



NTNU

Det skapende universitet

Ikke-parametriske metoder

KLMED 8001

25 november 2013

Stian Lydersen

www.ntnu.no

Om du ønsker, kan du sette inn navn, tittel på foredraget, o.l. her.

2

Referanser:

- Aalen, O. O. et al: "Statistiske metoder i medisin og helsefag." Gyldendal akademisk, 2005.
- Rosner, B.: "Fundamentals of biostatistics" 7th ed. Brooks/Cole, 2010.
- Altman, D. G.: "Practical statistics for medical research." Chapman & Hall / CRC, 1991.

NTNU
Det skapende universitet

www.ntnu.no

3

Parametriske vs ikke-parametriske metoder

- Parametriske metoder:
 - Forutsetter bestemt parametrisk sannsynlighetsfordeling, f.eks normalfordelingen
- Ikke-parametriske metoder
 - Få eller ingen forutsetninger om fordelingen
 - Kunne kalles fordelings-uavhengige metoder
 - Forutsetter også *uavhengige* observasjoner!

Valg av parametrisk vs ikke-parametrisk metode baseres *ikke* på størrelsen på utvalget!

NTNU
Det skapende universitet

www.ntnu.no

Om du ønsker, kan du sette inn navn, tittel på foredraget, o.l. her.

4

Ikke-parametriske metoder

- Et utvalg eller to parede utvalg:
 - Tegntesten (The Sign Test)
 - Wilcoxon's signed rank test
- To uavhengige utvalg
 - Wilcoxon's rangsum-test (også kalt Wilcoxon-Mann-Whitney's test eller Mann-Whitney's U-test)
- Tre eller flere uavhengige utvalg:
 - Ikke-ordnede utvalg (tilsv. *enveis* ANOVA): Kruskal-Wallis' test
 - Ordnete utvalg: Jonckheere-Terpstra's test
- Tilsvarende *toveis* ANOVA:
 - Friedman's test
- Korrelasjon:
 - Spearman's rho
 - Kendall's tau

NTNU
Det skapende universitet

www.ntnu.no

5

Tegntesten (The sign test)

NTNU
Det skapende universitet

www.ntnu.no

6

Tabell 8.5 i Aalen et al
"Statistiske metoder i medisin og helsefag" 2006
Aggressivitetsscore for tvillingpar

Tvillingpar	Førstefødt	Sistefødt	Differense
1	88	86	2
2	71	77	-6
3	77	76	1
4	68	64	4
5	91	96	-5
6	72	72	0
7	77	65	12
8	91	90	1
9	70	65	5
10	80	71	9
11	88	81	7
12	89	72	17

NTNU
Det skapende universitet

www.ntnu.no

7

H_0 : Det er ingen forskjell i forventet score mellom første- og sistefødte tvilling
 H_1 : Det er forskjell i forventet score mellom første- og sistefødte tvilling

Tegntesten: Tell opp antall positive differanser C (her $c_{obs} = 9$)
 blant de n som er forskjellig fra 0 (her $n=11$)

Under nullhypotesen er C binomisk fordelt ($n, 1/2$)

Forkast nullhypotesen hvis C avviker "mye" fra $n/2$.

Her er $C > n/2$ så (Rosner eqn 9.3)

$$p\text{-verdi} = 2P(C \geq 9) = 2 \sum_{k=c_{obs}}^n \binom{n}{k} \left(\frac{1}{2}\right)^n = 2(0.0269 + .0054 + .0005) = 0.0656$$

NTNU
Det skapende universitet

www.ntnu.no

8

Eks 9.7, Rosner

Gir salve A eller salve B best beskyttelse mot sola?

Hver forøksperson får en salve på hver arm (randomisert)

Tre mulig utfall:

- 1) Arm A mindre rød enn arm B
- 2) Arm B mindre rød enn arm A
- 3) Begge like røde

NTNU
Det skapende universitet

www.ntnu.no

9

Resultat:

45 personer,
 22 stk A best, 18 stk B best, 5 stk begge like bra.

