

# Kapitel 13

## Reliabilitet og enighed

Peter Tibert Stoltze  
stat@peterstoltze.dk

Elementær statistik  
F2011

Version 11. april 2011

1 / 23

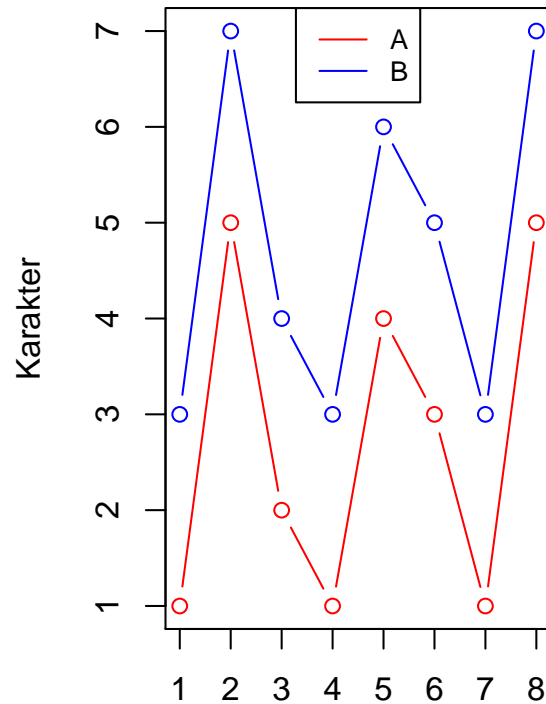
## Indledning

- ▶ En observation er sammensat af en sand værdi og en målefejl — vi ser selvfølgelig helst, at målefejlen er så lille som mulig
- ▶ Målefejlen kan vi forsøge at kvantificere, så vi kan vurdere om data er pålidelige (“reliable”)
- ▶ Hvis observationen er dannet ved et subjektivt skøn er det særlig vigtig at få dannet et skøn over målefejlen

2 / 23

# Pålidelighed og enighed

- ▶ Pålidelighed er ikke det samme som enighed.
- ▶ I eksemplet er A konsekvent to point under B, så data er pålidelige men der er ikke enighed
- ▶ Men er der enighed, så er der også pålidelighed
- ▶ Enighed medfører pålidelighed, men ikke omvendt



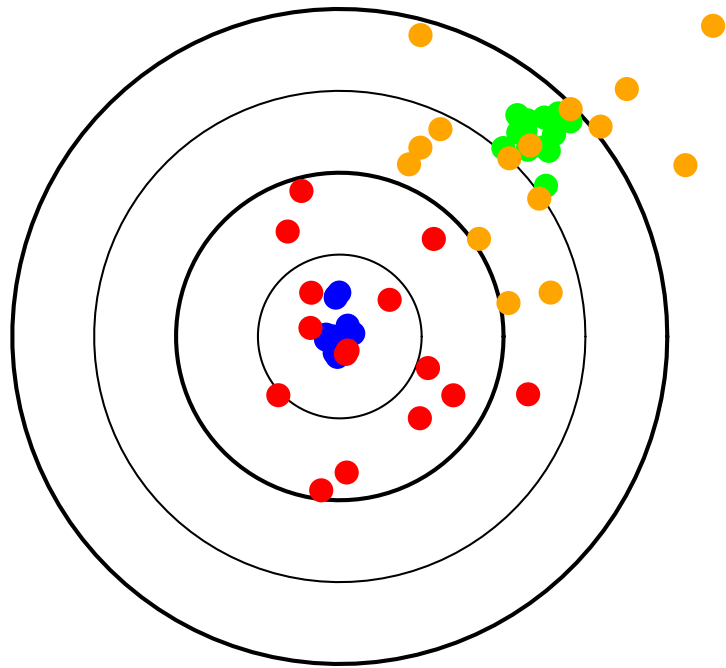
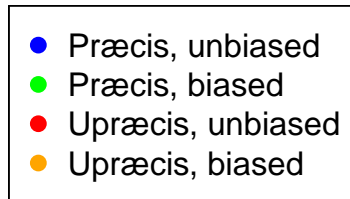
3 / 23

