偏差平方和の分解の式

$$\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(y_{ij}-\bar{y}\right)^2=\sum_{j=1}^{m}n_j\left(\bar{y}_j-\bar{y}\right)^2+\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(y_{ij}-\bar{y}_j\right)^2$$

測定値に等分散正規性 $y_{ij}{\sim}N(\mu,\sigma^2)$ が仮定できるとき，平方和の分解の左辺を母分散 σ^2 で割った式は次のように変形される。

$$\begin{aligned}
\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(\frac{y_{ij}-\bar{y}}{\sigma}\right)^2 &= \sum_{j=1}^{m}\sum_{i=1}^{n_j}\left\{\frac{\left(y_{ij}-\mu\right)-(\bar{y}-\mu)}{\sigma}\right\}^2 \\
&= \sum_{j=1}^{m}\sum_{i=1}^{n_j}\left\{\left(\frac{y_{ij}-\mu}{\sigma}\right)^2-\frac{2\left(y_{ij}-\mu\right)(\bar{y}-\mu)}{\sigma^2}+\left(\frac{\bar{y}-\mu}{\sigma}\right)^2\right\} \\
&= \sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(\frac{y_{ij}-\mu}{\sigma}\right)^2-\frac{2(\bar{y}-\mu)}{\sigma^2}\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(y_{ij}-\mu\right)+\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(\frac{\bar{y}-\mu}{\sigma}\right)^2 \\
&= \sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(\frac{y_{ij}-\mu}{\sigma}\right)^2-\frac{2(\bar{y}-\mu)}{\sigma^2}(n\bar{y}-n\mu)+n\left(\frac{\bar{y}-\mu}{\sigma}\right)^2 \\
&= \sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(\frac{y_{ij}-\mu}{\sigma}\right)^2-n\left(\frac{\bar{y}-\mu}{\sigma}\right)^2=\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(\frac{y_{ij}-\mu}{\sigma}\right)^2-\left(\frac{\bar{y}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2{\sim}\chi_{(n-1)}
\end{aligned}$$

これは，総ケース数と同じ *n* 個の「標準化された正規変量」の二乗和から，確率変数である標本平均を標準化したもの（第 3 章 1-2 参照），すなわち 1 個の標準正規変量の二乗を引いたものなので，定義上，自由度 *n*-1 のカイ二乗分布に従う[1]。

平方和の分解の右辺の第一項と第二項を母分散で割った$SS_{Between}/\sigma^2$，SS_{within}/σ^2もそれぞれ，自由度 *m*-1，*n*-*m* のカイ二乗分布に従う（次ページ以降）。

[1] 自由度 ν の χ^2 分布とは，互いに独立な ν 個の標準正規変量の二乗和として定義される。したがって，カイ二乗分布で近似できるために等分散正規性が必要とされる。

$$\sum_{j=1}^{m}\frac{n_j\left(\bar{y}_j-\bar{y}\right)^2}{\sigma^2}=\sum_{j=1}^{m}\left(\frac{\bar{y}_j-\bar{y}}{\frac{\sigma}{\sqrt{n_j}}}\right)^2=\sum_{j=1}^{m}\left\{\frac{\left(\bar{y}_j-\mu\right)-\left(\bar{y}-\mu\right)}{\frac{\sigma}{\sqrt{n_j}}}\right\}^2$$

$$=\sum_{j=1}^{m}\left\{\left(\frac{\bar{y}_j-\mu}{\frac{\sigma}{\sqrt{n_j}}}\right)^2-\frac{2\left(\bar{y}_j-\mu\right)\left(\bar{y}-\mu\right)}{\left(\frac{\sigma}{\sqrt{n_j}}\right)^2}+\left(\frac{\bar{y}-\mu}{\frac{\sigma}{\sqrt{n_j}}}\right)^2\right\}$$

$$=\sum_{j=1}^{m}\left(\frac{\bar{y}_j-\mu}{\frac{\sigma}{\sqrt{n_j}}}\right)^2-2\left(\bar{y}-\mu\right)\sum_{j=1}^{m}\frac{\left(\bar{y}_j-\mu\right)}{\frac{\sigma^2}{n_j}}+\sum_{j=1}^{m}\frac{\left(\bar{y}-\mu\right)^2}{\frac{\sigma^2}{n_j}}$$

