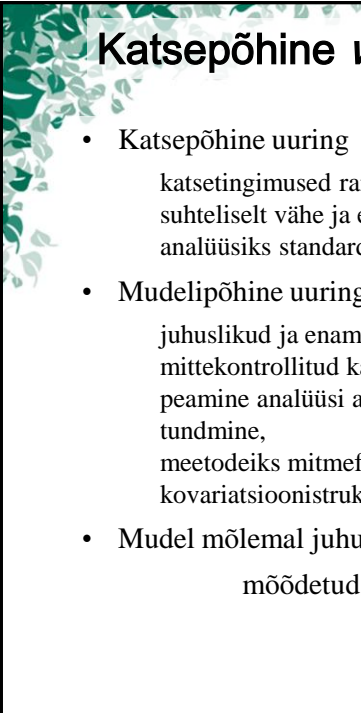


Matemaatiline statistika ja modelleerimine

Üldised lineaarsed mudelid [*general linear models, GLM*]

EMÜ doktorikool
DK.0007

Tanel Kaart



Katsepõhine vs mudelipõhine uuring

- Katsepõhine uuring
 - katsetingimused range kontrolli all, suhteliselt vähe ja enamasti tasakaalus [*balanced*] andmed, analüüsiks standardne regressioon- või dispersioonanalüüs (t-test).
- Mudelipõhine uuring
 - juhuslikud ja enamasti mittetasakaalus [*unbalanced*] andmed, mittekontrollitud katsetingimused, peamine analüüsi alus on uurija intuitsioon/analüüsitava materjali tundmine, meetodeiks mitmefaktorilised, sageli mittestandardse vaatluste kovariatsioonistruktuuriga mudelid.
- Mudel mõlemal juhul
 - mõõdetud väärtus = sobitatud väärtus + viga.

Sõltuvad ja sõltumatud tunnused

- **Uuritavad e sõltuvad tunnused** [*dependent variables*] – tunnused, mille käitumine huvi pakub.
- **Argument- e sõltumatud tunnused** [*independent variables*] e **faktorid** – tunnused, mille mõju uuritavatele tunnustele soovitakse selgitada.
- Faktortunnuse erinevaid väärtusi nimetatakse **tasemeteks e nivoodeks** [*levels*].
- Iga faktor jaguneb vastavalt oma tasemete iseloomule **diskreetseks** või **pidevaks, arviliseks** (kvantitatiivne) või **klassifitseerivaks** (kvalitatiivne).

Näiteks on lehma sünniaasta, laktatsioon jne diskreetsed arvilised faktorid;
farm tasemetega (väärtustega) 'Vorbuse', 'Ülenurme' jne on diskreetne klassifitseeriv faktor;
laktatsiooni pikkus, piimatoodang, pekipaksus jne (mõõdetud tunnused) on aga pidevad faktorid.

Faktorite vahekorid mudelis

- Faktorid on **lihtsad ja tuletatud**. Lihtsate faktorite väärtused on vahetult mõõdetud või registreeritud, tuletatud faktorid moodustatakse lihtsatest. Tüüpilised tuletatud faktorid on **interaktsioonid e** koosmõjud ning arviliste faktorite korrutised.

Näiteks on farm, isa ja laktatsiooni pikkus lihtsad faktorid; farm*isa (koosmõju) ja laktatsioon*laktatsioon (arvilise faktori kõrgem järk) aga tuletatud faktorid.

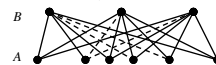
- Praktikas on faktorite vahel sageli ka **alluvusseosed**. Faktor A allub faktorile B , kui A iga nivoo (tase) esineb koos vaid ühe B nivooga.

Näiteks võime me tavaliselt lugeda farmi allutatuks maakonnale; kui iga ema on ristatud kindla isaga, on ema allutatud isale.



- Faktorid A ja B on **ristseoses**, kui A iga nivoo kombineerub (saab põhimõtteliselt kombineeruda) B kõigi nivoodega.

Näiteks kui viiel aastal on uuritud pullide tütarde jõudlusandmeid ja igal pullil on igal aastal tütreid, on pull ja aasta ristseoses; kui aga igal aastal on valitud uued pullid, allub pull aastale.