## Indledning

- ▶ *Intersubjektiv* pålidelighed eller enighed knytter sig til flere bedømmers observationer af samme objekt
- ▶ *Intrasubjektiv* pålidelighed eller enighed knytter sig til én bedømmers gentagne observationer af en gruppe af objekter
- ▶ Løst oversat: *Inter* betyder imellem og *intra* betyder indenfor
- ▶ Vi bruger samme metoder i de to tilfælde

4 / 23

# Præcision og nøjagtighed



5 / 23

## Reliabilitet

- ▶ Direkte oversat fra *reliability* (eng: pålidelighed), hvilket dækker om vi i bredeste forstand kan fæstne lid til vore data — virker data rimelige?
- ▶ Dette kan vurderes mekanisk uden videre stillingtagen til det egentlige indhold af data
- ▶ Der bør være en vis variation i de sande værdier, for ellers er det næsten givet, at den tilfældige variation (målefejlen) vil blive dominerende

6 / 23

## Cronbachs $\alpha$ : Teori

- ▶ Bygger faktisk på den variansanalysemodel for blokforsøg, som vi gennemgik sidst
- ▶ En tosidet variansanalyse med én observation for hver kombination af de to faktorerers niveauer
- ▶ Vores hovedinteresse er den første faktor — den anden er observatøren
- ▶ Parametrisk mål der forudsætter, at responset er målt på mindst intervallskala — benyttes dog ofte for ordinale data

7 / 23

## Cronbachs $\alpha$ : Teori (fortsat)

Variation	Frihedsgrader DF	Middelsum MS	F-værdi $F_{\text{obs}}$
Rækker (behandling)	$a - 1$	$MS_a = \frac{SAK_a}{a - 1}$	$F_a = \frac{MS_a}{MS_e}$
Søjler (blokke)	$b - 1$	$MS_b = \frac{SAK_b}{b - 1}$	$F_b = \frac{MS_b}{MS_e}$
Residual	$(a - 1)(b - 1)$	$MS_e = \frac{SAK_e}{(a - 1)(b - 1)}$	
Total	$ab - 1$		

- ▶ Middeltkvadratsummerne er vores variansestimater:

$$\hat{\sigma}_a^2 = MS_a \quad \text{og} \quad \hat{\sigma}_e^2 = MS_e$$

8 / 23

## Cronbachs $\alpha$ : Teori (fortsat)

- ▶ Fra variansanalysetabellen får vi variansestimater for  $s_a^2$  (behandling, række eller objekt) og  $s_e^2$  (residual)
- ▶ Disse samles til Cronbachs  $\alpha$  efter følgende formel:

$$\alpha = \frac{s_a^2 - s_e^2}{s_a^2} = 1 - \frac{s_e^2}{s_a^2} = 1 - \frac{1}{F_a}$$

- ▶ Vi vedtager, at  $\alpha > 0,6$  betyder en acceptabel grad af reliabilitet, mens  $\alpha > 0,9$  betyder en meget høj grad af reliabilitet (øvre grænseværdi for  $\alpha$  er 1)

