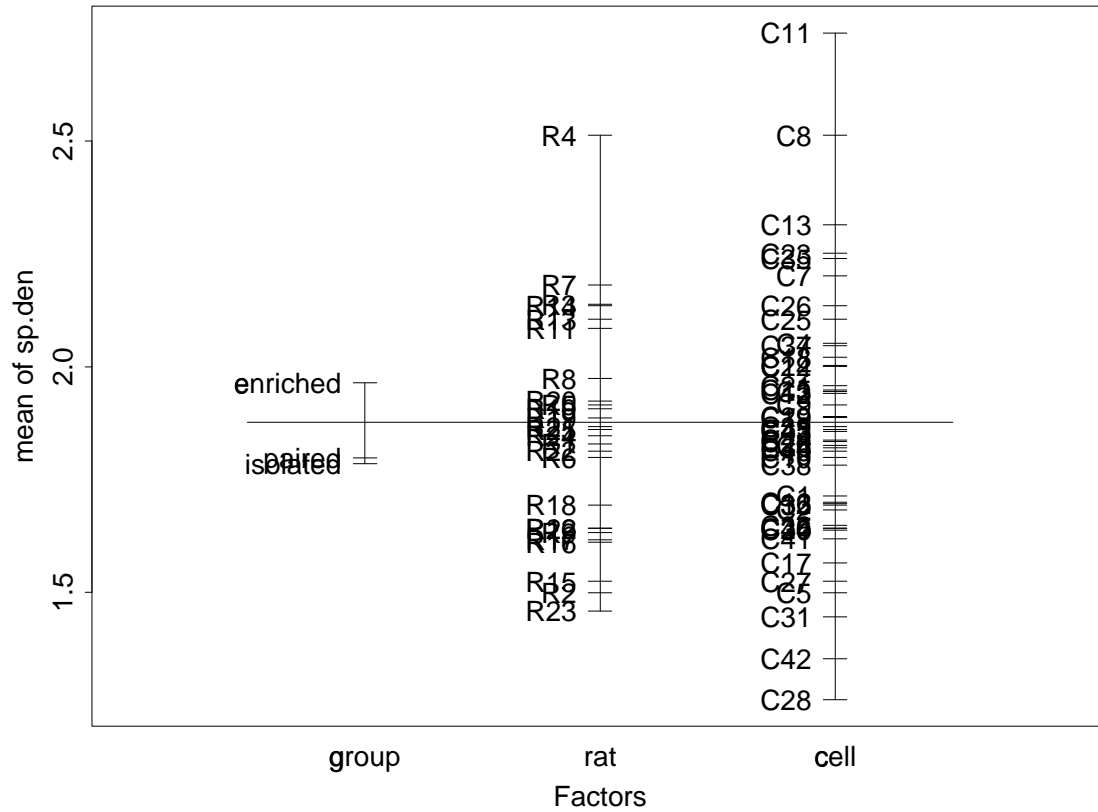


VARIANSANALYSE FOR PREKLINIKERE



Gjennomsnittet for rotte 4 (R4) er omtrent 2.5; for R8 omtrent 2.

Thore Egeland, Norsk Regnesentral
6 mars 1995. ¹

¹Tilgjengelig via Mosaic: <http://www.nr.no/home/SAND/thore/>

INNHOOLD FREM TIL LUNSJ

- Noen motiverende datamaterialer.
Hva vil vi finne ut?
- Sentrale begreper for å kunne *identifisere og spesifisere* modell:
 - Balansert forsøksplan.
 - Fast effekt vs. tilfeldig effekt.
 - Varianskomponent.
 - Nestet (hierarkisk) modell.
- Analyse av dataene for hånd.
Box, Hunter & Hunter s. 571–579.
- Presentasjon av resultater.
- RAT.DAT. (data fra BMDP).
- Presis modellspesifikasjon.
- Antagelser. Tilpasning til antagelser.
Kontroll av modellforutsetninger.
- Eksempel: spinatetthet.
 - Biologi ved Mari Trommald (ca 15 min).
 - Dataanalyse.

HVA VIL VI FINNE UT?

	group	rat	cell	sp.den
1	enriched	R1	C1	1.48
2	enriched	R1	C1	1.63
3	enriched	R1	C1	1.68
4	enriched	R1	C1	1.80
5	enriched	R1	C1	1.96
6	enriched	R1	C2	1.48
.
.
.
279	isolated	R26	C46	2.14

- Er det en behandlingseffekt?
- Hvilke komponenter bidrar mest til variasjon?

