

Multisample Inference – del 2 (Rosner 12.5 – 12.7)

Inger Johanne Bakken

Enhet for anvendt klinisk forskning, NTNU
og
Avdeling for forebyggende helsearbeid, SINTEF

NTNU

1

SINTEF

Inference oversettes med Sluttsats

inference n .

- a. The act or process of deriving logical conclusions from premises known or assumed to be true.
- b. The act of reasoning from factual knowledge or evidence.

NTNU

2

SINTEF

Innhold

- 12.1-12.4: One-Way Analysis of Variance (ANOVA)

- 12.5 Case study, effect of lead exposure
Sammenheng ANOVA og lineær regresjon
ANCOVA (En kategorivariabel + kovariater)
- 12.6 Two-Way Analysis of Variance
(To kategorivariabeler)
- 12.7 The Kruskal-Wallis Test
(Ikke-parametrisk versjon av one-way ANOVA)
- Altman Friedmans test
(Ikke-parametrisk versjon av two-way ANOVA)

NTNU

3

SINTEF

ANOVA

Modell: $y_{ij} = \mu_i + e_{ij} = \mu + \alpha_i + e_{ij}$

- y_{ij} : j-te observasjon i utvalg fra gruppe i
 μ_i : forventningsverdien i gruppe i
 μ : forventningsverdien for alle observasjonene "grand mean"
 α_i : avviket i gruppe i fra "grand mean"

NTNU

4

SINTEF

ANOVA

$$\text{Between SS} = \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{y}_i - \bar{y})^2$$

$$\text{Within SS} = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$

$$\text{Total SS} = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$$

NTNU

5

SINTEF

ANOVA (tabell)

Kilde	SS	df	MS	F
Between	A	k-1	A/(k-1)	$MS_{\text{between}}/MS_{\text{within}}$
Within	B	n-k	B/(n-k)	
Total	A+B	n-1		

NTNU

6

SINTEF

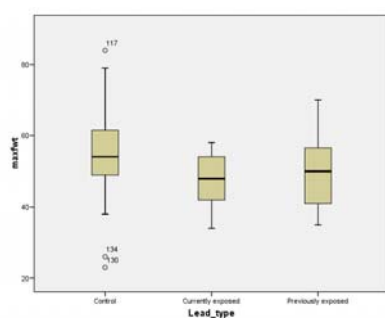
Sammenheng ANOVA og lineær regresjon

Eksempel: Lead exposure

- Data ligger på fagets hjemmeside:
 - lead without outliers.sav
- Hentet fra Rosner CD
 - Beregnet MAXFWT
 - Fjernet outliers (!)
- "Group" angir
 - kontroll (1)
 - tidligere eksponert (2)
 - eksponert nå (3)

<http://folk.ntnu.no/slyderse/medstat/KLMED8005/>

Box plot MAXFWT



ANOVA (Group)

Report

maxfwf			
Lead_type	Mean	N	Std. Deviation
Control	55,10	63	10,935
Currently exposed	47,59	17	7,080
Previously exposed	49,40	15	10,197
Total	52,85	95	10,638

ANOVA

maxfwf					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	966,791	2	483,395	4,598	,012
Within Groups	9671,146	92	105,121		
Total	10637,937	94			

Multipel lineær regresjon vs ANOVA

En kategorisk variabel C med k nivåer kan representeres med k-1 dummy variable

$x_1=1$ hvis $C=2$, ellers 0

$x_2=1$ hvis $C=3$, ellers 0

...

$x_k=1$ hvis $C=k$, ellers 0

Referansegruppen er gitt når alle x 'er = 0

Koding av indikatorvariable

C	x_1	x_2	...	x_{k-1}
1	0	0	...	0
2	1	0	...	0
3	0	1	...	0
...	...	...	...	...
k	0	0	...	1

Her har vi:

C	Grp2	Grp3
1	0	0
2	1	0
3	0	1

Lineær regresjon

$$\text{Modell: } y = \alpha_0 + \beta_1 * \text{grp2} + \beta_2 * \text{grp3} + e$$

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	966,791	2	483,395	4,598	,012 ^a
	Residual	9671,146	92	105,121		
	Total	10637,937	94			

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	55,095	1,292		42,652	,000
	grp2	-7,507	2,802	-,272	-2,679	,009
	grp3	-5,695	2,946	-,196	-1,933	,056

a. Dependent Variable: maxfwt

ANOVA vs Lineær regresjon

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	966,791	2	483,395	4,598	,012
Within Groups	9671,146	92	105,121		
Total	10637,937	94			

