COCHRAN–SATTERTHWAITE APPROXIMATION

FOR LINEAR COMBINATIONS OF MEAN SQUARES

Suppose $MS_1, \ldots, MS_k$ are independent random variables and that

$$\frac{\text{df}_i \, MS_i}{E(\text{MS}_i)} \sim \chi^2_{\text{df}_i} \quad \forall \ i = 1, \ldots, k.$$ 

Consider the random variable

$$S^2 = a_1 \, MS_1 + a_2 \, MS_2 + \ldots + a_k \, MS_k,$$

where $a_1, a_2, \ldots, a_k$ are known constants in $\mathbb{R}$.  

\[ E(S^2) = a_1 E(MS_1) + \ldots + a_k E(MS_k) \]

\[ \text{Var}(S^2) = a_1^2 \text{Var}(MS_1) + \ldots + a_k^2 \text{Var}(MS_k) \]

\[ = a_1^2 \left[ \frac{E(MS_1)}{df_i} \right]^2 2 df_i + \ldots + a_k^2 \left[ \frac{E(MS_k)}{df_k} \right]^2 2 df_k \]

\[ = 2 \sum_{i=1}^{k} a_i^2 \left[ \frac{E(MS_i)}{df_i} \right]^2 \]

A natural estimator of \( \text{Var}(S^2) \) is

\[ \text{Var}(S^2) = 2 \sum_{i=1}^{k} a_i^2 MS_i^2 / df_i \]
Recall that \( \frac{df_i MS_i}{E(MS_i)} \sim \chi^2_{df_i} \) \( \forall i=1,\ldots,k \).

If \( S^2 = \alpha_1 MS_1 + \cdots + \alpha_k MS_k \) is distributed like each of the random variables in the linear combination,

\[
\frac{df S^2}{E(S^2)} \sim \chi^2_{df}.
\]
Note that \( E \left[ \frac{df}{E(S^2)} \right] = df = E(X^2) \)

\[
\text{Var} \left[ \frac{df \cdot S^2}{E(S^2)} \right] = \frac{df^2}{[E(S^2)]^2} \text{ Var} \ (S^2)
\]

Equating this expression to \( \text{Var} \ (X^2) = 2df \)
and solving for \( df \) yields

\[
df = \frac{2 \ [E(S^2)]^2}{\text{Var} \ (S^2)}
\]
Replacing $E(S^2)$ with $S^2$ and $\text{Var}(S^2)$ with $\text{Var}(S^2)$ yields

$$\hat{df} = \frac{(S^2)^2}{\sum_{i=1}^{k} a_i^2 MS_i^2 / df_i}$$

$$= \frac{(\sum_{i=1}^{k} a_i MS_i)^2}{\sum_{i=1}^{k} a_i^2 MS_i^2 / df_i},$$

which is the Cochran-Satterthwaite formula for the approximate degrees of freedom of a linear combination of mean squares.