Problem: Observasjoner fra samme rotte er potensielt *avhengige*, observasjoner fra samme celle er *mer* avhengige.

t-tester, wilcoxon tester o.l. kan ikke brukes pga. avhengighet.

Avhengigheten *må* modelleres.

BALANSERT

Balansert forsøksplan:

Hver kategori har like mange observasjoner.

Behandling A og B.

Balansert:

A	B
2.5	2.9
2.2	3.1

Ubalansert:

A	B
2.5	2.9
2.2	3.1
2.4	

BETYDNINGEN AV BALANSERTE PLANER

- Balanserte planer er lettere å analysere og fortolke enn ubalanserte planer.
- I *våre* dager kan imidlertid ubalanserte planer håndteres.

Vi tenker ut fra ubalanserte planer i fortsettelsen.

FASTE OG TILFELDIGE EFFEKTER

	group	rat	cell	sp.den
1	enriched	R1	C1	1.48
2	enriched	R1	C1	1.63
3	enriched	R1	C1	1.68
4	enriched	R1	C1	1.80
5	enriched	R1	C1	1.96
6	enriched	R1	C2	1.48
.
.
.
279	isolated	R26	C46	2.14

- group (behandling): Typisk fast effekt.
Valgt. Noen få nivåer.
Nivåene av interesse som sådan.
- rat, cell: Typisk tilfeldige effekter.
'Trukket fra en hatt'.
(Uendelig) mange nivåer.
Nivåene ikke av interesse som sådan.

Eksempelet over:

Ubalansert og

blandet (både fast og tilfeldige effekter).

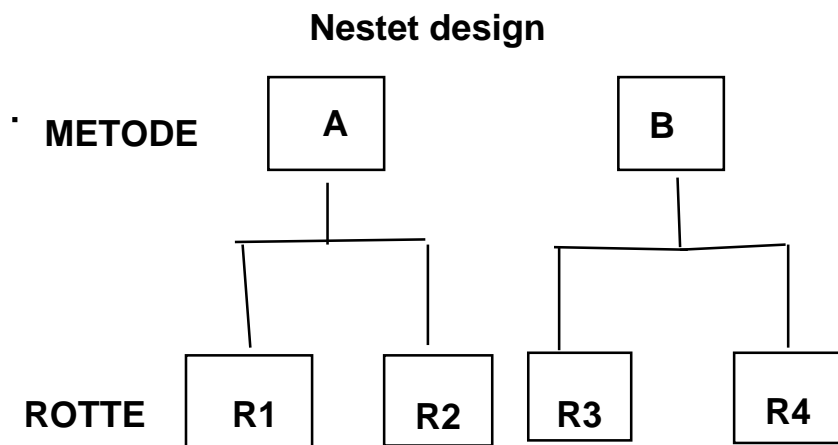
NESTET?

Ikke-nestet:

Behandling	Rotte	Respons
A	R1	2.5
A	R2	2.2
B	R1	2.9
B	R2	3.1

Nestet:

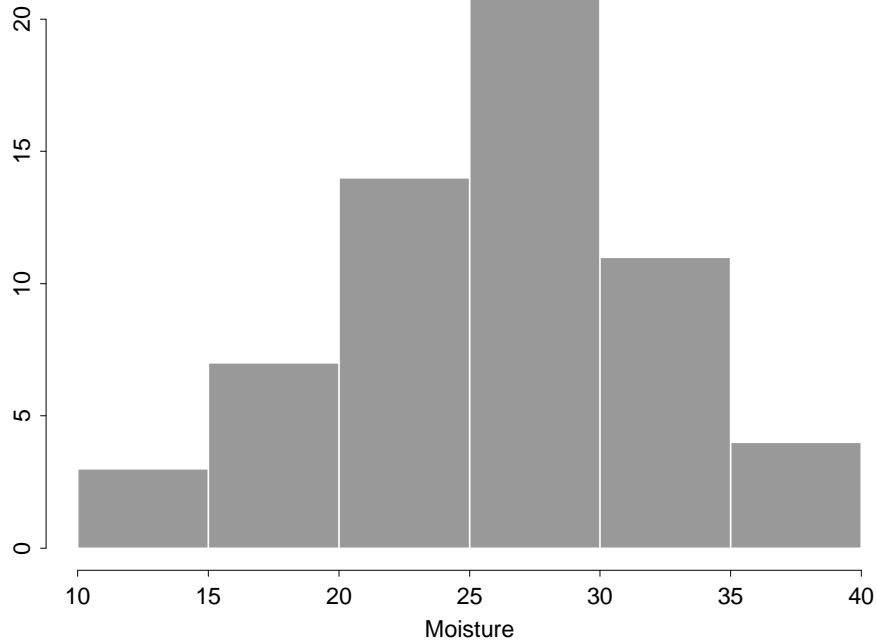
Behandling	Rotte	Respons
A	R1	2.5
A	R2	2.2
B	R3	2.9
B	R4	3.1



