Привет ☺

ANOVA-2

$y$ - непрерывная

$A$, $B$ — два дискретных фактора.

| A \ B | 1 | 2 | ... | $n_B$ |
|---|---|---|---|---|
| 1 | | $y_{121}$ <br> $y_{122}$ | | |
| 2 | | | | |
| ⋮ | | | | |
| $n_A$ | | | | |

В каждой клетке таблицы живут несколько

**Данные**

ANOVA-style

$y_{ijk}$ — $i$ по фактору $A$; $j$ по фактору $B$; $k$ номер внутри клетки.

Regression-style

$y_i$ — $i$-номер наблюдения

| $y_i$ | $d_i(a=1)$ | | $d_i(a=n_a-1)$ | $d_i(b=1)$ | | | $d_i(b=n_b-1)$ |
|---|---|---|---|---|---|---|---|
| | 1 | 0 | 0 | 0 | | | |
| | 0 | 1 | 0 | 0 | 1 | 0 | 0 |

Модель без взаим-свия

$$y_{ijk} = \mu + \alpha_i + \beta_j + u_{ijk}$$

$\mu, \alpha_i, \beta_j$ — const

$u_{ijk} \sim N(0; \sigma^2)$ незав

предп. идент.

$$\boxed{\sum_{i=1}^{n_a} \alpha_i = 0 \quad \sum_{j=1}^{n_b} \beta_j = 0}$$

$$y_i = \delta_1 + \gamma_1 d_i(a=1) + \ldots + \gamma_{n_a-1} d_i(a=n_a-1)$$
$$+ w_1 \cdot d_i(b=1) + \ldots + w_{n_b-1} \cdot d_i(b=n_b-1)$$
$$+ u_i$$

$u_i \sim N(0; \sigma^2)$ незав

всего пар-в:

$$\underset{\mu}{1} + \underset{\alpha_i}{(n_A - 1)} + \underset{\beta_j}{(n_B - 1)} + \underset{\sigma^2}{1}$$

$$\underset{\delta_1}{1} + \underset{\gamma_j}{(n_a - 1)} + \underset{w_j}{(n_b - 1)} + \underset{\sigma^2}{1}$$

$$\hat{y}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j$$

residual $\hat{u}_{ijk} = y_{ijk} - \hat{y}_{ijk}$

$$\hat{y}_i = \hat{\delta}_1 + \hat{\gamma}_1 \cdot d_i(a=1) + \ldots$$
$$+ \hat{w}_1 \cdot d_i(b=1) + \ldots$$
$$\hat{u}_i = y_i - \hat{y}_i$$

Q. Зачем условия идент-сти в ANOVA?

A. чтобы не было ситуации, когда разные парам-ры по сути означают одну и ту же модель.

Маша | Вовочка
---|---

| A \ B | $\beta=1$ | $\beta=2$ | $\beta=3$ |
|---|---|---|---|
| $a=1$ | • | • | • |
| $a=2$ | • | • | • |

| A \ B | $\beta=1$ | $\beta=2$ | $\beta=3$ |
|---|---|---|---|
| $a=1$ | • | • | • |
| $a=2$ | • | • | • |

Маша:

$\mu=5$, $\alpha_1=2$, $\alpha_2=3$

$\beta_1=2$, $\beta_2=6$, $\beta_3=4$.

$$y_{11k}=5+2+2+u_{11k}$$

$$y_{11k}=9+u_{11k}$$

Вовочка:

$\mu=11.5$ ($5+2.5+4$), $\alpha_1=-0.5$, $\alpha_2=0.5$

$\beta_1=-2$, $\beta_2=2$, $\beta_3=0$

$$y_{11k}=11.5-0.5-2+u_{11k}$$

$$y_{11k}=9+u_{11k}.$$

$$y_{ijk}=\mu+\underset{\Sigma}{\alpha_i}+\underset{\Sigma}{\beta_j}+u_{ijk}$$

| A \ B | $\beta=1$ | $\beta=2$ | $\beta=3$ |
|---|---|---|---|
| $a=1$ | • • | • • | • • |
| $a=2$ | • • | • • | • • |

$$\text{TotSS}=\sum_{ijk}(y_{ijk}-\bar{y})^2$$

$$\text{SSA}=\sum_{ijk}(\bar{y}_{i\cdot}-\bar{y})^2=n_B\cdot K\cdot\sum_i(\bar{y}_{i\cdot}-\bar{y})^2$$

$$\text{SSB}=\sum_{ijk}(\bar{y}_{\cdot j}-\bar{y})^2=n_A\cdot K\cdot\sum_j(\bar{y}_{\cdot j}-\bar{y})^2$$

$$\text{ResSS}=\sum_{ijk}(y_{ijk}-\hat{y}_{ijk})^2$$

Error SS

$\bar{y}$ – общее среднее

$$\bar{y}=\frac{\sum_{ijk}y_{ijk}}{n}$$

$$\bar{y}_{\cdot j}=\frac{\sum_{ik}y_{ijk}}{n_A\cdot K}$$

по всем элементам столбца j взяли среднее

$$\bar{y}_{i\cdot}=\bar{y}_{i\cdot\cdot}=\frac{\sum_{jk}y_{ijk}}{n_B\cdot K}$$

по всем элементам строки i взяли среднее

K – кол-во элементов в отдельной клетке

Т. Пифагора
по ст. своб.

$$TotSS = SSA + SSB + ResSS$$
$$n-1 = (n_A-1)+(n_B-1)+(n-n_A-n_B+1)$$

11

$\hat{\mu}$ $\hat{\alpha}_i$ $\hat{\beta}_j$ как получить?

$$\hat{\mu} = \bar{y}$$
$$\hat{\alpha}_i = \bar{y}_{i\cdot} - \bar{y}$$
$$\hat{\beta}_j = \bar{y}_{\cdot j} - \bar{y}$$

$$SSA = \sum_{ijk} \hat{\alpha}_i^2$$
$$SSB = \sum_{ijk} \hat{\beta}_j^2$$

$$\ln L = \sum_{ijk} \ln f(y_{ijk} \mid \theta)$$

$$y_{ijk} \sim N(\mu+\alpha_i+\beta_j;\ \sigma^2)$$
$$\theta = (\mu, \alpha_1, \ldots, \alpha_{n_a}, \beta_1 \ldots \beta_{n_B}, \sigma^2)$$
$$\sum \alpha_i = 0 \quad \sum \beta_j = 0$$

$$\ln f(y_{ijk} \mid \theta) = \ln \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2}(y_{ijk}-\mu-\alpha_i-\beta_j)^2$$

Гипотезы

Фактор A не влияет:

$H_0: \alpha_1 = \alpha_2 = \ldots \alpha_{n_a} = 0$

$H_A$: хотя бы один $\alpha_i \neq 0$.

$$F = \frac{SSA/(n_A-1)}{ResSS/(n-n_A-n_B+1)} \sim F_{n_A-1,\ n-n_A-n_B+1}$$

Фактор B не влияет

$H_0: \beta_1 = \beta_2 = \ldots \beta_{n_B} = 0$

$H_A$: хотя бы один $\beta_j \neq 0$

$$F = \frac{SSB/(n_B-1)}{ResSS/(n-n_A-n_B+1)} \sim F_{n_B-1,\ n-n_A-n_B+1}$$

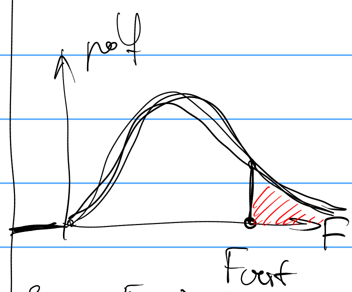
Ни фактор A, ни фактор B не влияют

$H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_{n_a} = 0,\ \beta_1 = \beta_2 = \ldots = \beta_{n_B} = 0$

$H_A$: хотя бы один $\alpha_i$ или хотя бы один $\beta_j$ не равен нулю.

$$F = \frac{(SSA+SSB)/(n_A+n_B-2)}{ResSS/(n-n_A-n_B+1)} \sim F_{n_A+n_B-2,\ n-n_A-n_B+1}$$

$F_{obs}$ v $F_{crit}$




если $F_{obs} > F_{crit}$ то $H_0$ отвергается.

