

Anvendt medisinsk statistikk, vår 2010

Repeterte målinger, del II

Eirik Skogvoll

Overlege, Klinikk for anestesi og akuttmedisin
1. amanuensis, Enhet for anvendt klinisk forskning
(med bidrag fra Harald Johnsen)

Dagens tema

- Varianskomponenter
- Sammensatt lineær modell ("linear mixed effects model") med faste ("fixed") og tilfeldige ("random") effekter
- Faktoriell design med repeterte målinger

Klassisk ANOVA/ lineær modell

- Utgangspunkt: grupperte observasjoner av kontinuerlig, tilfeldig variabel
- Problemstilling: er gruppenes forventningsverdi lik?
 H_0 testes med F-test
- Hvilke grupper skiller seg ut?
Post-hoc tester avhengig av problemstilling (Bonferroni, Tukey, Sidak m.fl.)

Klassisk ANOVA/ lineær modell

- Målet er å generalisere fra en fast effekt/ faktor til populasjonen.
- Nivået til gruppene har interesse i seg selv.
Eks. Rosner Ex. 12.1 (s. 557). Lungefunksjon (FEF) blant røykere:

FEF data for smoking and nonsmoking males				
Group number, i	Group name	Mean FEF (L/s)	sd FEF (L/s)	n_i
1	NS	3.78	0.79	200
2	PS	3.30	0.77	200
3	NI	3.32	0.86	50
4	LS	3.23	0.78	200
5	MS	2.73	0.81	200
6	HS	2.59	0.82	200

Source: Reprinted by permission of the *New England Journal of Medicine*, 302(13), 720–723, 1980.

Varianskomponenter

- Utgangspunkt: gjentatte observasjoner av kontinuerlig tilfeldig variabel
Nøkkelbegreper:
- Repeterbarhet (eng. "repeatability")
 - Samme analyse under så like forhold som overhodet mulig. Uttrykker den grunnleggende, usystematiske feil.
- Reproduserbarhet (eng. "reproducibility")
 - Samme analyse under skiftende forhold, for eksempel dag/natt, type utstyr, operatør. Uttrykker den "realistiske" feil fra mange kilder

Varianskomponenter

- Eks. 1
Reproduserbarhet av plasma østradiol
(Rosner table 12.20, s. 607)

Subj	Rep 1	Rep 2	Mean
1	25,50	30,40	27,95
2	11,10	15,00	13,05
3	8,00	8,10	8,05
4	20,70	16,90	18,80
5	5,80	8,40	7,10

- Observerer først at differensen synes å øke med middelveiden. Analyserer derfor på log skala isteden.
- Plott ...

Varianskomponenter

- Problemstilling:
er variasjonen mellom pasientene større enn variasjonen av gjentatte analyser av samme pasient?
- Ingen egentlig interesse av å kjenne nivået til den enkelte pasient!
 - vanlig ANOVA uinteressant
- Variansen til den enkelte måling oppfattes som en sum av to uavhengige komponenter:
 - Between subjects
 - Within subjects (residual)

Varianskomponenter

Modell:

$$y_{ij} = \mu + b_i + e_{ij} \quad b_i \sim N(0, \sigma_B^2), \quad e_{ij} \sim N(0, \sigma^2), \text{ uavh av } b_i.$$

Altså er $y_{ij} \sim N(\mu, \sigma_B^2 + \sigma^2)$

$i = 1, 2, \dots, k$ (gruppe, klasse el. individ)

$j = 1, 2, \dots, n_i$ (observasjoner innen hver gruppe el. klasse)

NB! α_i i Rosner Eq. 12.29 kalles b_i i denne presentasjonen.

Variasjonskoeffisient – CV Eks. I

$$CV \text{ (målt i \%)} = 100 \cdot \frac{\sigma}{\mu}$$

$$\text{Estimeres fra ANOVA som } 100 \cdot \frac{\sqrt{MSE}}{\bar{y}}$$

$$\text{eller } 100 \cdot \frac{SD(y)}{\bar{y}}$$

Ekspl. Rosner 6.ed, tabell 12.20 (obs. log-skala)

$$\hat{\sigma}^2 = MSE = 0.030, \hat{\sigma}_B^2 = 0.317, \bar{y} = 2.57, \text{ gir } CV = 100 \cdot \frac{\sqrt{0.030}}{2.57} = 6.74\%$$