$c_{obs}=18, n=40$

Binomisk fordeling:
 $p\text{-verdi} = 2 * P(C \leq 18) = 0.636$

Rosner bruker normalfordelingstilnærming (hvorfor?)
 $p\text{-verdi} \approx 0.635$

NTNU
Det skapende universitet

www.ntnu.no

10

Wilcoxon's test for paradata
 (Wilcoxon's signed rank test)
 = Wilcoxon's ettutvalgstest

NTNU
Det skapende universitet

www.ntnu.no

11

9.3
 Wilcoxon's test for paradata (Wilcoxon's signed rank test)
 = Wilcoxon's ettutvalgstest

For forsøksenhet nr i (f.eks pasient):

x_i (f.eks grad av rødhet ved salve A)
 y_i (f.eks grad av rødhet ved salve B)
 $d_i = x_i - y_i$

Se bort fra de d_i som er lik 0
 Ranger de øvrige etter økende absoluttverdi (avstand fra 0)
 La R_1 være rangsum for positive d_i -er.

NTNU
Det skapende universitet

www.ntnu.no

12

Tabell 8.6 i Aalen et al

Tvillingpar i	Førstefødt x_i	Sistefødt y_i	Differanse d_i	Rang r_i
1	88	86	2	3
2	71	77	-6	7
3	77	76	1	1,5
4	68	64	4	4
5	91	96	-5	5,5
6	72	72	0	
7	77	65	12	10
8	91	90	1	1,5
9	70	65	5	5,5
10	80	71	9	9
11	88	81	7	8
12	89	72	17	11

$R_1 = 3+1,5+4+10+1,5+5,5+9+9+11=53,5$

NTNU
Det skapende universitet

www.ntnu.no

13

n observasjoner (forskjellig fra 0)
Summen av alle rangene er $1+2+\dots+n = n(n+1)/2$

Under nullhypotesen:

$$E(R_1) = n(n+1)/4$$

Forkast nullhypotesen hvis R_1 avviker "mye" fra $n(n+1)/4$

NTNU
Det skapende universitet

www.ntnu.no

14

R_1 er tilnærmet normalfordelt (Rosner: Hvis $n \geq 16$)

Hvis ingen sammenfallende observasjoner ("ties"):
 $SD(R_1) = \sqrt{n(n+1)(2n+1)/24}$

Hvis t_i sammenfallende i gruppe nr i:
 $SD(R_1) = \sqrt{n(n+1)(2n+1)/24 - \sum_{i=1}^g (t_i^3 - t_i)/48}$

Alternativ formel:
 $SD(R_1) = \sqrt{\sum_{j=1}^n r_j^2}$

Hvor r_j er rangen til observasjon nr j, summer over alle observasjonene.

NTNU
Det skapende universitet

www.ntnu.no

15

Eksempel Aalen et al:

$$E(R_1) = 11(11+1)/4 = 33$$

$$SD(R_1) = \sqrt{11(11+1)(2 \cdot 11 + 1)/24 - [(2^3 - 2) + (2^3 - 2)]/48}$$

$$= \sqrt{126.5 - 0.25} = 11.24$$

$$T = \frac{R_1 - E(R_1)}{SD(R_1)} = \frac{53.5 - 33}{11.24} = 1.824$$

P-verdi = $2P(Z > 1.824) = 0.068$

NTNU
Det skapende universitet

www.ntnu.no

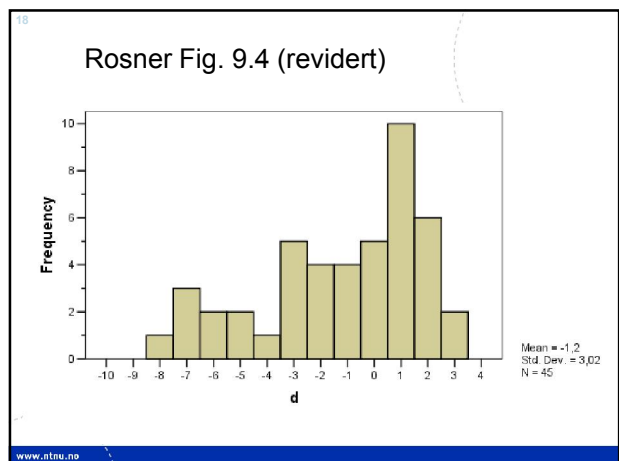
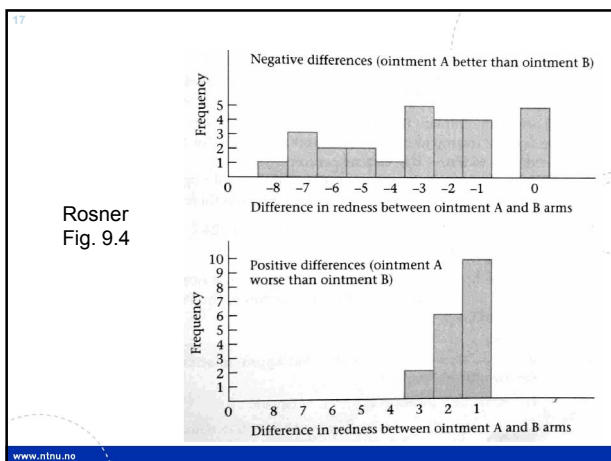
16