Lineaarsed mudelid

- Lineaarne mudel sisaldab komplekti faktoreid, mis mõjutavad vaatlusi aditiivselt, kusjuures mingi muutuja faktori siseselt võib olla näiteks ruutu võetud.
- Lineaarseid mudeleid sobib rakendada enamustes bioloogilistes uuringutes.
- Mittelineaarsed seosed on tihti lähendatavad lineaarse mudeliga.
- Traditsiooniline lineaarne mudel koosneb kolmest osast:
 - võrrand – mudeli esitus faktorite mõjude summana;
 - juhuslike muutujate keskvärtused ja dispersioonistruktuur;
 - eeldused, kitsendused ja piirangud.

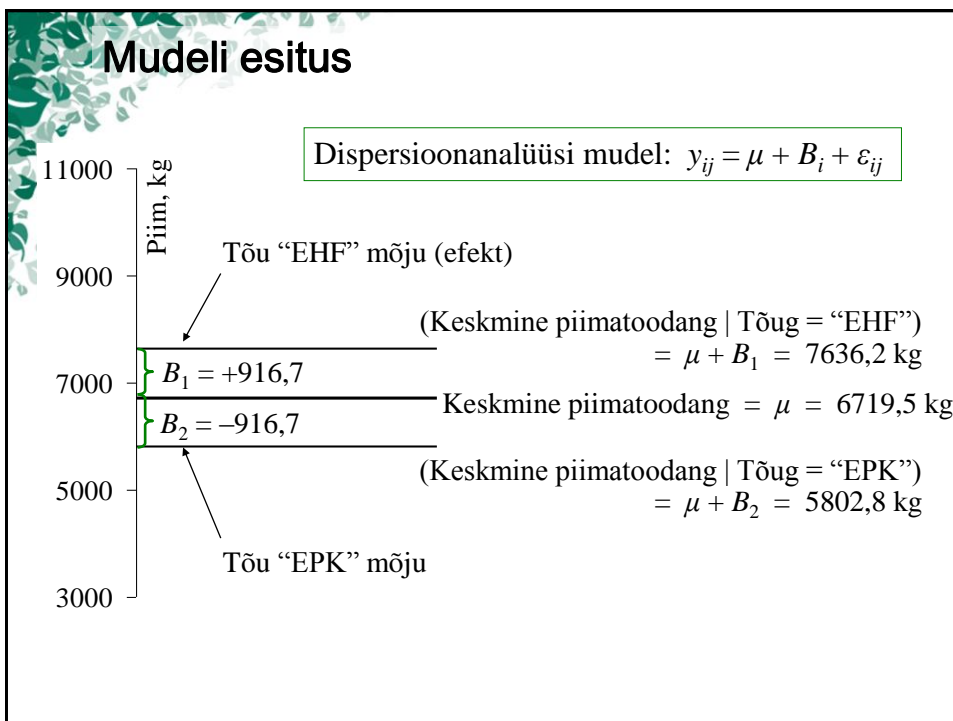
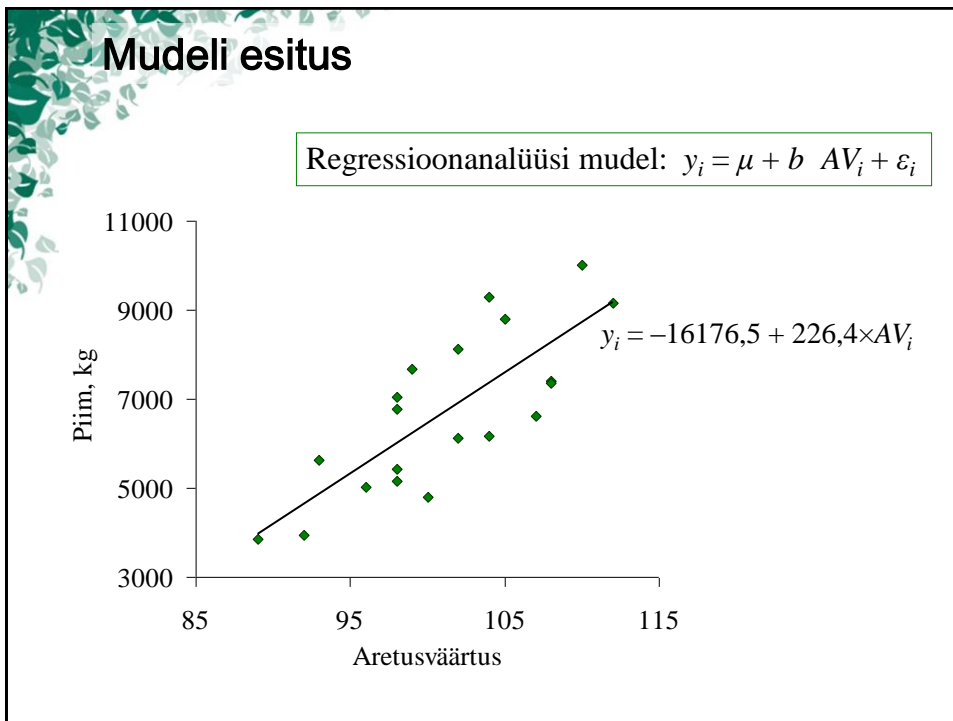
Näiteandmestik