EKSEMPEL. PIGMENT DATA

			Respons	
	Batch	Sample	Test	Moisture
1	B1	S1	T1	40
2	B2	S1	T1	26
3	B3	S1	T1	29
4	B4	S1	T1	30
5	B5	S1	T1	19
6	B6	S1	T1	33
7	B7	S1	T1	23
8	B8	S1	T1	34
9	B9	S1	T1	27
10	B10	S1	T1	13
11	B11	S1	T1	25
12	B12	S1	T1	29
13	B13	S1	T1	19
14	B14	S1	T1	23
15	B15	S1	T1	39
16	B1	S2	T1	37
.
.
.
60	B15	S2	T2	28

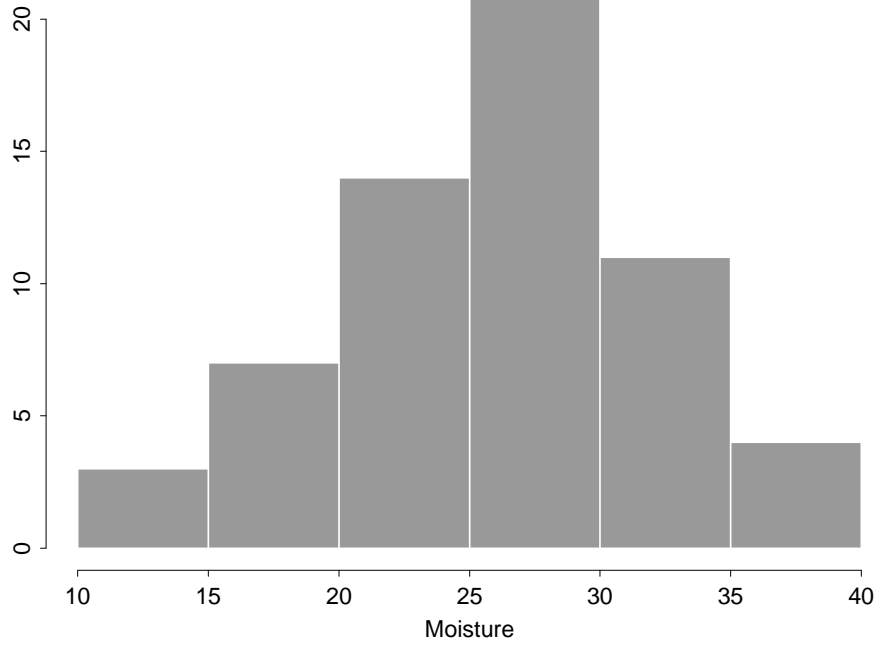
RESPONSEN: MOISTURE



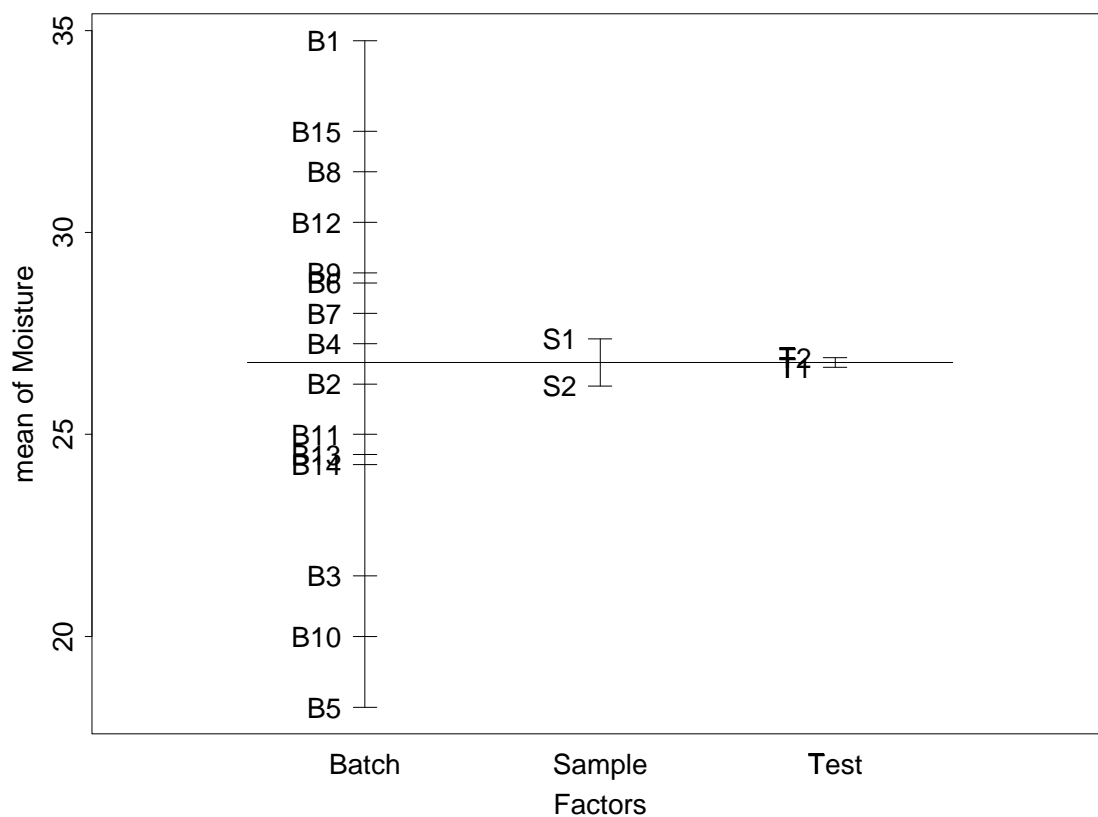
Normalfordeling kan passe.

Transformasjon kan være aktuelt:
logaritmisk, kvadratro, kvadrere.

PROBLEM: FOR STOR SPREDNING



HVILKEN FAKTOR BIDRAR MEST TIL VARIANSEN?



Forklaring:

Gjennomsnittet av data fra B10 er omtrent 20; fra B1 i underkant av 35.

MODELLEN. FIGUR S. 571

IDENTIFIKASJON AV MODELLEN

Det er

- 15 prøver (Batch),
- 2 stikkprøver (Sample) fra hver prøve,
- 2 tester (Test) for hver stikkprøve.

Modellen

- er balansert,
- består av tre tilfeldige effekter,
- er nestet.

FIGUR SIDE 572

DEKOMPOSISJON AV VARIANSEN

Faktor	Varians
Test (residual)	σ_T^2
Sample	σ_S^2
Batch	σ_B^2

