自由度m_1の χ 2 乗分布に従うχ_1^2と自由度m_2の χ 2 乗分布に従うχ_2^2の比で構成される確率変数Gを

$$G = \frac{\chi_1^2/(m_1)}{\chi_2^2/(m_2)}$$

がF分布に従うと定義する。

いま、$k \in \{1,2,\ldots,m\}$のm個の母集団分布から、i.i.d.でそれぞれからサイズnの標本Xを得ることを考える。母集団分布は正規分布で、m個の分布は同分散を仮定する。つまり、

$$X^k \sim\ N(\mu_k, \sigma^2)$$

3 つの群それぞれは、標本分布は正規分布の再生性の定理から正規分布となり、

$$\overline{X^k}\ \sim\ N\left(\mu_k, \frac{\sigma^2}{n}\right)$$

となる。各群の標本平均を以下の線形変換を施すと、正規分布の線形変換の性質から、

$$\frac{\overline{X^k} - \mu_k}{\sqrt{\sigma^2/n}} \sim N(0,1)$$

と標準正規分布に従う。また、3 つの群の標準化した標本平均の平方和は、 χ 2 乗分布の定義から、

$$\sum_{k=1}^{m}\left(\frac{\overline{X^k} - \mu_k}{\sqrt{\sigma^2/n}}\right)^2 \sim\ \chi^2(m)$$

自由度mの χ 2 乗分布に従う。

さて、帰無仮説、$\mu_1 = \mu_2 = \cdots = \mu_k$のもとで、各母平均を$\mu_0$と仮定すると、

$$\sum_{k=1}^{m}\left(\frac{\overline{X^k} - \mu_0}{\sqrt{\sigma^2/n}}\right)^2 \sim \chi^2(m)$$

が自由度mの χ 2 乗分布に従う。このμ_0を不偏推定量である全体平均$\bar{X}$で置き換えると、

$$\sum_{k=1}^{m}\left(\frac{\overline{X^k} - \bar{X}}{\sqrt{\sigma^2/n}}\right)^2 \sim \chi^2(m-1)$$

となり、自由度$m-1$の χ 2 乗分布に従う。

続いて、母分散σ^2を不偏分散$U^2 = \frac{1}{m(n-1)}\sum_{k=1}^{m}\sum_{i=1}^{n}\left(X_i - \overline{X^k}\right)^2$に置き換え、$m-1$で割った確率変数

$$F = \frac{1}{m-1}\sum_{k=1}^{m}\left(\frac{\overline{X^k} - \bar{X}}{\sqrt{U^2/n}}\right)^2$$

を、Fと呼ぶことにする。Fは、次の変換により、

$$\frac{1}{m-1}\sum_{k=1}^{m}\left(\frac{\overline{X^k} - \bar{X}}{\sqrt{U^2/n}}\right)^2 = \frac{1}{m-1}\frac{n}{U^2}\sum_{k=1}^{m}\left(\overline{X^k} - \bar{X}\right)^2$$

$$= \frac{1}{m-1}\frac{n}{U^2}\sum_{k=1}^{m}\left(\overline{X^k} - \bar{X}\right)^2\frac{\sigma^2 m(n-1)}{\sigma^2 m(n-1)} = \frac{1}{m-1}\frac{n}{\sigma^2}\sum_{k=1}^{m}\left(\overline{X^k} - \bar{X}\right)^2\frac{\sigma^2 m(n-1)}{U^2 m(n-1)}$$

$$= \frac{n\sum_{k=1}^{m}\left(\overline{X^k} - \bar{X}\right)^2}{\sigma^2(m-1)} \Big/ \frac{m(n-1)U^2}{\sigma^2 m(n-1)} \ = \frac{n\sum_{k=1}^{m}\left(\frac{\overline{X^k} - \bar{X}}{\sqrt{\sigma^2/n}}\right)^2}{(m-1)} \Big/ \frac{\sum_{k=1}^{m}\sum_{i=1}^{n}\left(\frac{X_i - \overline{X^k}}{\sqrt{\sigma^2}}\right)^2}{m(n-1)}$$

となる。分子は自由度$M-1$の χ 2 乗分布に従うχ_1^2を$(m-1)$で割ったもの、分母は自由度$m(n-1)$の χ 2

乗分布に従うχ_2^2を$m(n-1)$で割ったものであり、その比Fは定義より、F分布に従う。

証明終わり

Fについて、標本Xの実現値xを用いて、Fの実現値であるf値を

$$f=\frac{1}{m-1}\sum_{k=1}^{m}\left(\frac{\overline{x^k}-\bar{x}}{\sqrt{u^2/n}}\right)^2=\frac{n\sum_{k=1}^{m}\left(\overline{x^k}-\bar{x}\right)^2}{(m-1)}\Big/\frac{\sum_{k=1}^{m}\sum_{i=1}^{n}\left(x_i-\overline{x^k}\right)^2}{m(n-1)}$$

と計算できる。

よって、f値は、標本から計算でき、群間の平均平方と郡内の平均平方の比で計算できる。