Lehm	Tõug	Farm	Aretusväärnus	Piim. kg
1	EHF	F1	105	8804.256
2	EHF	F3	112	9152.284
3	EHF	F4	98	7055.046
4	EHF	F2	89	3856.88
5	EHF	F1	98	6768.067
6	EHF	F4	99	7676.258
7	EHF	F4	104	9282.086
8	EHF	F1	102	8126.694
9	EHF	F1	110	10017.95
10	EHF	F1	93	5622.356
11	EPK	F2	98	5431.155
12	EPK	F2	108	7406.513
13	EPK	F4	98	5152.659
14	EPK	F3	100	4797.637
15	EPK	F3	96	5011.426
16	EPK	F4	108	7369.143
17	EPK	F2	107	6626.611
18	EPK	F2	104	6170.835
19	EPK	F2	92	3948.281
20	EPK	F3	102	6113.998

↙ ↘
Diskreetsed
faktorid

↑
Pidevad
faktorid

↑
Sõltuv
muutuja



Mudeli esitus

$$y_{ijk} = \mu + T_i + F_j + b \times AV_{ijk} + \varepsilon_{ijk}$$

$$y_{211} = \mu + 0 \times T_1 + 1 \times T_2 + 1 \times F_1 + 0 \times F_2 + 0 \times F_3 + 0 \times F_4 + b \times AV_{211} + \varepsilon_{211}$$

$y = X \times \beta + \varepsilon$

$$\begin{pmatrix} 8804,3 \\ 9152,3 \\ 7055,0 \\ 3856,9 \\ 6768,1 \\ 7676,3 \\ 9282,1 \\ 8126,7 \\ 10018,0 \\ 5622,4 \\ 5431,2 \\ 7406,5 \\ 5152,7 \\ 4797,6 \\ 5011,4 \\ 7369,1 \\ 6626,6 \\ 6170,8 \\ 3948,3 \\ 6114,0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 105 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 112 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 98 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 89 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 98 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 99 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 104 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 102 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 110 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 93 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 98 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 108 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 98 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 100 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 96 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 108 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 107 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 104 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 92 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 102 \end{pmatrix} \times \begin{pmatrix} \mu \\ Toug_1 \\ Toug_2 \\ Farm_1 \\ Farm_2 \\ Farm_3 \\ Farm_4 \\ b \end{pmatrix} + \varepsilon$$

Hinnatavad efektid, reparametriseerimine

$$y = X\beta + \varepsilon \Rightarrow \hat{\beta} = (X^T X)^{-1} X^T y$$

Probleem on, et β ei ole üheselt hinnatav.

Vaatame näiteks ANOVA-mudelit: $y_{ij} = \mu + B_i + \varepsilon_{ij}$

Keskmine piimatoodang (Tõug = "EHF") = $\mu_1 = \mu + B_1 = 7636,2$ kg

Keskmine piimatoodang (Tõug = "EPK") = $\mu_2 = \mu + B_2 = 5802,8$ kg

Meil on 2 võrrandit ja 3 tundmatut parameetrit.

