

Lack of independence for observations sharing a block

In the completely randomized design, $Y_{ij} = \mu + \alpha_i + e_{ij}$, the covariance between observations Y_{ij} and $Y_{i'j'}$ is:

$$\begin{aligned} \text{cov}(Y_{ij}, Y_{i'j'}) &= E(Y_{ij}Y_{i'j'}) - E(Y_{ij})E(Y_{i'j'}) \\ &= E[(\mu + \alpha_i + e_{ij})(\mu + \alpha_{i'} + e_{i'j'})] - E(\mu + \alpha_i + e_{ij})E(\mu + \alpha_{i'} + e_{i'j'}) \\ &= E(\mu^2 + \mu\alpha_i + \mu e_{i'j'} + \alpha_i\mu + \alpha_i^2 + \alpha_i e_{i'j'} + e_{ij}\mu + e_{ij}\alpha_{i'} + e_{ij}e_{i'j'}) - (\mu + \alpha_i)(\mu + \alpha_{i'}) \\ &= \mu^2 + 2\mu\alpha_i + \alpha_i^2 - (\mu^2 + 2\mu\alpha_{i'} + \alpha_{i'}^2) = 0 \end{aligned}$$

We could have looked at $Y_{i'j}$ instead of $Y_{i'j'}$ and gotten the same result. So therefore any two different observations have a covariance of 0. For normally distributed random variables, a covariance of 0 is equivalent to independence.

In the randomized block design with random blocks, $Y_{ij} = \mu + \alpha_i + b_j + e_{ij}$, the covariance between two observations in the same block is:

$$\begin{aligned} \text{cov}(Y_{ij}, Y_{i'j}) &= E(Y_{ij}Y_{i'j}) - E(Y_{ij})E(Y_{i'j}) \\ &= E[(\mu + \alpha_i + b_j + e_{ij})(\mu + \alpha_{i'} + b_j + e_{i'j})] - \\ &E(\mu + \alpha_i + b_j + e_{ij})E(\mu + \alpha_{i'} + b_j + e_{i'j}) \\ &= E(\mu^2 + \mu\alpha_{i'} + \mu\alpha_i + \alpha_i\alpha_{i'} + b_j^2 + e_{ij}e_{i'j} + \text{other terms}) - (\mu + \alpha_i)(\mu + \alpha_{i'}) \\ &= E(b_j^2) + E(e_{ij}e_{i'j}) = \sigma_b^2 + 0 = \sigma_b^2 \end{aligned}$$

Thus two observations within the same block do covary. If random assignment was performed within each block, we should tend to have equal covariance between any two treatments within the same block, because the randomization helps ensure that $E(e_{ij}e_{i'j}) = 0$. If we also have equal variances per treatment, the resulting covariance structure of the set of observations within each block has what is known as **compound symmetric** structure. The usual ANOVA F tests are still valid when we have compound symmetry, or even if a more general condition called the **Huynh-Feldt condition** is satisfied. In a situation where random blocks are used and treatments are not randomized within each block, such as when the treatment factor is time, then typically $E(e_{ij}e_{i'j}) \neq 0$, and neither compound symmetry nor the Huynh-Feldt condition may be satisfied. In this case we have to change our approach for analyzing the data.