ANOVA^a (fra lineær regresjon)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	966,791	2	483,395	4,598	,012 ^a
	Residual	9671,146	92	105,121		
	Total	10637,937	94			

a. Predictors: (Constant), grp3, grp2

b. Dependent Variable: maxfwt

ANCOVA

ANCOVA

"ANOVA" + justering for kovariater (en lineær regresjonsmodell)

Fordeler:

Justering for konfunderende faktorer
Mer presis estimering

Eksempel 12.21 s 589:

$$y = \alpha_0 + \beta_1 * \text{grp2} + \beta_2 * \text{grp3} + \beta_3 * \text{age} + \beta_4 * \text{sex} + e$$

Regression -> Linear

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6017,809	4	1504,452	29,307	,000 ^a
	Residual	4620,128	90	51,335		
	Total	10637,937	94			

a. Predictors: (Constant), sex, grp3, age, grp2

b. Dependent Variable: maxfwt

Regression - > Linear (forts)

Coefficients ^a					
Model	Unstandardized Coefficients		Standardized Coefficients		Sig.
	B	Std. Error	Beta	t	
1 (Constant)	33,958	3,130		10,851	,000
grp2	-5,278	1,974	-,191	-2,674	,009
grp3	-4,927	2,061	-,170	-2,391	,019
age	,024	,002	,705	9,919	,000
sex	-2,408	1,515	-,112	-1,589	,116

a. Dependent Variable: maxfwt

General linear model - > Univariate

Tests of Between-Subjects Effects

Dependent Variable: maxfwt

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6017,809 ^a	4	1504,452	29,307	,000
Intercept	5157,872	1	5157,872	100,475	,000
age	5050,289	1	5050,289	98,380	,000
sex	129,633	1	129,633	2,525	,116
group	549,784	2	274,892	5,355	,006
Error	4620,128	90	51,335		
Total	276011,000	95			
Corrected Total	10637,937	94			

a. R Squared = ,566 (Adjusted R Squared = ,546)

General linear model - > Univariate (forts)

Parameter Estimates

Dependent Variable: maxfwt

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval		Partial Eta Squared
					Lower Bound	Upper Bound	
Intercept	29,032	3,511	8,270	,000	22,057	36,006	,432
age	,024	,002	9,919	,000	,020	,029	,522
sex	-2,408	1,515	-1,589	,116	-5,418	,602	,027
[group=1]	4,927	2,061	2,391	,019	,833	9,021	,060
[group=2]	-,351	2,548	-,138	,891	-5,413	4,710	,000
[group=3]	0 ^a	.	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

Linear regression vs General linear model

Samme modell og dermed samme resultater

Ved bruk av linear regression:
Dummyvariable for kategoriske variable med mer enn to nivåer (group i vårt eksempel)

Artikler om ANCOVA

Pocock SJ et al:
Subgroup analysis, covariate adjustment and baseline comparison in clinical trial reporting: current practice and problems.
Statistics in Medicine 2002; 21: 2917-2930

Vickers AJ and Altman DG:
Statistics notes: Analysing controlled trials with baseline and follow-up measurements.
BMJ 2001; 323: 1123-1124

Klinisk, randomisert utprøving, sammenligning av ulike behandlinger

Måler en gitt variabel før og etter behandling

Alternative analyser av effekt av behandling:

Follow-up scores
Change scores
Analysis of covariance (ANCOVA)

Follow-up vs Change Score

Higher baseline scores (by chance) in treatment group?

- Follow-up score analysis overestimates effect
- Change score analysis underestimates effect (regression to the mean)

Average baseline scores equal?