Lahendus? Reparametrisatsioon = lisakitsendused

- Klassikaline reparametrisatsioon: $B_1 + B_2 = 0$
($\mu = 6719,5$; $B_1 = 916,7$; $B_2 = -916,7$)
- SAS-i reparametrisatsioon: $B_2 = 0$ ($B_1 = 1833,4$; $\mu = 5802,8$)
- R-i reparametrisatsioon: $B_1 = 0$ ($B_2 = -1833,4$; $\mu = 7636,2$)

Hinnatavad efektid, reparametriseerimine

SAS

```
proc glm data=dk0007;
class breed farm;
model milk = breed farm by / solution;
run;
```

The SAS System 14:16 Wednesday, February 13, 2008 8

The GLM Procedure

Dependent Variable: milk milk

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	-16637.09733 B	1893.708817	-8.79	<.0001
breed EHF	1527.26447 B	291.258313	5.24	0.0001
breed EPK	0.00000 B	.	.	.
farm F1	-95.49905 B	327.411139	-0.29	0.7748
farm F2	-678.20733 B	321.650812	-2.11	0.0535
farm F3	-753.46783 B	340.016142	-2.22	0.0438
farm F4	0.00000 B	.	.	.
bv	227.09833	18.304106	12.41	<.0001

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

Hinnatavad efektid, reparametriseerimine

```

RGui

> cow.model1 <- lm(milk ~ breed + farm + bv, data=dk0007_glm)
> summary(cow.model1)

Call:
lm(formula = milk ~ breed + farm + bv, data = dk0007_glm)

Residuals:
    Min       1Q   Median       3Q      Max
-566.84 -373.07   36.75  312.64  773.69

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -15205.3     1872.2   -8.122 1.15e-06 ***
breed[T.EPK] -1527.3       291.3   -5.244 0.000124 ***
farm[T.F2]   -582.7       384.2   -1.517 0.151575
farm[T.F3]   -658.0       390.5   -1.685 0.114160
farm[T.F4]    95.5       327.4    0.292 0.774809
bv           227.1        18.3   12.407 6.09e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 483.5 on 14 degrees of freedom
Multiple R-Squared:  0.9455,    Adjusted R-squared:  0.9261
F-statistic: 48.6 on 5 and 14 DF,  p-value: 2.354e-08
    
```

Hinnatavad funktsioonid, kontrastid

Kontrast on mudeli parameetrite hinnatav lineaarkombinatsioon.

Kontrastide esitamiseks sobib kasutada maatrikskorutist kujul $\mathbf{I}\beta$.

Näiteks kontrast, hindamaks tõugudevahelist erinevust, on esitatav kujul

$$\mathbf{1} \times \beta = \begin{pmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \times \begin{pmatrix} \mu \\ \tilde{Tõug}_1 \\ \tilde{Tõug}_2 \\ Farm_1 \\ Farm_2 \\ Farm_3 \\ Farm_4 \\ b \end{pmatrix} = 1 \times \tilde{Tõug}_1 - 1 \times \tilde{Tõug}_2$$

Milline efekt (erinevus) on hinnatav reavektori \mathbf{I} abil:

$$\mathbf{I} = (0 \ 0 \ 0 \ 0,5 \ 0,5 \ -0,5 \ -0,5 \ 0)?$$

Vähimruutkeskmised [*least square means*]

Vähimruutkeskmise [LSM] kujutab enesest mingi faktori mingile tasemele vastavate väärtuste keskmist, mis on hinnatud mudelist sobivalt defineeritud kontrasti kujul.

Näiteks 2. farmi lehmade piimatoodangu vähimruutkeskmise hinnatakse kujul

$$\begin{aligned}
 \text{LSM}(\text{Farm}_2) &= 1 \begin{vmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 1 & 0 & 0 \end{vmatrix} \bar{a}v \times \begin{pmatrix} \mu \\ T\ddot{o}ug_1 \\ T\ddot{o}ug_2 \\ \text{Farm}_1 \\ \text{Farm}_2 \\ \text{Farm}_3 \\ \text{Farm}_4 \\ b \end{pmatrix} \\
 &= 1 \times \mu + \frac{T\ddot{o}ug_1 + T\ddot{o}ug_2}{2} + 1 \times \text{Farm}_2 + b \times \bar{a}v
 \end{aligned}$$

Vähimruutkeskmised [*least square means*]

SAS

```

proc glm data=dk0007;
class breed farm;
model milk = breed farm bv / solution;
lsmeans breed / stderr pdiff;
run;
    
```

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	-16637.09733 B	1893.708817	-8.79	<.0001
breed EHF	1527.26447 B	291.258313	5.24	0.0001
breed EPK	0.00000 B	.	.	.
farm F1	-95.49905 B	327.411139	-0.29	0.7748
farm F2	-678.20733 B	321.650812	-2.11	0.0535
farm F3	-753.46783 B	340.016142	-2.22	0.0438
farm F4	0.00000 B	.	.	.
bv	227.09839	18.304106	12.41	<.0001

