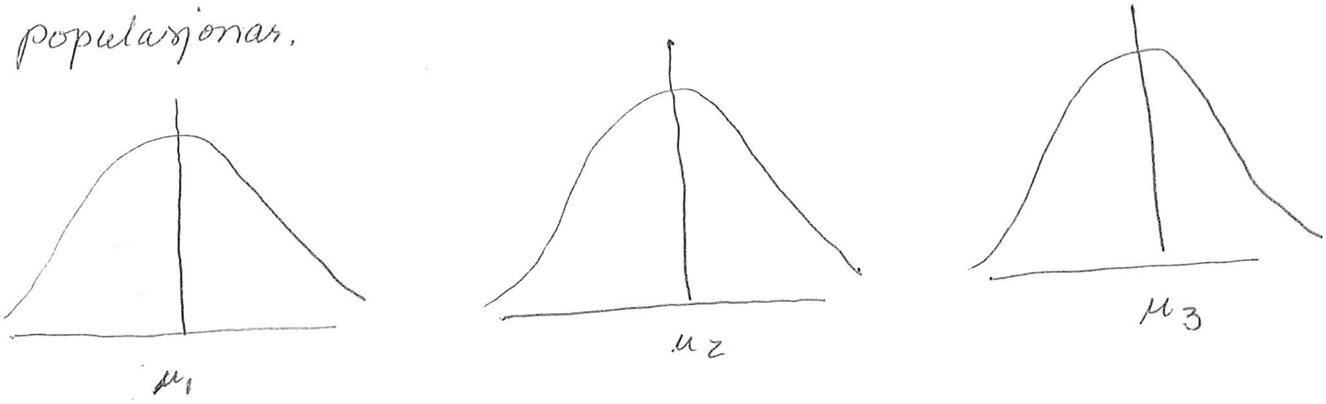
Ronald Aylmer Fisher 1890 – 1962
Father of modern statistics
greatest of Darwin's successors

Kap 12. Variansanalyse

Korleis ein kan gjere slutningar om meir enn to populasjonar.



Generell modell

La $Y_{ij} = \mu_j + \varepsilon_{ij} \left\{ \begin{array}{l} N(0,\sigma^2) \text{ og uavh.} \\ i = 1,2,\dots,n_j, \quad j = 1,2,\dots,k \end{array} \right.$

Meir vanleg

La $\mu = \sum_{j=1}^{k} n_j \mu_j \Big/ \sum_{j=1}^{k} n_j$

Vi får $\mu_j = \mu + \underbrace{\mu_j - \mu}_{\alpha_j} = \mu + \alpha_j$

der $\sum_{j=1}^{k} n_j \alpha_j = \sum_{j=1}^{k} n_j(\mu_j - \mu) = \sum_{j=1}^{k} n_j\mu_j - \sum_{j=1}^{k} n_j \mu$

$= \sum_{j=1}^{k} n_j \mu_j - \sum_{j=1}^{k} n_j \mu_j = 0$

Dette gjev $Y_{ij} = \mu + \alpha_j + \varepsilon_{ij} \left\{ \begin{array}{l} N(0,\sigma^2) \text{ og uavh.} \\ i = 1,2,\dots,n_j, \quad j = 1,2,\dots,k. \end{array} \right.$

<u>Dekomponering av variasjon</u>

La $\bar{Y}_{\cdot j} = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij}$ og $\bar{Y}_{\cdot\cdot} = \frac{1}{\sum_{j=1}^{k} n_j} \sum_{j=1}^{k} \sum_{i=1}^{n_j} Y_{ij}$

Vi får $Y_{ij} = \bar{Y}_{..} + \underbrace{\bar{Y}_{.j} - \bar{Y}_{..}}_{\text{estimator for } \alpha_j} + Y_{ij} - \bar{Y}_{.j}$

For totalvariasjonen får vi:

$$\sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{..})^2 = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{.j} + \bar{Y}_{.j} - \bar{Y}_{..})^2$$

$$= \sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{.j})^2 + \sum_{j=1}^{k} \sum_{i=1}^{n_j} (\bar{Y}_{.j} - \bar{Y}_{..})^2 + 2 \sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{.j})(\bar{Y}_{.j} - \bar{Y}_{..})$$

Variasjon innan grupper

Variasjon mellom grupper

Vi har $\sum_{j=1}^{k} (\bar{Y}_{.j} - \bar{Y}_{..}) \underbrace{\sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{.j})}_{0}$

slik at $$\sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{..})^2 = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{.j})^2 + \sum_{j=1}^{k} \sum_{i=1}^{n_j} (\bar{Y}_{.j} - \bar{Y}_{..})^2$$

eller

$$SSTOT = SSE + SSTR$$ (Teorem 12.2.4

SSE og SSTR

da $S_j^2 = \dfrac{\sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{.j})^2}{n_j - 1}$, S_j^2, $j = 1, \dots, k$ er uavh.

og $\dfrac{(n_j - 1) S_j^2}{\sigma^2} \sim \chi^2(n_j - 1)$, S_j^2, $j = 1, \dots, k$ er uavh.

$\dfrac{SSE}{\sigma^2} = \sum_{j=1}^{k} (n_j - 1) \dfrac{S_j^2}{\sigma^2} \sim \chi^2\left(\sum_{j=1}^{k} (n_j - 1)\right) \sim \chi^2(n - k)$ (der $n = \sum_{j=1}^{k} n_j$) Teorem 12.2.3

$\bar{Y}_{.j}$ uavh av S_j^2 og S_i^2, $i \neq j$. $\Rightarrow$ SSE og SSTR er uavh

Hypoteser

$H_0: \mu_1 = \mu_2 = \cdots = \mu_k = \mu$

$\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$

H_1: Minst 2 er forskjellige

H_1: minst ein α_j, $j=1,\dots,k$ er $\neq 0$.