Rosner Table 9.1

d _j	Negative		Positive		Number of people with same absolute value	Range of ranks	Average rank
	d _j	f _j	d _j	f _j			
10	-10	0	10	0	0	—	—
9	-9	0	9	0	0	—	—
8	-8	1	8	0	1	40	40.0
7	-7	3	7	0	3	37-39	38.0
6	-6	2	6	0	2	35-36	35.5
5	-5	2	5	0	2	33-34	33.5
4	-4	1	4	0	1	32	32.0
3	-3	5	3	2	7	25-31	28.0
2	-2	4	2	6	10	15-24	19.5
1	-1	4	1	10	14	1-14	7.5
		22		18			
0	0	5					

NTNU
Det skapende universitet

www.ntnu.no



19

Eks 9.12

$$R_1 = 10(7.5) + 6(19.5) + 2(28.0) = 248$$

$$E(R_1) = 40(41)/4 = 410$$

$$\text{Var}(R_1) = 40(41)(81)/24 - [(14^3-14) + (10^3-10) + (7^3-7) + (1^3-1) + (2^3-2) + (2^3-2) + (3^3-3) + (1^3-1)]/48$$

$$= 5535 - 4092/48 = 5535 - 85.25 = 5449.75$$

Alternativt:

$$\text{Var}(R_1) = [14(75)^2 + 10(19.5)^2 + 7(28.8)^2 + 32^2 + 2(33.5)^2 + 2(35.5)^2 + 3(38.0)^2 + 40.0^2]/4 = 5449.75$$

$$T = \frac{248 - 410}{\sqrt{5449.27}} = -2.19$$

NTNU
Det skapende universitet

www.ntnu.no

20

Oppsummering eks 9.8 og 9.12:

Tegntesten: $p=0.636$

Wilcoxon-testen: $p=0.029$

Wilcoxon er sterkere enn tegntesten, men krever at data er symmetrisk fordelt under nullhypotesen.

Merk:
Rosner bruker kontinuitetskorreksjonen $-1/2$ i tellerne.
Er omdiskutert.

NTNU
Det skapende universitet

www.ntnu.no

21

9.4. Wilcoxon's toutvalgstest (Wilcoxon's Rank-Sum test) = Wilcoxon-Mann-Whitney's test

NTNU
Det skapende universitet

www.ntnu.no

22

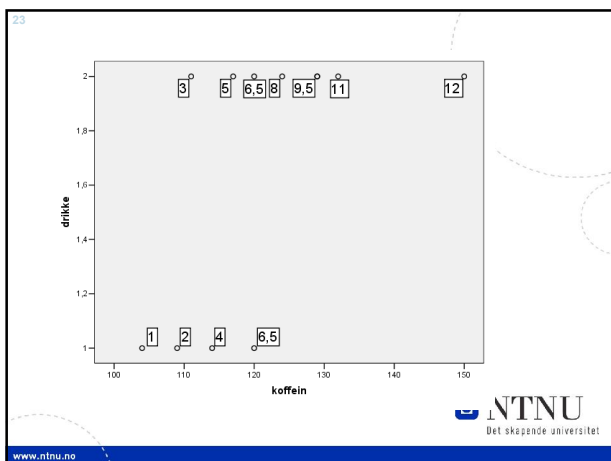
Fra Aalen et al 2006. Prøver fra 4 flasker Coca-Cola og 8 flasker Tab. Koffeininnhold i mg/l

Tabell 8.4 Tilordning av rangtall til observerte verdier. Rangtallene i Coca-Cola-gruppen er satt i kursiv.