The GLM Procedure
Least Squares Means

breed	milk LSMEAN	Standard Error	H0:LSMEAN=0 Pr > t	H0:LSMean1=LSMean2 Pr > t
EHF	7479.37593	180.89567	<.0001	0.0001
EPK	5952.11146	183.24601	<.0001	

$$\begin{aligned}
 \text{LSM}_{\text{EHF}} &= \text{Intercept} + 1 \times \text{EHF} + 0 \times \text{EPK} + \frac{(F1 + F2 + F3 + F4)}{4} + \text{coef} \times \bar{BV} \\
 &= -16637,1 + 1527,3 + \frac{(-95,5 - 678,2 - 753,5 + 0)}{4} + 227,1 \times 101,15 \approx 7479,5
 \end{aligned}$$

keskmine
aretusväärtus

I ja III tüüpi ruutude summad

SAS

```
proc glm data=dk0007;
class breed farm;
model milk = breed farm bv / solution;
run;
```

The SAS System 14:16 Wednesday, February 13, 2008 8

The GLM Procedure

Dependent Variable: milk milk

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	56822518.79	11364503.76	48.60	<.0001
Error	14	3273463.43	233818.82		
Corrected Total	19	60095982.22			

	R-Square	Coeff Var	Root MSE	milk Mean
	0.945529	7.196185	483.5482	6719.507

Source	DF	Type I SS	Mean Square	F Value	Pr > F
breed	1	16806079.82	16806079.82	71.88	<.0001
farm	3	4024063.57	1341354.52	5.74	0.0089
bv	1	35992375.40	35992375.40	153.93	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
breed	1	6429116.53	6429116.53	27.50	0.0001
farm	3	1510502.69	503500.90	2.15	0.1392
bv	1	35992375.40	35992375.40	153.93	<.0001