Men ift. originalskala benytter man MSE fra log-analysen direkte:

$$CV = 100 \cdot \sqrt{MSE} = 100 \cdot \sqrt{0.030} = 17.3\%$$

Intra-klasse korrelasjon – ICC Eks. I

- ICC $\in [0,1]$
- Et mål på reproduserbarhet:
I hvor stor grad er klassene/
gruppene (her: pasientene)
"likere seg selv enn andre"?

$$ICC = \frac{\sigma_B^2}{\sigma_B^2 + \sigma^2}$$

Eks. Rosner tabell 12.20 (log østradiol)

$$ICC = \frac{\hat{\sigma}_B^2}{\hat{\sigma}_B^2 + \hat{\sigma}^2} = \frac{0.317}{0.317 + 0.03} = 0.91$$

Lineær, sammensatt modell

- Eng: "linear mixed effects model"
- Både faste ("fixed") og tilfeldige ("random") faktorer/ effekter
- Utgangspunkt: grupperte observasjoner av kontinuerlig tilfeldig variabel
- Gruppene er "likere seg selv enn andre": avhengighet (korrelasjon) mellom observasjoner fra samme gruppe
- Gruppene er trukket tilfeldig fra populasjonen av tilsvarende grupper
- Målet er å generalisere til denne populasjonen
- Eksempler:
 - Gjentatte observasjoner på samme individ
 - Observasjoner fra forskjellige skoler, sykehus, byer, land...

Lineær, sammensatt modell

- Problemstilling: bidrar gruppene til økt varians? Hvor mye?
- Ingen egentlig interesse av å kjenne nivået til den enkelte gruppe!
 - vanlig ANOVA uinteressant
- Målet er å estimere bidrag til variansen fra gruppestrukturen.
- Eksplisitt modellere og kontrollere for avhengighet (korrelasjon) i datamaterialet.

Lineær, sammensatt modell

$$y_{ij} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + b_i + e_{ij} \quad b_i \sim N(0, \sigma_b^2), \quad e_{ij} \sim N(0, \sigma^2), \text{ uavh av } b_i.$$

Altså er $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots$

$i = 1, 2, \dots, k$ (gruppe, klasse el. individ)

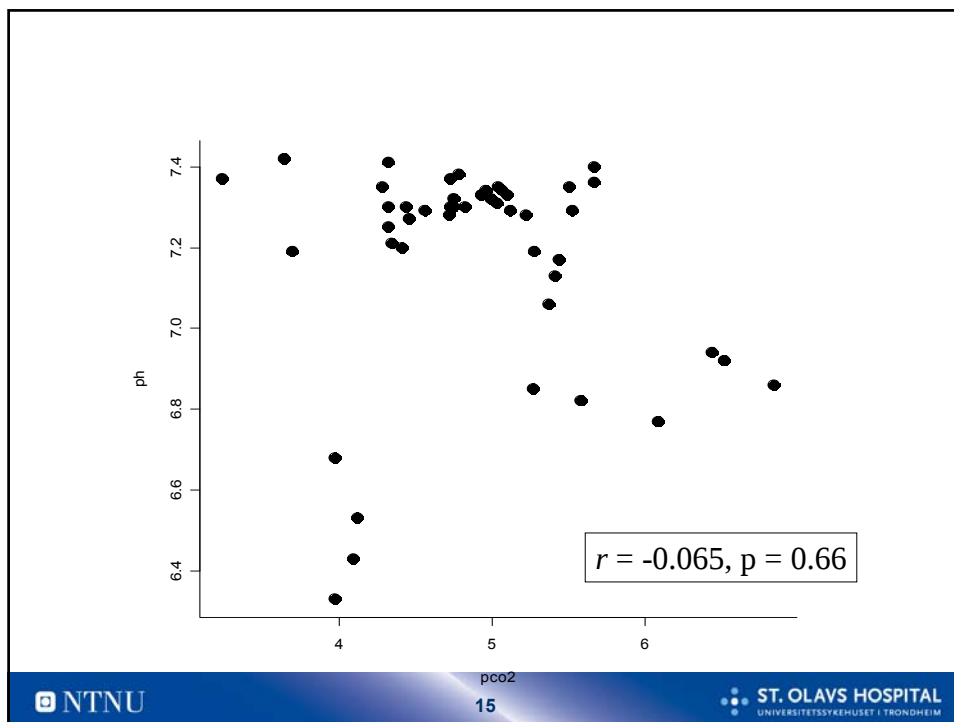
$j = 1, 2, \dots, n_i$ (observasjoner innen hver gruppe el. klasse)

