ANOVA (1way, 2way)

ANOVA 1way
ANOVA 2way → regression (opinion)

Gelman ANOVA 1way
ANOVA 2way → multilevel models !!

ANOVA 1-way. [ANalysis Of VAriance]

one continuous var-l y
one discrete var-l a

ANOVA-setup

a	[1] Green	[2] Blue	[3] Red
observ	y_{11}	y_{21}	y_{31}
	y_{12}	y_{22}	$\vdots$
	$\vdots$	$\vdots$	
	y_{1n_1}	y_{2n_2}	y_{3n_3}

Regression setup.

y_i	a_i
y_1	a_1
y_2	a_2
y_3	a_3
$\vdots$	
y_n	a_n

Model

classic ANOVA language

$$y_{ij} = \mu_i + u_{ij}$$
$$u_{ij} \sim N(0;\sigma^2) \text{ iid}$$

regression language

$$y_i = \mu_1 \cdot I(a_i=1) + \mu_2 \cdot I(a_i=2) + \dots + \mu_{n_a} \cdot I(a_i=n_a) + u_i$$
$$u_i \sim N(0;\sigma^2) \text{ iid}$$

Question

$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_{n_a}$
H_1: At least one eq is broken

—//—

Anova - language.

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

$\bar{y}_{i\cdot}$ ← the average value of y_{ij} for all j $[a = i]$

$$\bar{y}_{i\cdot} = \frac{\sum_j y_{ij}}{n_i}$$

← total

$$\text{Tot } SS = \sum_{ij} (y_{ij} - \bar{y})^2$$

$$\begin{pmatrix}\text{Between } SS\\ \text{Treat } SS\end{pmatrix} = \sum_{ij} (\bar{y}_{i\cdot} - \bar{y})^2 = \sum_i n_i (\bar{y}_{i\cdot} - \bar{y})^2$$

$$\begin{pmatrix}\text{Within } SS\\ \text{Error } SS\end{pmatrix} = \sum_{ij} (y_{ij} - \bar{y}_{i\cdot})^2$$

$H_0: \mu_1 = \mu_2 = \dots = \mu_{n_a}$

$H_A: \exists\, \mu_i \neq \mu_j$

$$F = \frac{\text{Betw } SS \;/(n_a - 1)}{\text{With } SS \;/(n - n_a)}$$

$$\text{Tot } SS = \text{With } SS + \text{Bet } SS$$

Regression language.

unrestricted (UR-model) $[n_a = 3]$

$$y_i = \mu_1 \cdot I(a_i = 1) + \mu_2 \cdot I(a_i = 2) + \mu_3 \cdot I(a_i = 3) + u_i$$

→ $\text{Res } SS_{UR}$

Restricted (R-model)

$$y_i = \mu + u_i$$

$\text{Res } SS_R = TSS = \sum (y_i - \hat{\mu})^2 =$

$\text{Tot } SS_{UR} = \sum (y_i - \bar{y})^2$

$\text{Res } SS$ = residual sum of squares $= \sum_i (y_i - \hat{y}_i)^2$

$\text{Exp } SS_{UR}$

$\text{Res } SS_{UR}$

$$F = \frac{(\text{Res } SS_R - \text{Res } SS_{UR})/(p_{UR} - p_R)}{\text{Res } SS_{UR}/(n - p_{UR})}$$

p_R - number of coeff in R-model.

p_{UR} - — // — in UR-model

number of est-d coeff-ts:

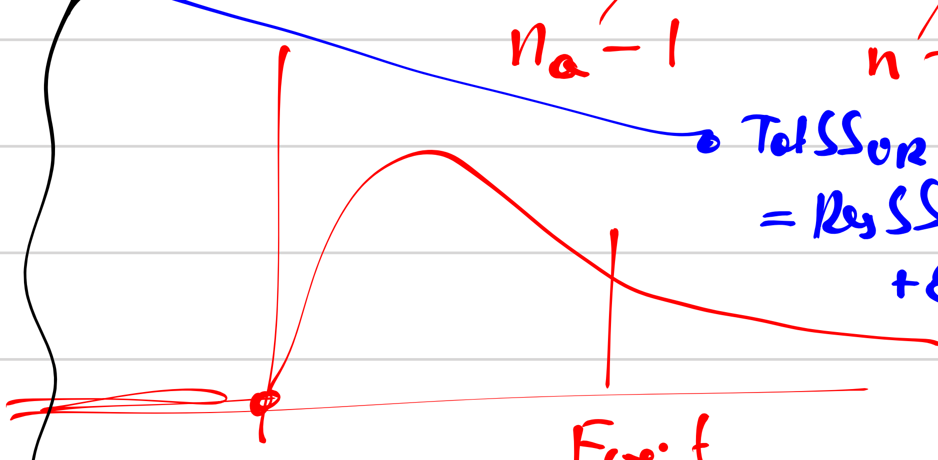
$p_R = 1 \quad [\mu]$

$p_{UR} = n_a \quad [\mu_1, \mu_2, \mu_3 \dots \mu_{n_a}]$

$$F \overset{H_0}{\sim} F_{p_{UR} - p_R,\; n - p_{UR}}$$

$p_{UR} - p_R$ ← $n_a - 1$; $n - p_{UR}$ ← $n - n_a$

$\text{Tot } SS_{UR} = \text{Res } SS_{UR} + \text{Expl } SS_{UR}$



If $F > F_{crit}$ then $H_0: \mu_1 = \mu_2 = \dots$ is rejected

Two-way ANOVA.

Two discrete var-s

$a \in \{1, 2, 3, \dots n_a\}$
$b \in \{1, 2, 3 \dots n_a\}$

$y \in \mathbb{R}$

Model without interactions.

classic ANOVA lang. | regression language.

	$b=1$	$b=2$	$b=n_b$
$a=1$	y_{111} $y_{112}\dots$	y_{121} $y_{122}\dots$	
$a=2$			
$\vdots$			
$a=n_a$			

Balanced case: K obs-ns in every cell

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

$u_{ijk} \sim N(0; \sigma^2)$ iid

$\sum_i \alpha_i = 0 \quad \sum_j \beta_j = 0$

example

$n_a = 2 \quad n_b = 2$

$\mu = 5 \quad \alpha_1 = 2 \quad \alpha_2 = 3$
$\beta_1 = 5 \quad \beta_2 = 6$

$y_{ijk} = 5 + 2.5 + 5.5 + (\alpha_i - 2.5) + (\beta_j - 5.5) + u_{ijk}.$

identification condition
$\sum \alpha_i = 0 \quad \sum \beta_j = 0$

$y_{ijk} = \tilde{\mu} + \tilde{\alpha}_i + \tilde{\beta}_j + u_{ijk}$
$\sum \tilde{\alpha}_i = 0 \quad \sum \tilde{\beta}_j = 0$

„dummy trap"

$n_a = 3$
$n_b = 4$

$y_i = w +$
$\delta_2 \cdot I(a_i = 2) + \delta_3 \cdot I(a_i = 3) +$
$+ \rho_2 I(b_i = 2) + \rho_3 I(b_i = 3) +$
$\rho_4 \cdot I(b_i = 4) + u_i$

Base level: $a=1 \quad b=1$

$u_i \sim N(0; \sigma^2)$ iid

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

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

[for balanced case]

$= K \cdot n_j \sum_i (\bar{y}_{i\cdot} - \bar{y})^2$

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

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

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

regr

Anova lang

$$\hat{\mu} = \frac{\sum y_{ijk}}{n}$$

$$\hat{\alpha}_i = \bar{y}_{i\cdot} - \bar{y}$$

$$\hat{\beta}_j = \bar{y}_{\cdot j} - \bar{y}$$

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

Pyth:

$$Tot\,SS = SSA + SSB + Res\,SS$$

regr. lang

„A does not influence y"

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

H_A: not all α_i are 0.