I ja III tüüpi ruutude summad

RGui

```
> anova(cov.model1)
Analysis of Variance Table

Response: milk
      Df Sum Sq Mean Sq F value Pr(>F)
breed   1 16806080 16806080  71.8765 6.925e-07 ***
farm    3  4024064  1341355   5.7367 0.008943 **
bv      1 35992375 35992375 153.9328 6.089e-09 ***
Residuals 14  3273463  233819
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> drop1(cov.model1, test="F")
Single term deletions

Model:
milk ~ breed + farm + bv
      Df Sum of Sq    RSS   AIC F value    Pr(>F)
<none>          3273463    252
breed  1   6429117  9702580    272  27.4961 0.0001243 ***
farm   3   1510503  4783966    254   2.1534 0.1392264
bv     1   35992375 39265839    300 153.9328 6.089e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Korduvad mõõtmised

ID	Lakt.	Piim, kg
3396	1	4119
3396	2	5857
3396	3	6260
3990	1	3106
3990	2	3934
3990	3	5171
4390	1	2473
4390	2	3301
4390	3	2958
...

Vaikimisi:

$$\text{var}(\boldsymbol{\varepsilon}) = \sigma_{\varepsilon}^2 \mathbf{I} = \begin{pmatrix} \sigma_{\varepsilon}^2 & 0 & 0 & \dots \\ 0 & \sigma_{\varepsilon}^2 & 0 & \dots \\ 0 & 0 & \sigma_{\varepsilon}^2 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Kompaund-sümmeetriline kovariatsioonistruktuur:

$$\text{var}(\boldsymbol{\varepsilon}) = \begin{pmatrix} \sigma_{\varepsilon}^2 + \sigma^2 & \sigma^2 & \sigma^2 & 0 & 0 & \dots \\ \sigma^2 & \sigma_{\varepsilon}^2 + \sigma^2 & \sigma^2 & 0 & 0 & \dots \\ \sigma^2 & \sigma^2 & \sigma_{\varepsilon}^2 + \sigma^2 & 0 & 0 & \dots \\ 0 & 0 & 0 & \sigma_{\varepsilon}^2 + \sigma^2 & \sigma^2 & \dots \\ 0 & 0 & 0 & \sigma^2 & \sigma_{\varepsilon}^2 + \sigma^2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Esimest järku autoregressiivne kovariatsioonistruktuur:

$$\text{var}(\boldsymbol{\varepsilon}) = \sigma_{\varepsilon}^2 \times \begin{pmatrix} 1 & \rho & \rho^2 & 0 & 0 & 0 & \dots \\ \rho & 1 & \rho & 0 & 0 & 0 & \dots \\ \rho^2 & \rho & 1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 1 & \rho & \rho^2 & \dots \\ 0 & 0 & 0 & \rho & 1 & \rho & \dots \\ 0 & 0 & 0 & \rho^2 & \rho & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, |\rho| \leq 1$$

Juhuslikud faktorid, segamudel

Näide 1

Testitakse kolme erineva ravusti mõju (kolmel sarnasel patsientide grupil)

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

α_i – ravimi i mõju, $i=1,2,3$, kusjuures kõik kolm testitavat ravimit on eelnevalt võrdlemiseks välja valitud

↓

vaatluse all on **iga ravimi** efekt

fikseeritud efektid,
fikseeritud mudel

Näide 2

Uut tüüpi süstitavat insuliini testitakse kolmes kliinikus

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

α_i – kliiniku i mõju, $i=1,2,3$, kusjuures uuringus osalenud kliinikud on valitud juhuslikult

↓

huvi pakub kliiniku **osa** ravimi mõju **koguvareeruvusest**

juhuslikud efektid,
juhuslik mudel

Juhuslikud faktorid, segamudel

- Kui palju on faktoril tasemeid?
Kui vähe, siis on ilmselt tegu fikseeritud faktoriga.
Kui palju, siis võib tegu olla juhusliku faktoriga.
- **Kas faktori tasemete arv populatsioonis on potentsiaalselt lõpmatu?**
Kui jah, on tegu juhusliku faktoriga.
- Kas eksperimendi kordamisel on (võivad olla) uuritavad samad tasemed?
Kui jah, on faktor fikseeritud.
- Huvi pakub kõigi (ka andmetes esindamata) tasemete keskmine mõju, ehk see, kui suur osa uuritava tunnuse koguvarieeruvusest on kirjeldatud antud faktori poolt?
Kui jah, on ilmselt mõttekas käsitleda faktorit juhuslikuna.
- Kas faktori tasemed on valitud sihipäraselt (mitte juhuslikult)?
Kui jah, siis ilmselt tuleb faktorit käsitleda fikseerituna.

Juhuslikud faktorid, segamudel

- **Eelised**
 - Segamudelid võimaldavad tunduvalt üldistada tehtavaid statistilisi järeldusi.
 - Segamudelid võimaldavad paindlikult modelleerida vaatluste kovariatsioonistruktuuri.
 - Segamudelite alusel tehtavad järeldused on vähem tundlikud andmete ebatäielikkuse ja/või mittetasakaalulisuse suhtes.
- **Puudused**
 - Tehtavate tõenäosusjaotuslike eelduste ja kasutatavate lähenduste hulk on suurem, mis võib viia nihkega hinnanguteni.
 - Mudelite suurem komplitseeritus võib muuta keeruliseks andmete töötlemise ja tulemuste esitamise.

Fikseeritud ja juhusliku mudeli karakteristikud ühefaktorilisel DA-I matemaatika mõistes

Karakteristik	Fikseeritud mudel	Juhuslik mudel