Eksempel II: intramural pH vs. $P_a\text{CO}_2$

obs	id	ph	pcO2
1	1	6.68	3.97
2	1	6.53	4.12
3	1	6.43	4.09
4	1	6.33	3.97
5	2	6.85	5.27
6	2	7.06	5.37
7	2	7.13	5.41
8	2	7.17	5.44
9	3	7.40	5.67
10	3	7.42	3.64
11	3	7.41	4.32
12	3	7.37	4.73
13	3	7.34	4.96
14	3	7.35	5.04
15	3	7.28	5.22
16	3	7.30	4.82
17	3	7.34	5.07

obs	id	ph	pcO2
18	4	7.36	5.67
19	4	7.33	5.10
20	4	7.29	5.53
21	4	7.30	4.75
22	4	7.35	5.51
23	5	7.35	4.28
24	5	7.30	4.44
25	5	7.30	4.32
26	5	7.37	3.23
27	5	7.27	4.46
28	5	7.28	4.72
29	5	7.32	4.75
30	5	7.32	4.99
31	6	7.38	4.78
32	6	7.30	4.73
33	6	7.29	5.12
34	6	7.33	4.93
35	6	7.31	5.03
36	6	7.33	4.93

obs	id	ph	pcO2
37	7	6.86	6.85
38	7	6.94	6.44
39	7	6.92	6.52
40	8	7.19	5.28
41	8	7.29	4.56
42	8	7.21	4.34
43	8	7.25	4.32
44	8	7.20	4.41
45	8	7.19	3.69
46	8	6.77	6.09
47	8	6.82	5.58



Eksempel II: intramural pH vs. P_aCO_2

$$y_{ij} = \beta_0 + \beta_1 x_1 + b_i + e_{ij} \quad b_i \sim N(0, \sigma_B^2), \quad e_{ij} \sim N(0, \sigma^2), \text{ uavh av } b_i.$$

Altså er $E(Y) = \beta_0 + \beta_1 x_1$

$i = 1, 2, \dots, 15$ (individ, id)

$j = 1, 2, \dots, n_i$ (observasjoner innen hver gruppe)

$$pH_{ij} = \beta_0 + \beta_1 \cdot PCO_2 + b_i + e_{ij}$$

Altså er $E(pH) = \beta_0 + \beta_1 PCO_2$

$i = 1, 2, \dots, 15$ (individ, id)

$j = 1, 2, \dots, n_i$ (observasjoner innen hver gruppe)

Eksempel II: intramural pH vs. P_aCO_2 SPSS: Analyze → Mixed models → Linear

Estimates of Fixed Effects^a

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	7,631356	,184785	31,901	41,299	,000	7,254915	8,007797
pco2	-,103271	,029354	40,565	-3,518	,001	-,162572	-,043971

a. Dependent Variable: ph.

Estimates of Covariance Parameters^a

Parameter	Estimate	Std. Error
Residual	,008785	,002016
Intercept [subject = id] Variance	,097099	,053187

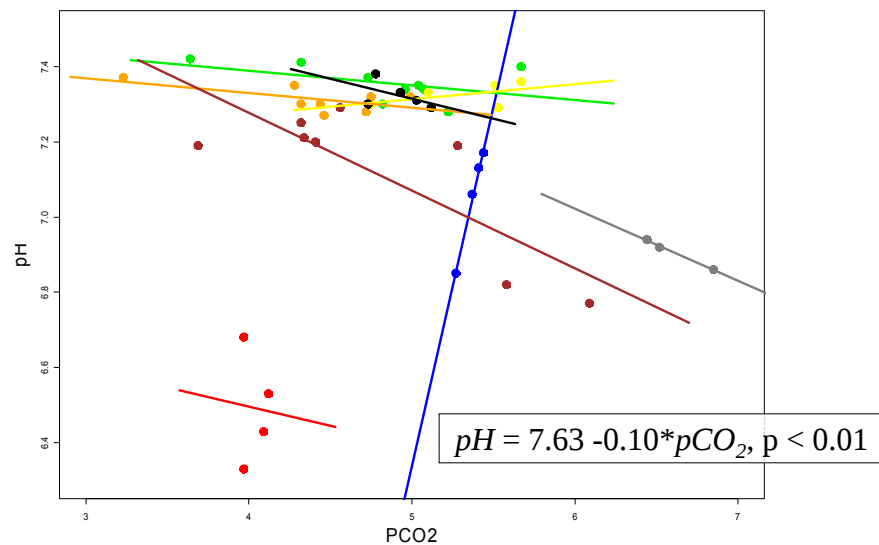
a. Dependent Variable: ph.