Variansen til en observasjon kan dekomponeres:

$$\sigma_B^2 + \sigma_S^2 + \sigma_T^2$$

Primært mål:

Estimere alle tre *varianskomponenter*

DATA SIDE 574-575

DETALJER

$$V_S = \frac{(39.5 - 30)^2/2 + \cdots + (38 - 27)^2/2}{15} = 28.99.$$

$$V_B = \frac{(34.75 - 26.783)^2 + \cdots + (32.50 - 26.783)^2}{14} = 21.64.$$

Det kan vises at

$$EV_T = \sigma_T^2$$

$$EV_S = \sigma_S^2 + \frac{\sigma_T^2}{2}$$

$$EV_B = \sigma_B^2 + \frac{\sigma_S^2}{2} + \frac{\sigma_T^2}{4}$$

RESULTATER, SIDE 577

FIGUR SIDE 578

SKALERTE BIDRAG TIL VARIANSEN

Faktor	Varians	Estimat	Andel av tot.
Test (residual)	σ_T^2	0.92	2.5%
Sample	σ_S^2	28.5	78.0%
Batch	σ_B^2	7.1	19.4%

BMDP-aktig notasjon

$$\begin{aligned} \text{MOISTURE} &= \text{mean} + \text{Batch} + \text{Sample} + \text{Test} \\ y_{bst} &= \mu + \alpha_b + \beta_{s(b)} + \epsilon_{t(bs)} \end{aligned}$$

Paranteser indikerer nesting.

- $b = 1, 2, \dots, 15$. Prøver (Batch),
- $s = 1, 2$. Stikkprøver (Sample) fra hver prøve,
- $t = 1, 2$. Tester (Test) for hver stikkprøve.