- Unbiased effect estimate in both analyses, but low power (wide confidence intervals)

Vickers & Altman

Follow-up vs Change Score

Results of trial of acupuncture for shoulder pain

	Placebo (mean)	Acupuncture (mean)	Difference	p
Baseline	53,9	60,4	6,5	
Analysis				
Follow-up	62,3	79,6	17,3	0,008
Change score	8,4	19,2	10,8	0,014

Vickers & Altman 2001

Follow-up vs Change Score vs ANCOVA

Results of trial of acupuncture for shoulder pain

	Placebo (mean)	Acupuncture (mean)	Difference	p
Baseline	53,9	60,4	6,5	
Analysis				
Follow-up	62,3	79,6	17,3	0,008
Change score	8,4	19,2	10,8	0,014
ANCOVA			12,7	0,005

Vickers & Altman 2001

ANCOVA

Follow-up score

= constant + a x baseline score + b x group

= constant + 0,71 x baseline score + 12,7 x group

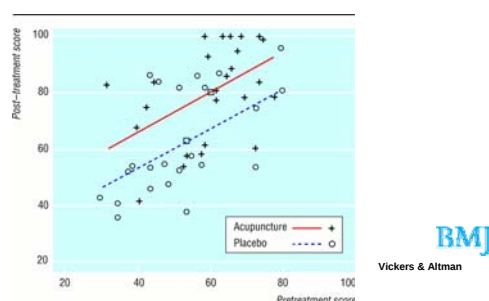
Results of trial of acupuncture for shoulder pain*

	Pain scores (mean and SD)		Difference between means (95% CI)	P value
	Placebo group (n=27)	Acupuncture group (n=25)		
Baseline	53.9 (14)	60.4 (12.3)	6.5	
Analysis				
Follow up	62.3 (17.9)	79.6 (17.1)	17.3 (7.5 to 27.1)	0.008
Change score*	8.4 (14.6)	19.2 (16.1)	10.8 (2.3 to 19.4)	0.014
ANCOVA			12.7 (4.1 to 21.3)	0.005

Analysis reported by authors.

Vickers & Altman 2001

Pretreatment and post-treatment scores in each group showing fitted lines



Estimert forskjell mellom gruppene fra ANCOVA er vertikal avstand mellom linjene

Korrelasjon og justering

Svak korrelasjon ($\rho < 0,3$)

Manglende balanse i kovariatene uten betydning

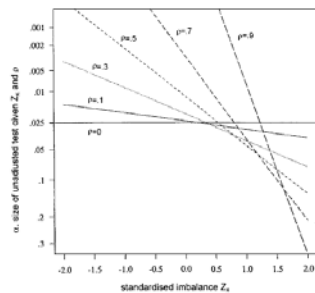
Sterk korrelasjon ($\rho > 0,5$)

Viktig å justere for kovariater selv ved balanse mellom gruppene

Pocock, tredje avsnitt s 2924

NB Vanligvis sterk korrelasjon mellom baseline og outcome (eks fedme)

Korrelasjon mellom kovariat og behandlingseffekt (1:2)

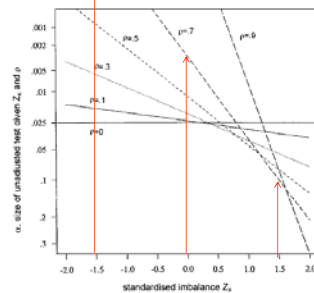


Sann $\alpha = 0,025$

$\rho = 0$:
manglende balanse har
ingen betydning for
resultatene

S. J. POCKOCK ET AL.

Betydning av korrelasjon mellom kovariat og behandlingseffekt (2:2)



Sann $\alpha = 0,025$ (effekt av
behandling)

$\rho = 0,7$: x har stor
betydning for resultatene,
også ved perfekt balanse

$Z_x = 0$: $p = 0,003$
 $Z_x = 1,5$: $p = 0,102$
 $Z_x = -1,5$: $p < 0,0001$

S. J. POCKOCK ET AL.

12.6 To-veis variansanalyse

Eksempel fra Rosner

Table 12.14 s 592 (12.13 i 5th ed)
Mean systolic blood pressure by dietary group and sex

Dietary group		Male	Female
SV	mean	109.9	102.6
	n	138	88
LV	mean	115.5	105.2
	n	26	37
Normal	mean	128.3	119.6
	n	240	220

To-veis variansanalyse

(Modell en-veis: $y_{ij} = \mu + \alpha_i + e_{ij}$)

Modell to-veis: $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$

Hvor:

- y_{ijk} Målt verdi for person nr k i rad i og kolonne j
- μ Forventningsverdien alle obs under ett "Grand mean"
- α_i Effekt av rad i
- β_j Effekt av kolonne j
- γ_{ij} Interaksjon/samspill mellom rad i og kolonne j
- e_{ijk} Støyledd (Normalfordelt med forventningsverdi 0 og variasjon σ^2)

To-veis variansanalyse

(Modell en-veis: $y_{ij} = \mu + \alpha_i + e_{ij}$)