Mudel	$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$	$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$
$E(y_{ij})$	$E(y_{ij}) = \mu + \alpha_i$	$E(y_{ij} \alpha_i) = \mu + \alpha_i$ $E(y_{ij}) = \mu$
α_i	fiks. tundmatu konstant $\hat{\alpha}_i = \bar{y}_i - \mu$	$\alpha_i \sim \text{i.i.d.}(0, \sigma_\alpha^2)$ $\tilde{\alpha}_i = \frac{n_i \sigma_\alpha^2}{\sigma_\varepsilon^2 + n_i \sigma_\alpha^2} (\bar{y}_i - \mu)$
ε_{ij}	$\varepsilon_i \sim \text{i.i.d.}(0, \sigma_\varepsilon^2)$	$\varepsilon_i \sim \text{i.i.d.}(0, \sigma_\varepsilon^2)$
$\text{var}(y_{ij})$	$\text{var}(y_{ij}) = \sigma_\varepsilon^2$	$\text{var}(y_{ij}) = \sigma_\alpha^2 + \sigma_\varepsilon^2$
$\text{cov}(y_{ij}, y_{i'j'})$	$\text{cov}(y_{ij}, y_{i'j'}) = \begin{cases} \sigma_\varepsilon^2, & i = i', j = j' \\ 0, & \text{mujal} \end{cases}$	$\text{cov}(y_{ij}, y_{i'j'}) = \begin{cases} \sigma_\alpha^2 + \sigma_\varepsilon^2, & i = i', j = j' \\ \sigma_\alpha^2, & i = i', j \neq j' \\ 0, & \text{mujal} \end{cases}$

Juhuslikud faktorid, segamudel

2 pulli, kummaltki 3 järglast.

Mudel, hindamaks isa mõju järglaste: $y_{ij} = \mu + s_i + \varepsilon_{ij}$

Isa, kui fikseeritud faktor

$$\text{var}(\mathbf{y}) = \begin{matrix} & y_{11} & y_{12} & y_{13} & y_{21} & y_{22} & y_{23} \\ \begin{matrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{matrix} & \begin{pmatrix} \sigma_\varepsilon^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_\varepsilon^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_\varepsilon^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_\varepsilon^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_\varepsilon^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_\varepsilon^2 \end{pmatrix} \end{matrix}$$

Isa, kui juhuslik faktor

$$\text{var}(\mathbf{y}) = \begin{matrix} & y_{11} & y_{12} & y_{13} & y_{21} & y_{22} & y_{23} \\ \begin{matrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{matrix} & \begin{pmatrix} \sigma_s^2 + \sigma_\varepsilon^2 & \sigma_s^2 & \sigma_s^2 & 0 & 0 & 0 \\ \sigma_s^2 & \sigma_s^2 + \sigma_\varepsilon^2 & \sigma_s^2 & 0 & 0 & 0 \\ \sigma_s^2 & \sigma_s^2 & \sigma_s^2 + \sigma_\varepsilon^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_s^2 + \sigma_\varepsilon^2 & \sigma_s^2 & \sigma_s^2 \\ 0 & 0 & 0 & \sigma_s^2 & \sigma_s^2 + \sigma_\varepsilon^2 & \sigma_s^2 \\ 0 & 0 & 0 & \sigma_s^2 & \sigma_s^2 & \sigma_s^2 + \sigma_\varepsilon^2 \end{pmatrix} \end{matrix}$$

Segamudeli üldkuju

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

$$E \begin{pmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} \text{ ja } \text{var} \begin{pmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R} & \mathbf{G}\mathbf{Z}^T & \mathbf{R} \\ \mathbf{Z}\mathbf{G} & \mathbf{G} & \mathbf{0} \\ \mathbf{R} & \mathbf{0} & \mathbf{R} \end{pmatrix}, \text{var}(\mathbf{u}) = \mathbf{G}, \text{var}(\mathbf{e}) = \mathbf{R}$$

Fikseeritud efektide $\boldsymbol{\beta}$ üldistatud vähimruutude hinnangud [*GLS, generalized least squares*]: $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$

Juhuslike faktorite realiseerunud väärtuste \mathbf{u} parim lineaarne nihketa prognoos [*BLUP, best linear unbiased prediction*]: $\hat{\mathbf{u}} = \mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$

(Hendersoni) segamudeli võrrand [*mixed model equation*]:

$$\begin{pmatrix} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}^T\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}$$

Kui $\mathbf{R} = \mathbf{I}_n\sigma_e^2$ ja $\mathbf{G} = \mathbf{I}_a\sigma_u^2$, siis $\begin{pmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \frac{\sigma_e^2}{\sigma_u^2}\mathbf{I} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{pmatrix}$

Mudelite võrdlemine

Hierarhiliste mudelite võrdlemiseks

- ⌘ lihtsamal juhul dispersioonanalüüs (*F*-statistik \rightarrow *p*-väärtus)
- ⌘ keerulisemal juhul tõepärasuhte test [*likelihood ratio test*]

$$\Lambda(\mathbf{z}) = -2\ln \left[\frac{\hat{L}_k(\mathbf{z})}{\hat{L}(\mathbf{z})} \right] = -2 \ln \left[\hat{L}_k(\mathbf{z}) \right] - \ln \left[\hat{L}(\mathbf{z}) \right] \underset{H_0}{\sim} \chi^2(r)$$

Üldisema võrdlemise tarvis

- ⌘ AIC (Akaike informatsiooni kriteerium)
- ⌘ BIC (Bayesi informatsiooni kriteerium)

NB! Testida ei saa, mida väiksem väärtus, seda parem.