## Нос. оценки регрессии

Un Restr. $y_i = \delta_1 + \nu_1 \cdot d_i(a=1) + \ldots \nu_{n_a-1} \cdot d_i(a=n_a-1) + w_1 d_i(b=1) + \ldots w_{n_b-1} d_i(b=n_b-1) + u_i$

МНК

ResSS$_{UR}$

фактор А не влияет

$H_0: \nu_1 = \nu_2 = \ldots \nu_{n_a-1} = 0$

$H_A$: хотя бы одна $\nu_j \neq 0$.

$$F = \frac{(RSS_R - RSS_{UR})/(n_a-1)}{RSS_{UR}/(n - \underbrace{1-(n_A-1)-(n_B-1)}_{\text{число коэф.}})}$$

число набл.

Restr. $y_i = \delta_1 + w_1 \cdot d_i(b=1) + \ldots + w_{n_b-1} d_i(b=n_b-1) + u_i$

ResSS$_R$

Фактор В (по аналогии)

$H_0$: ни один фактор не влияет все $\nu_j = w_j = 0$.

$H_A$: хотя бы один $\nu_j$ или $w_j \neq 0$.

Restr. $y_i = \delta_1 + u_i$ $\quad RSS_R = TSS$

$$F = \frac{(TSS - RSS_{UR})/(n_a - 1 + n_b - 1)}{RSS_{UR}/(n - n_a - n_b + 1)}$$

# ANOVA со взаимодействиями

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + u_{ijk}$$

$u_{ijk} \sim N(0; \sigma^2)$ незав.

идент-сть: $\sum \alpha_i = 0 \quad \sum \beta_j = 0$

$\sum_i \gamma_{ij} = 0 \quad \sum_j \gamma_{ij} = 0$

$$TotSS = SSA + SSB + SSAB + ResSS$$

df: $n-1 \quad n_A - 1 \quad n_B - 1 \quad (n_A-1)(n_B-1) \quad ? = (n-1) - (n_A-1) - (n_B-1) - (n_A-1)(n_B-1)$

$$SSAB = \sum_{ijk} \left( \underbrace{(\bar{y}_{ij} - \bar{y})} - \underbrace{(\bar{y}_{i\cdot} - \bar{y})} - \underbrace{(\bar{y}_{\cdot j} - \bar{y})} \right)^2$$

смысл

$$= \sum_{ijk} (\bar{y}_{ij} + \bar{y} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j})^2$$

Новая $H_0$: нет взаим-ия между факторами A и B

$H_A$: есть взаим-ие

$H_0$: все $\gamma_{ij} = 0$

$H_A$: хотя бы один $\gamma_{ij} \neq 0$.

$$F \overset{?}{=} \frac{SSAB / ((n_A-1)(n_B-1))}{ResSS / (n - n_A n_B)} \sim F_{(n_A-1)(n_B-1),\, n - n_A n_B}$$

$$df(ResSS) = n - 1 - (n_A - 1) - (n_B - 1) - (n_A-1)\cdot(n_B-1) =$$

$$= n - \cancel{1} - \cancel{n_A} + \cancel{1} - \cancel{n_B} + \cancel{1} - n_A n_B + \cancel{n_B} + \cancel{n_A} - \cancel{1} = n - n_A n_B$$

всего голов — мат. ож-ия для ячеек.