Modell to-veis: $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$

I vårt eksempel har vi:

- tre rader (diettgrupper)
- to kolonner (to kjønn)

Fremgangsmåte

Test først samspill, dvs test hypotesen $H_0: \gamma_{ij}=0$ for alle i og j

- Hvis samspill er signifikant, gå videre med modellen (Eq 12.22)

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

Konvensjon: $\sum_{i=1}^r \alpha_i = \sum_{j=1}^c \beta_j = 0 \quad \sum_{j=1}^c \gamma_{ij} = \sum_{i=1}^r \gamma_{ij} = 0$

- Hvis samspillet ikke er signifikant brukes modell uten samspill

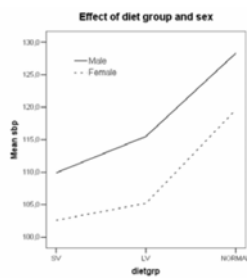
$$y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk}$$

Samspill

Hvis det er samspill er effekt av rad forskjellig i forskjellige kolonner.

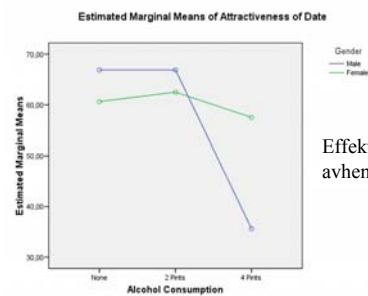
(I dette tilfellet: effekten av diett ville vært avhengig av kjønn)

Samspill (1)



Linjene er (tilnærmet) parallelle, ikke samspill

Samspill (2)



Effekt av alkohol er avhengig av kjønn: samspill

12.6 (forts) To-veis ANCOVA

Diett og kjønn, betydning for blodtrykk
Justerer for alder og vekt

Ser bort fra interaksjon, modellen blir:

$$y = \mu + \beta_1 * x_1 + \beta_2 * x_2 + \beta_3 * \text{kjønn} + \beta_4 * \text{alder} + \beta_5 * \text{vekt} + e$$

12.7 Rosner & Altman 12.3.5: Ikke-parametriske tester

Sammenlikning k grupper

	Normalfordelte data	Ikke normalfordelt
k = 2	To-utvalgs t-test 1)	Wilcoxon-Mann-Whitney's test
k > 2	enveis ANOVA 1)	Kruskall-Wallis' test

1) Alternativt multipel lineær regresjon med indikatorvariable for gruppene. Kan også brukes ved justering for kovariater (f.eks alder, kjønn).

Eksempel

Tabell 12.17 Rosner.
Ocular anti-inflammatory effects of 4 drugs on lid closure

Rabbit no	Indomethicin	Aspirin	Piroxicam	BW775C
	Score	Score	Score	Score
1	+2	+1	+3	+1
2	+3	+3	+1	0
3	+3	+1	+2	0
4	+3	+2	+1	0
5	+3	+2	+3	0
6	0	+3	+3	-1

NB! 24 kaniner!

Table 12.18 (12.17 in 5th ed). Assignment of ranks

Lid-closure score	Frequency	Range of ranks	Average rank
-1	1	1	1.0
0	5	2 – 6	4.0
1	5	7 – 11	9.0
2	4	12 – 15	13.5
3	9	16 – 24	20.0

Table 12.17 (12.16 in 5th ed):
Ocular anti-inflammatory effects of 4 drugs on lid closure

Rabbit no	Indomethicin		Aspirin		Piroxicam		BW775C	
	Score	Rank	Score	Rank	Score	Rank	Score	Rank
1	+2	13.5	+1	9.0	+3	20.0	+1	9.0
2	+3	20.0	+3	20.0	+1	9.0	0	4.0
3	+3	20.0	+1	9.0	+2	13.5	0	4.0
4	+3	20.0	+2	13.5	+1	9.0	0	4.0
5	+3	20.0	+2	13.5	+3	20.0	0	4.0
6	0	4.0	+3	20.0	+3	20.0	-1	1.0

NB "Six rabbits in each group" Totalt 24 kaniner

Kruskall-Wallis test (s 601, 603)