μ	Ukjent tall
α_b	Stokastisk variabel med forv. 0 std. avvik σ_α^2
$\beta_{s(b)}$	Stokastisk variabel med forv. 0 std. avvik σ_β^2
$\epsilon_{t(bs)}$	Stokastisk variabel med forv. 0 std. avvik σ_ϵ^2

ANTAGELSER

- (i) Alle stokastiske variable antas uavhengige.
- (ii) Variansen til α_b er en (ukjent) konstant.
Tilsvarende for $\beta_{s(b)}$ og $\epsilon_{t(bs)}$.
- (iii) Inferens
(p-verdier eller konfidensintervaller)
krever store utvalg eller normalfordeling.

Vanlig avvik fra antagelse (ii):
variansen øker med størrelsen på observasjonene.

Transformasjon kan løse dette problemet (og ofte (iii) samtidig).

EKSEMPEL PÅ TRANSFORMASJON (SIDE 335)

RESIDUALPLOT

PREDIKERT MOT RESIDUAL

P-VERDIER. TILFELDIGE EFFEKTER

	Df	Mean Sq	Est.	Var.	F	P
Batch	14	86.5		7.1	1.5	22.6%
Sample %in% Batch	15	58.0		28.5	64.4	0%
Residuals	30	0.9		0.9		

Eksempel:

Vi ønsker å påvise *formelt* at det er en signifikant effekt fra Sample.

Teststørrelse: $58.0/0.9=64.4$

P-verdi: 0.

P-verdiberegninger vesentlig mer komplisert i det ubalanserte tilfellet.

Løsning: BMDP.

NYTT EKSEMPEL: BLANDET MODELL

Data: BMDP manual s. 1149: rat.dat.

	A	B	Y
1	1	1	3
2	1	1	6
3	1	1	3
4	1	1	3
5	1	2	1
6	1	2	2
.	.	.	.
.	.	.	.
.	.	.	.
31	2	8	9
32	2	8	11

SIDE 1150, BMDP-MANUALEN

15 MINUTTER MARI

BEHANDLING, ROTTER, CELLER, ...

	group	rat	cell	sp.den
1	enriched	R1	C1	1.48
2	enriched	R1	C1	1.63
3	enriched	R1	C1	1.68
4	enriched	R1	C1	1.80
5	enriched	R1	C1	1.96
6	enriched	R1	C2	1.48
.
.
.
279	isolated	R26	C46	2.14

MODELL

$$y_{ijkl} = \mu + \tau_i + \beta_{(j)i} + \gamma_{(ij)k} + \epsilon_{(ijk)l} \quad \begin{cases} i = 1, \dots, a \\ j = 1, \dots, b_i \\ k = 1, \dots, n_{ij} \\ l = 1, \dots, n_{ijk} \end{cases} \quad (1)$$

- y_{ijkl} responsverdi nr. 1 for celle k i rotte j i gruppe (behandling) i .
- $a = 3$,
- $b_1 = 11, b_2 = 8, b_3 = 7$
- $n_{11} = 19, n_{12} = 4, \dots$

MODELLIDENTIFIKASJON

Modellen er

- (a) blandet. Behandling er fast effekt; de øvrige tilfeldige.
- (b) nestet; celler og rotter kan ikke brukes om igjen.
- (c) ubalansert. Det er ikke like mange observasjoner i hver kategori.

Faktor	Estimat	Andel av tot.
Rotte	0.0212	16.1%
Cell	0.0386	29.4%
Residualer	0.0717	54.5%

HYPOTESEPRØVING

Tester for faste effekter og tilfeldige effekter blir forskjellige. Vi har tidligere testet tilfeldige effekter; nå kommer faste:

τ_1 måler ‘enriched’ effekt.

Hypotesesituasjonen:

$$H_0 : \tau_1 \leq 0 \text{ vs. } H_A : \tau_1 > 0 \tag{2}$$

(3)

Alternativet svarer til positiv ‘enriched’ effekt.

P-VERDIER

BMDP ²:

.	.	.	.	
.	.	.	.	
.	.	.	.	
PARAM	EST	STD.E	EST/STD.E	P (ONE)
ERR.VAR.	0.072	0.007		
CONST.	1.835	0.047	39.204	0.000
gruppe	0.132	0.061	2.172	0.015
.	.	.	.	
.	.	.	.	
.	.	.	.	

Teststørrelse:

$$\mathbf{EST/STD.ERR} = 0.132/0.061 = 2.172.$$

²Redigert av plasshensyn.

BONFERRONI

Merk at vi i utgangspunktet bare var interessert i denne hypotesen; dersom vi tester alle mulige hypoteser, bør vi dele signifikansnivået med antall hypoteser.

Tester man 100 hypoteser, bør kravet til signifikans reduseres fra 5% til $5\%/100=0.05\%$.

DIVERSE PLOT