n} - \frac{1}{mn} \right) \\
&= \frac{n}{q-1} \sum_{l=1}^q r_l \tau_l^2 + \frac{n \sigma_b^2}{q-1} \left(\sum_{l=1}^q 1 - \frac{\sum_{l=1}^q r_l}{m} \right) + \frac{\sigma_e^2}{q-1} \left(\sum_{l=1}^q 1 - \frac{\sum_{l=1}^q r_l}{m} \right) \\
&= \frac{n}{q-1} \sum_{l=1}^q r_l \tau_l^2 + \frac{n \sigma_b^2}{q-1} (q-1) + \frac{\sigma_e^2}{q-1} (q-1) \\
&= \frac{n}{q-1} \sum_{l=1}^q r_l \tau_l^2 + n \sigma_b^2 + \sigma_e^2
\end{aligned}$$

And there we are.