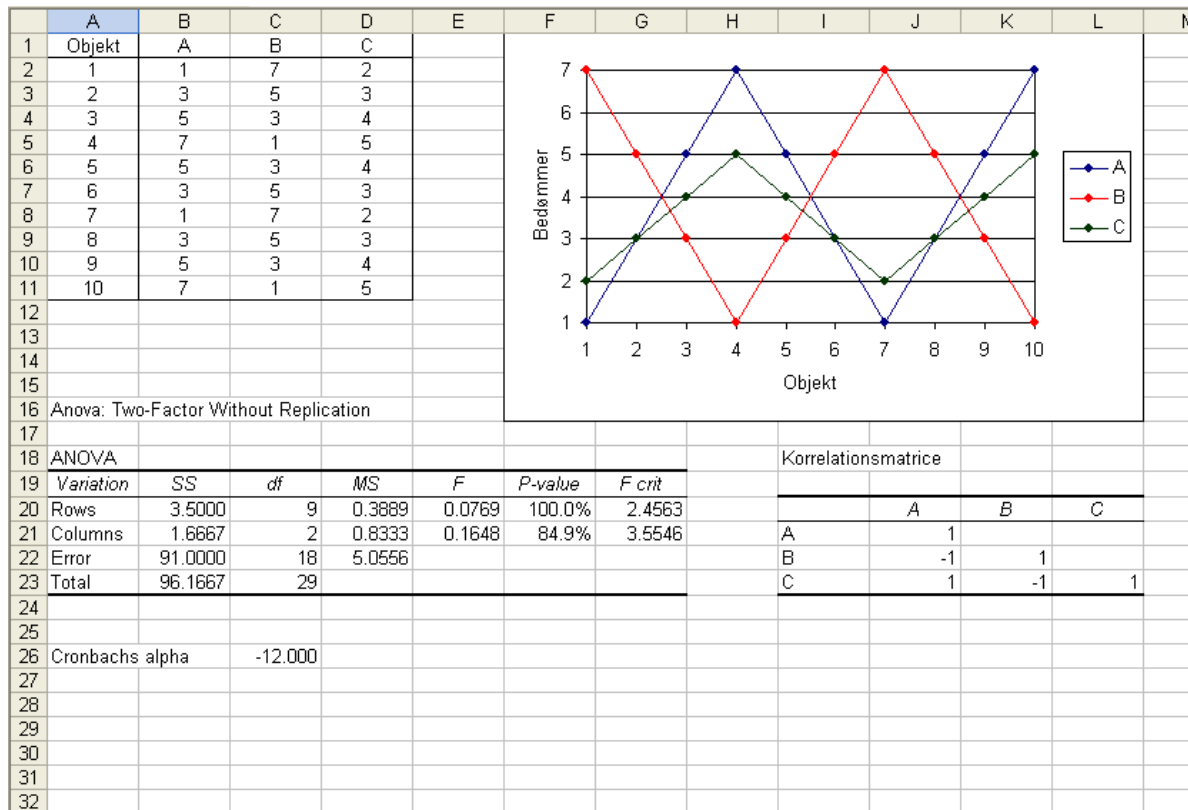
9 / 23

## Cronbachs $\alpha$ : Eksempel 1 (Tabel 13.2)

	A	B	C	D	E	F	G	H	I	J	K	L
1	Objekt	A	B	C								
2	1	1	4	2								
3	2	5	7	6								
4	3	2	1	2								
5	4	1	1	1								
6	5	4	3	5								
7	6	3	4	4								
8	7	1	2	1								
9	8	5	7	6								
10	9	3	2	4								
11	10	2	4	3								
12												
13												
14												
15												
16	Anova: Two-Factor Without Replication											
17												
18	ANOVA					Korrelationsmatrice						
19	Variation	SS	df	MS	F	P-value	F crit					
20	Rows	84.133	9	9.348	13.078	0.0%	2.456					
21	Columns	3.800	2	1.900	2.658	9.7%	3.555					
22	Error	12.867	18	0.715								
23	Total	100.800	29									
24												
25												
26	Cronbachs alpha:		0.924									
27												
28												
29												
30												
31												
32												

10 / 23

## Cronbachs $\alpha$ : Eksempel 2 (Figur 13.1)



11 / 23

## Enighed

- ▶ Enighed beskriver identitet mellem observerede værdier
- ▶ Passer til data målt på nominal skala og for diskrete variable målt på de højere skalatyper
- ▶ Problematisk for kontinuerte data (fx. 184, 186, 188 ms)
- ▶ Hvis sandsynligheden  $p$  for enighed ved tilfældighed er meget høj, så er fuldstændig enighed ikke meget værd — og vores enighedsmål inkluderer derfor dette  $p$

12 / 23

## Tinsley-Weiss' $T$ : Teori

- ▶ Tinsley-Weiss'  $T$  benyttes når data er målt på ordinal skala eller højere og beregnes som

$$T = \frac{N_1 - pN}{N - pN}$$

hvor  $N_1$  er antal objekter med samme observationsværdi,  $N$  er total antal objekter, og  $p$  er sandsynligheden for enighed ved en tilfældighed

- ▶ Formlen udtrykker altså forholdet mellem den observerede enighed og den maksimalt opnåelige enighed, *idet der fratrækkes den grad af enighed, der må forventes at være pga tilfældighed*