$$=\sum_{j=1}^{m}\left(\frac{\bar{y}_j-\mu}{\frac{\sigma}{\sqrt{n_j}}}\right)^2-2\frac{\left(\bar{y}-\mu\right)}{\sigma^2}\sum_{j=1}^{m}n_j\left(\bar{y}_j-\mu\right)+\frac{1}{\sigma^2}\sum_{j=1}^{m}n_j\left(\bar{y}-\mu\right)^2$$

$$=\sum_{j=1}^{m}\left(\frac{\bar{y}_j-\mu}{\frac{\sigma}{\sqrt{n_j}}}\right)^2-2\frac{\left(\bar{y}-\mu\right)}{\sigma^2}n\left(\bar{y}-\mu\right)+\frac{1}{\sigma^2}n\left(\bar{y}-\mu\right)^2=\sum_{j=1}^{m}\left(\frac{\bar{y}_j-\mu}{\frac{\sigma}{\sqrt{n_j}}}\right)^2-\frac{n\left(\bar{y}-\mu\right)^2}{\sigma^2}$$

$$=\sum_{j=1}^{m}\left(\frac{\bar{y}_j-\mu}{\frac{\sigma}{\sqrt{n_j}}}\right)^2-\left(\frac{\bar{y}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2\sim\chi_{(m-1)}$$

最後の辺りで，$n_j\bar{y}_j$が j 群の合計，よってそれを j=1 から m まで合計したものは全体の合計になる事などに気付く事が重要である。

$$\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(\frac{y_{ij}-\bar{y}_j}{\sigma}\right)^2=\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left\{\frac{\left(y_{ij}-\mu\right)-\left(\bar{y}_j-\mu\right)}{\sigma}\right\}^2$$

$$=\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left\{\left(\frac{y_{ij}-\mu}{\sigma}\right)^2-2\frac{\left(y_{ij}-\mu\right)\left(\bar{y}_j-\mu\right)}{\sigma^2}+\left(\frac{\bar{y}_j-\mu}{\sigma}\right)^2\right\}$$

$$=\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(\frac{y_{ij}-\mu}{\sigma}\right)^2-\frac{2}{\sigma^2}\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left\{y_{ij}\bar{y}_j-\mu\left(y_{ij}+\bar{y}_j\right)+\mu^2\right\}+\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(\frac{\bar{y}_j-\mu}{\sigma}\right)^2$$

$$=\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(\frac{y_{ij}-\mu}{\sigma}\right)^2-\frac{2}{\sigma^2}\sum_{j=1}^{m}\left(n_j\bar{y}_j{}^2-2n_j\bar{y}_j\mu+n_j\mu^2\right)+\sum_{j=1}^{m}n_j\left(\frac{\bar{y}_j-\mu}{\sigma}\right)^2$$

$$=\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(\frac{y_{ij}-\mu}{\sigma}\right)^2-\sum_{j=1}^{m}n_j\left(\frac{\bar{y}_j-\mu}{\sigma}\right)^2=\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(\frac{y_{ij}-\mu}{\sigma}\right)^2-\sum_{j=1}^{m}\left(\frac{\bar{y}_j-\mu}{\frac{\sigma}{\sqrt{n_j}}}\right)^2\sim\chi_{(n-m)}$$

これも最後の辺りで，i を含まない項は i についての Σ の中では定数扱いである事に気付く必要がある。

ここで,「カイ二乗分布に従う統計量÷その自由度」という統計量が二つあったとき,その比が F 分布として定義されることを利用する。

$$F_{(df1,df2)} = \frac{\chi_{(df1)}/df1}{\chi_{(df2)}/df2}$$

平方和(Sum of Squares, *SS*)の分解の式の右辺の第一項と第二項をそれぞれの自由度で割った統計量を**平均平方**(Mean Square, *MS*)という。これは,自由度一つあたりの平方和という意味である。

$$MS_{Between} = \frac{SS_{Between}}{m-1}$$

$$MS_{within} = \frac{SS_{within}}{n-m}$$

そしてこの二つの平均平方の比は,自由度 m-1,n-m の F 分布に従うことが分かる。

$$\frac{MS_{Between}}{MS_{within}} = \frac{SS_{Between}/(m-1)}{SS_{within}/(n-m)} = \frac{(SS_{Between}/\sigma^2)/(m-1)}{(SS_{within}/\sigma^2)/(n-m)} = F_{(m-1,n-m)}$$

母分散が分母と分子で約分されて消え去り,平均平方の比自体はデータから具体的に計算可能な値になっているので,平均平方の比である F 統計量によってゼロ仮説を検定することが可能になる。