B does not influence y

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

H_A: not all β_j are 0

neither A, neither B

H_0: $\alpha_1 = \dots = \alpha_{n_a} = \beta_1 = \dots = \beta_{n_B} = 0$

H_A: At least one α_i or one β_j is not zero.

$n_a = 3$

$n_B = 4$

$$y_i = w + \delta_2 \cdot I(a_i = 2) + \delta_3 \cdot I(a_i = 3) + \rho_2 I(b_i = 2) + \rho_3 I(b_i = 3) + \rho_4 \cdot I(b_i = 4) + u_i$$

„A does not influence y"

H_0: $\delta_2 = \delta_3 = \dots \delta_{n_a} = 0$

H_A: not all $\delta_i = 0$.

B does not influency y

H_0: $\rho_2 = \rho_3 = \dots = \rho_{n_B} = 0$

H_A: not all $\rho_j = 0$.

$$F_A = \frac{SSA / (n_A - 1)}{Res\,SS / (n - n_A - n_B + 1)}$$

?

$$\frac{(Res\,SS_R - Res\,SS_{UR}) / (p_{UR} - p_R)}{Res\,SS_{UR} / (n - p_{UR})}$$

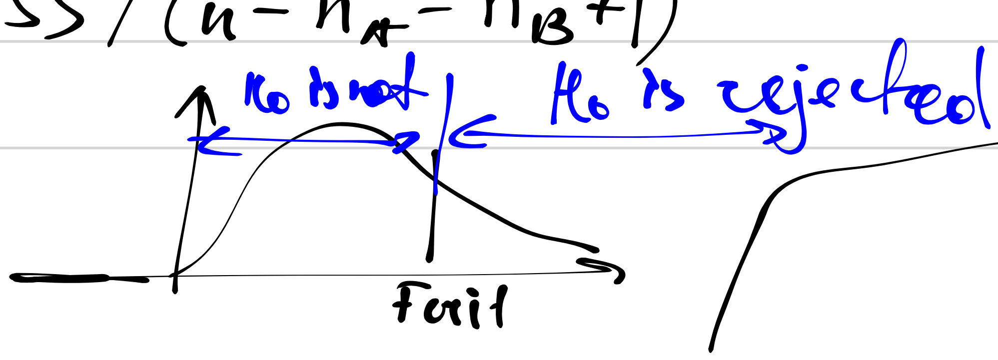
$p_{UR} = 1 + (n_a - 1) + (n_B - 1)$

$w \quad \delta_2, \delta_3 \dots \quad \rho_2, \rho_3 \dots$

$p_{UR} = n_a + n_B - 1$

$$F_B = \frac{SSB / (n_B - 1)}{Res\,SS / (n - n_A - n_B + 1)}$$

$$F_{AB} = \frac{(SSA + SSB) / (n_A + n_B - 2)}{Res\,SS / (n - n_A - n_B + 1)}$$



$b = j$

$a = i \rightarrow \boxed{y_{ij1} \quad y_{ij2} \; \ldots \; y_{ijk}}$

2-way ANOVA with interactions.

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

$$u_{ijk} \sim N(0; \sigma^2)$$

identification constr:

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

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

$$SSAB = \sum_{ijk} \left((\bar{y}_{ij} - \bar{y}) - \hat{\alpha}_i - \hat{\beta}_j \right)^2$$

Hypo: „No interactions between influence of A and influence B"

H_0: $\gamma_{11} = \gamma_{12} = \ldots\ldots = \gamma_{n_a n_b} = 0$

H_A: not all γ_{ij} are zero

Pyth:

$$TotSS = SSA + SSB + \underline{SSAB} + ResSS$$

$$F_{stat} = \frac{SSAB / (n_A - 1)\cdot(n_B - 1)}{ResSS / (n - n_A \cdot n_B)}$$