Går ut i frå at H_0 er sann.

$\bar{Y}_{..}$ er då ein estimator for $E[Y_{ij}] = \mu$ $i=1,2,\dots,n_j$, $j=1,\dots,k$.

og $\dfrac{SSTOT}{\sigma^2} = \dfrac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij} - \bar{Y}_{..})^2}{\sigma^2}$ er $\chi^2(n-1)$, $\left(\text{der } n = \sum_{j=1}^{k} n_j\right)$

$$SSTOT = \underbrace{SSE + SSTR}_{\text{uavh}} \Rightarrow M_{\frac{SSTOT}{\sigma^2}}(t) = M_{\frac{SSE}{\sigma^2}}(t) \cdot M_{\frac{SSTR}{\sigma^2}}(t)$$

$$M_{\frac{SSTOT}{\sigma^2}}(t) = \left(\frac{1}{1-2t}\right)^{\frac{(n-1)}{2}}, \quad M_{\frac{SSE}{\sigma^2}}(t) = \left(\frac{1}{1-2t}\right)^{\frac{n-k}{2}}$$

$$\Rightarrow M_{\frac{SSTR}{\sigma^2}}(t) = \left(\frac{1}{1-2t}\right)^{\left(\frac{n-1}{2} - \frac{n-k}{2}\right)} = \left(\frac{1}{1-2t}\right)^{\frac{k-1}{2}}$$

$$\text{d: } \frac{SSTR}{\sigma^2} \sim \chi^2(k-1)$$ (Teorem 12.2.2)

<u>SSTR dersom H_1 er sann</u>

$$SSTR = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(\bar{Y}_{.j} - \bar{Y}_{..})^2 = \sum_{j=1}^{k} n_j(\bar{Y}_{.j} - \bar{Y}_{..})^2$$

$$= \sum_{j=1}^{k} n_j\left[(\bar{Y}_{.j} - \mu) - (\bar{Y}_{..} - \mu)\right]^2 = \sum_{j=1}^{k} n_j(\bar{Y}_{.j} - \mu)^2 + \sum_{j=1}^{k} n_j(\bar{Y}_{..} - \mu)^2$$

$$- 2(\bar{Y}_{..} - \mu)\sum_{j=1}^{k} n_j(\bar{Y}_{.j} - \mu)$$

$$\parallel$$

$$- 2(\bar{Y}_{..} - \mu)\left(\sum_{j=1}^{k} n_j \bar{Y}_{.j} - \sum_{j=1}^{k} n_j \mu_j\right)$$

$$\parallel$$

$$n\bar{Y}_{..} - n\mu = n(\bar{Y}_{..} - \mu)$$

Dette gir $SSTR = \sum_{j=1}^{n} n_j(\bar{Y}_{.j} - \mu)^2 + n(\bar{Y}_{..} - \mu)^2 - 2n(\bar{Y}_{..} - \mu)^2$

$$= \sum_{j=1}^{k} n_j(\bar{Y}_{.j} - \mu)^2 - n(\bar{Y}_{..} - \mu)^2$$

$$E[SSTR] = \sum_{j=1}^{k} n_j E[\bar{Y}_{.j} - \mu]^2 - nE[\bar{Y}_{..} - \mu]^2$$

$$E[\bar{Y}_{.j} - \mu]^2 = Var[\bar{Y}_{.j} - \mu] + (E[\bar{Y}_{.j} - \mu])^2 = \frac{\sigma^2}{n_j} + (\mu_j - \mu)^2$$

$$E[\bar{Y}_{..} - \mu]^2 = Var[\bar{Y}_{..}] = \frac{\sigma^2}{n}$$

and $E[SSTR] = \sum_{j=1}^{k} n_j\left(\frac{\sigma^2}{n_j} + (\mu_j - \mu)^2\right) - n \cdot \frac{\sigma^2}{n}$

$$= k\sigma^2 + \sum_{j=1}^{k} n_j(\mu_j - \mu)^2 - \sigma^2 = (k-1)\sigma^2 + \sum_{j=1}^{k} n_j(\mu_j - \mu)^2$$

Teorem 12.2.1

$$E\left[\frac{SSE}{n-k}\right] = \sigma^2 E\left[\frac{SSE}{\sigma^2} \cdot \frac{1}{n-k}\right] = \sigma^2 E\left[\chi^2(n-k) \cdot \frac{1}{n-k}\right] = \sigma^2 \text{ alltid.}$$

$$E\left[\frac{SSTR}{k-1}\right] = \sigma^2 E\left[\frac{SSTR}{\sigma^2} \cdot \frac{1}{k-1}\right] = \sigma^2 \cdot E\left[\chi^2(k-1) \cdot \frac{1}{k-1}\right] = \sigma^2 \text{ dersom}$$

H_0 er sann.

SSE og SSTR er uavh. Dersom H_0 er sann er

$$F = \frac{\frac{SSTR}{k-1}}{\frac{SSE}{n-k}} = \frac{\sigma^2 \cdot \frac{SSTR}{\sigma^2(k-1)}}{\sigma^2 \cdot \frac{SSE}{\sigma^2(n-k)}} \sim \frac{\frac{\chi^2(k-1)}{k-1}}{\frac{\chi^2(n-k)}{n-k}} \sim F_{k-1,\,n-k}.$$

$F \approx 1$ dersom H_0 er sann og forventa > 1 dersom H_0 ikkje er sann

o: Forkast: $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ dersom $F \geq F_{1-\alpha,\,k-1,\,n-k}$.