$$\hat{\beta}_0 = 7.63 \quad \hat{\beta}_1 = -0.10$$

$$\hat{\sigma}^2 = 0.009 \quad \hat{\sigma}_B^2 = 0.097$$

$$ICC = \frac{\hat{\sigma}_B^2}{\hat{\sigma}_B^2 + \hat{\sigma}^2} = \frac{0.097}{0.097 + 0.009} = 0.92$$

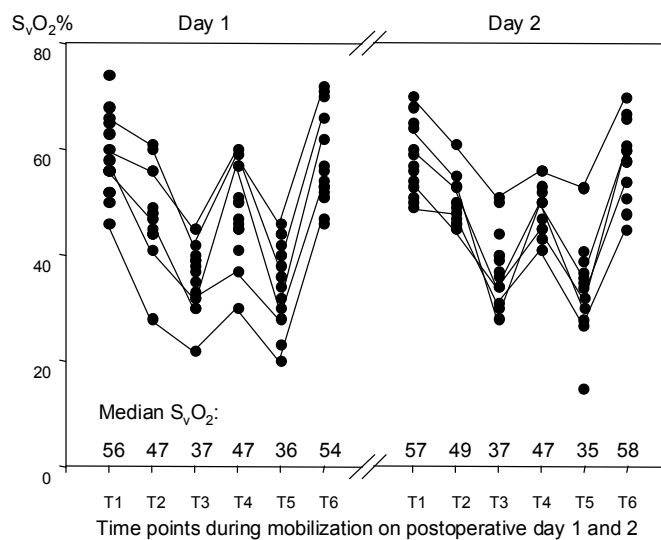
Estimert stigningskoeffisient (β) fra enkel lineær regresjon for hvert individ



Eksempel III: mobilisering etter hjertekirurgi

- Observasjonsstudie ved St. Elisabeth
- Blandet venøs oksygenmetning målt på samme pasient ved tidspunktene T1-T6:
 - Hvilende, i seng T1
 - Sittende i stol T2
 - Stå/gå ved senga T3
 - Sittende i stol T4
 - Stå/gå ved senga T5
 - Hvilende, i seng T6

Gjentatte observasjoner på samme individ:



Eksempel III: mobilisering etter hjertekirurgi.

$$y_{ij} = \beta_j + b_i + e_{ij} \quad b_i \sim N(0, \sigma_B^2), \quad e_{ij} \sim N(0, \sigma^2), \text{ uavh av } b_i.$$

Altså er $E(Y) = \beta_j$

$i = 1, 2, \dots, 15$ (individ, id)

$j = 1, 2, \dots, 6$ (fast ["fixed"] faktor; tidspunkt for mobilisering,
kun en observasjon for hver situasjon)

Eksempel III: mobilisering etter hjertekirurgi. SPSS: Analyze → Mixed models → Linear

Estimates of Fixed Effects^a

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
[Index1=1]	58,533333	2,022048	24,517	28,948	,000	54,364685	62,701982
[Index1=2]	47,866667	2,022048	24,517	23,672	,000	43,698018	52,035315
[Index1=3]	35,933333	2,022048	24,517	17,771	,000	31,764685	40,101982
[Index1=4]	48,000000	2,022048	24,517	23,738	,000	43,831352	52,168648
[Index1=5]	35,533333	2,022048	24,517	17,573	,000	31,364685	39,701982
[Index1=6]	57,400000	2,022048	24,517	28,387	,000	53,231352	61,568648

^a. Dependent Variable: SVO2 Day 1 premob, oxymetri.

Estimates of Covariance Parameters^a

Parameter	Estimate	Std. Error
Residual	18,608254	3,145369
Intercept [subject = id] Variance	42,721905	17,327504

^a. Dependent Variable: SVO2 Day 1 premob, oxymetri.

$$\hat{\beta}_{1..6} = 58.5, 47.8, \dots, 57.4$$

$$\hat{\sigma}^2 = 18.6 \quad \hat{\sigma}_B^2 = 42.7$$

$$ICC = \frac{\hat{\sigma}_B^2}{\hat{\sigma}_B^2 + \hat{\sigma}^2} = \frac{42.7}{42.7 + 18.6} = 0.70$$