13 / 23

## Tinsley-Weiss' $T$ : Teori (fortsat)

- ▶ Sandsynligheden  $p$  for enighed ved tilfældighed blandt  $k$  bedømmere og en skala med  $n$  trin beregnes på følgende måde:
  - ▶ Kræves der fuldstændig enighed beregnes  $p$  som

$$p = \left(\frac{1}{n}\right)^{k-1} = n^{1-k}$$

- ▶ Accepteres en afvigelse på ét skalatrin beregnes  $p$  som

$$p = \frac{(n-1) \sum_{i=1}^k 2^{i-1} + 1}{n^k}$$

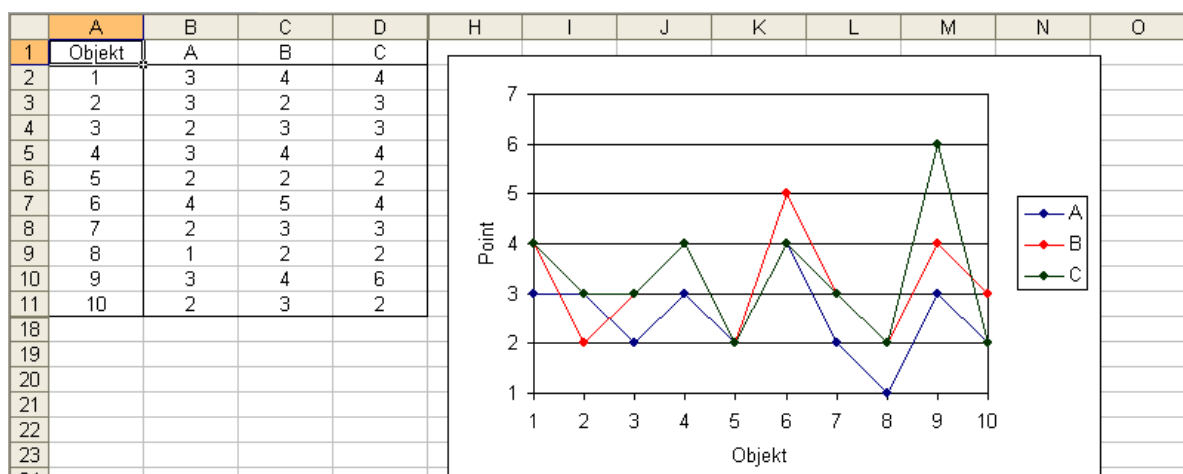
14 / 23

## Tinsley-Weiss' $T$ : Teori (fortsat)

- ▶ Tinsley-Weiss'  $T$  kan højst have værdien 1 (når  $N_1 = N$ )
- ▶ Hvis omfanget af enighed svarer til antallet ved tilfældighed (altså  $N_1 \approx pN$ ) er  $T$  cirka 0
- ▶ I den (tænkte) situation, at omfanget af enighed er mindre end ved tilfældighed (altså  $N_1 < pN$ ), da bliver  $T$  negativ
- ▶ Der er ingen nedre grænse for  $T$ , da  $p$  kan være vilkårligt tæt på 1 hvorved nævneren går mod nul fra højre

15 / 23

## Tinsley-Weiss' $T$ : Eksempel (Tabel 13.3)



16 / 23

## Tinsley-Weiss' $T$ : Eksempel (Tabel 13.3, fortsat)

	A	B	C	D	E	F	G	H
1	Objekt	A	B	C		Fuldstændig enighed	Partiel enighed	
2	1	3	4	4		FALSK	SAND	
3	2	3	2	3		FALSK	SAND	
4	3	2	3	3		FALSK	SAND	
5	4	3	4	4		FALSK	SAND	
6	5	2	2	2		SAND	SAND	
7	6	4	5	4		FALSK	SAND	
8	7	2	3	3		FALSK	SAND	
9	8	1	2	2		FALSK	SAND	
10	9	3	4	6		FALSK	FALSK	
11	10	2	3	2		FALSK	SAND	
12	Antal enige				N1	1	9	
13	Antal i alt				N	10	10	
14	Antal kategorier				n	7	7	
15	Antal bedømmere				k	3	3	
16	Sandsynlighed for enighed ved tilfældighed				p	0,0204	0,1254	
17	Forventet antal enige ved tilfældighed				p*N	0,20	1,25	
18					T	<b>0,0813</b>	<b>0,8857</b>	
19								