Uten "ties":
$$H = H^* = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

$$H^* = \frac{12}{N(N+1)} \sum_{i=1}^k n_i (\bar{R}_i - \bar{R})^2$$

Med "ties":

$$H = \frac{H^*}{1 - \frac{\sum_{j=1}^g (t_j^3 - t_j)}{N^3 - N}}$$

NB! Alle $n_i \geq 5$

hvor t_j er antall sammenfallende observasjoner i klynge nr j
Under H_0 er H khikvadratfordelt med k-1 frihetsgrader

SPSS: Nonparametric tests -> k independent samples

Ranks				Test Statistics ^{a,b}	
score	med	N	Mean Rank		score
1		6	16,25	Chi-Square	11,804
2		6	14,17	df	3
3		6	15,25	Asymp. Sig.	,008
4		6	4,33	Exact Sig.	,003
Total		24		Point Probability	,000

a. Kruskal Wallis Test
b. Grouping Variable: med

Hvilke grupper er forskjellige?

Dunn prosedyren (ikke i SPSS):

Regn ut:

$$z_{ij} = \frac{\bar{R}_i - \bar{R}_j}{\sqrt{\frac{N(N+1)}{12} x \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}}$$

Forkast H_0 hvis $|z| > z_{1-\alpha^*}$ hvor $\alpha^* = \alpha/k(k-1)$

Mer om 24 kaniner...

$$\begin{aligned} z_{12} &= 0,51 & z_{13} &= 0,24 \\ z_{14} &= 2,92 & z_{23} &= -0,27 \\ z_{24} &= 2,41 & z_{34} &= 2,67 \end{aligned}$$

$$\begin{aligned} \alpha^* &= \alpha/k(k-1) = 0,005/(4*3) = 0,0042 \\ z_{1-0,0042} &= 2,64 \end{aligned}$$

Altså:

Gruppe 1 og 4 er statistisk signifikant forskjellige
Gruppe 3 og 4 er statistisk signifikant forskjellige

Friedmans to-veis ANOVA

- Krever ikke normalfordeling
- n subjekter og k grupper. Få eller ingen sammenfallende observasjoner (for hvert subjekt)
- H_0 : Ingen forskjell på gruppene

Tabell 12.9 Altman

- 8 personer tester 4 forskjellige redningsdrakter (A, B, C, D)

id	A	B	C	D	id	A	B	C	D
1	308	132	454	64	1	3	2	4	1
2	102	526	0	28	2	3	4	1	2
3	182	134	96	30	3	4	3	2	1
4	268	324	264	90	4	3	4	2	1
5	166	228	134	34	5	3	4	2	1
6	332	296	458	6	6	3	2	4	1
7	198	350	200	90	7	2	4	3	1
8	28	274	16	24	8	3	4	1	2
snitt	198	283	302	45,7	R(tot)	24	27	19	10

Friedmans test

$$H = \frac{12}{nk(k+1)} \sum_{i=1}^k R_i^2 - 3n(k+1)$$

Hvor H er tilnærmet χ^2_{k-1} under H_0

I eksempelet er:

$$H = \frac{12}{8*4*(4+1)} * (24^2 * 27^2 * 19^2 * 10^2) - 3*8*(4+1) = 12,45$$

$$p = P(\chi^2_{k-1} \geq 12,45) = 0,006$$

Friedmans test i SPSS

Analyze -> Nonparametric tests -> k related samples
Velg opsjonen "exact"

Ranks		Test Statistics ^a	
	Mean Rank		
A	3,00	N	8
B	3,38	Chi-Square	12,450
C	2,38	df	3
D	1,25	Asymp. Sig.	,006
		Exact Sig.	,003
		Point Probability	,000

a. Friedman Test

Hvilke grupper er forskjellige?

- Friedmans test viser om minst to grupper er statistisk signifikant forskjellige
- Par av grupper kan sammenliknes ved hjelp av Wilcoxon test for matchede par. Juster for multiple sammenlikninger
- Friedmans test for 2 grupper tilsvarer tegntesten

Noen eksempler i SPSS

Diettdatasettet fra Rosner-boka ikke tilgjengelig

Velger eksempeldatasett fra

Andy Field: Discovering Statistics with SPSS:

Libido.sav (ANCOVA med en "fixed" variabel og justering for én kovariat)

Googles.sav (to-veis ANOVA med interaksjon)

Dagens tema var

- ANOVA en-veis og to-veis
- ANCOVA
- Sammenhengen med lineær regresjon
- I SPSS: ANOVA one-way, linear regression, General linear model UNIVARIATE
- Ikke-parametriske tester for multiple sammenlikninger (ikke-parede og parede)