17 / 23

## Tinsley-Weiss' $T$ : Eksempel (Tabel 13.3, fortsat)

- ▶ Med stramt krav til enighed er  $T = 0,0813$
- ▶ Med løsere krav til enighed er  $T = 0,8857$
- ▶ Vi må konkludere, at der har været en ganske høj grad af enighed, hvilket støttes af grafen
- ▶ Men i eksemplet gives der  $3 \times 10 = 30$  bedømmelser, hvoraf kun tre karakterer ikke er 2, 3 eller 4 — dette kunne tyde på (for?) lav objektvariation

18 / 23

## Cohens $\kappa$ : Teori

- ▶ Cohens  $\kappa$  anvendes når data er målt på nominalskala, og enighed defineres her som **fuldstændig** enighed (eventuelt beslægtede kategorier kan slås sammen)
- ▶ Princippet i beregningen er analogt til Tinsley-Weiss  $T$ , nemlig den observerede enighed i forhold til den maksimalt opnåelige enighed idet begge fratrækkes den enighed, man kan forvente ved tilfældighed

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

19 / 23

## Cohens $\kappa$ : Teori (fortsat)

- ▶ Antag  $N$  objekter,  $k$  bedømmere og en skala med  $v$  kategorier
- ▶ Da beregnes andelen af parvise sammenligninger med enighed  $p_0$  som

$$p_0 = \frac{\sum_{i=1}^N \sum_{j=1}^v n_{ij}^2 - Nk}{Nk(k-1)}$$

idet  $n_{ij}$  er antal bedømmere, der placerer objekt  $i$  i kategori  $j$

- ▶ Enighed ved tilfældighed  $p_e$  beregnes som

$$p_e = \sum_{j=1}^v p_j^2 \quad \text{hvor} \quad p_j = \frac{\sum_{i=1}^N n_{ij}}{Nk}$$

20 / 23

# Cohens $\kappa$ : Teori (fortsat)

- ▶ Af definitionen

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

kan vi se at  $p_0 \approx 1$  (meget stor andel af enighed) giver  $\kappa \approx 1$

- ▶ En generel vurdering af  $\kappa$  kan være

$0,4 \leq \kappa < 0,6$  Moderat enighed

$0,6 \leq \kappa < 0,8$  Substantiel enighed

$0,8 \leq \kappa$  Næsten perfekt enighed

# Cohens $\kappa$ : Eksempel (Tabel 13.4)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Objekt	Bedømmer			Antal svar i kategori (j)								
2	(i)	A	B	C	a	b	c						
3	1	b	b	b	0	3	0						
4	2	a	a	a	3	0	0						
5	3	a	a	a	3	0	0						
6	4	c	c	b	0	1	2						
7	5	b	b	b	0	3	0						
8	6	b	b	b	0	3	0						
9	7	a	a	a	3	0	0						
10	8	c	a	c	1	0	2						
11				Sum	10	10	4						
12													
13													
14													
15				N	8								
16				k	3								
17													
18				$p_0$	0.8333								
19				$p_j$	0.4167	0.4167	0.1667						
20				$p_e$	0.3750								
21													
22				kappa	0.7333								
23													
24													
25													
26													
27													
28													
29													
30													
31													
32													

## Cohens $\kappa$ : Eksempel (Tabel 13.4, *fortsat*)

- ▶ Vi finder  $\kappa = 0,733$  hvilket svarer til substantiel enighed (grænsende til det perfekte)
- ▶ De mulige konfigurationer af  $n_{ij}$  for givet  $i$  er
  - ▶ 3:0:0 (perfekt enighed)
  - ▶ 2:1:0 (partiel enighed)
  - ▶ 1:1:1 (ingen enighed)

og pointen i beregningen af  $p_0$  er at  $3^2 > 2^2 + 1^2 > 3 \times 1^2$