Coca-Cola mg/l	104	109		114		120			
Tab mg/l			111	117	120	124	129/129	132	150
Rangtall	<i>1</i>	<i>2</i>	3	4	5	6.5	8	9.5	11

NTNU
Det skapende universitet

www.ntnu.no



24

Wilcoxon's toutvalgstest

Hypotese:

Rosner:
 H_0 Median₁ = Median₂
 H_1 Median₁ ? Median₂

Mer generelt:
 H_0 De to fordelingene er like
 H_1 Observasjonene fra den ene fordelingen tenderer til å være større

NTNU
Det skapende universitet

www.ntnu.no

25

To grupper med n_1 og n_2 observasjoner

Ranger alle observasjonene fra minste til største

R_1 = rangsum i Gruppe 1.

Under H_0 :

$$E(R_1) = \frac{n_1(n_1 + n_2 + 1)}{2}$$

www.ntnu.no

26

Hvis ingen sammenfallende observasjoner:

$$SD(R_1) = \sqrt{\left(\frac{n_1 n_2}{2}\right)(n_1 + n_2 + 1)}$$

Hvis sammenfallende observasjoner

$$SD(R_1) = \sqrt{\left(\frac{n_1 n_2}{2}\right) \left[n_1 + n_2 + 1 - \frac{\sum_{i=1}^g t_i(t_i^2 - 1)}{(n_1 + n_2)(n_1 + n_2 - 1)} \right]}$$

www.ntnu.no

27

Eksempel koffein i Coca-cola versus Tab;

$$R_1 = 1+2+4+6.5 = 13.5$$

$$E(R_1) = \frac{4(4+8+1)}{2} = 26$$

$$SD(R_1) = \sqrt{\left(\frac{4 \cdot 8}{2}\right)(4+8-1)} = 5.87$$

$$T = \frac{13.5 - 26}{5.87} = -2.13$$

Tosidig p-verdi=0.033

www.ntnu.no

28

Rosner Table 9.3

Comparison of visual acuity in people age 10–19 with dominant and sex-linked retinitis pigmentosa

Visual acuity	Dominant	Sex-linked	Combined sample	Range of ranks	Average rank
20–20	5	1	6	1–6	3.5
20–25	9	5	14	7–20	13.5
20–30	6	4	10	21–30	25.5
20–40	3	4	7	31–37	34.0
20–50	2	8	10	38–47	42.5
20–60	0	5	5	48–52	50.0
20–70	0	2	2	53–54	53.5
20–80	0	1	1	55	55.0
	25	30	55		

www.ntnu.no

29

Kruskall-Wallis test

www.ntnu.no

30

Sammenlikning mellom 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 multiplert lineær regresjon med indikatorvariable for gruppene. Kan også brukes ved justering for kovariater (f.eks alder, kjønn).

30

31

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	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

Note: There are 6x4 = 24 rabbits!

NTNU
Det skapende universitet

www.ntnu.no

31

32

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

NTNU
Det skapende universitet

www.ntnu.no

32

33

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
1	+2	13.5	+1	9.0
2	+3	20.0	+3	20.0
3	+3	20.0	+1	9.0
4	+3	20.0	+2	13.5
5	+3	20.0	+2	13.5
6	0	4.0	+3	20.0

NTNU
Det skapende universitet

www.ntnu.no

33

34

Kruskal-Wallis' test:

k grupper, gruppe nr i har n_i observasjoner og rangsum R_i . $N = \sum_{i=1}^k n_i$.

$$H^* = \frac{12}{N(N+1)} \sum_{i=1}^k n_i \left[R_i / n_i - \frac{(N+1)/2}{E(R_i / n_i) \text{ under } H_0} \right]^2 = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

Lettere å regne ut

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

hvor t_j er antall sammenfallende observasjoner i klynge nr j

er tilnærmet χ^2_{k-1} under H_0 .

NTNU
Det skapende universitet

www.ntnu.no

34

35

SPSS:

Nonparametric tests -> k independent samples

		Rank	
		DRUG	Mean Rank
LIDSCORE	Indomethilin	6	16,25
	Aspirin	6	14,17
	Piroxicam	6	15,25
	BW755C	6	4,33
	Total	24	

Test Statistics^{a,b}

	LIDSCORE
Chi-Square	11,804
df	3
Asymp. Sig.	,008
Exact Sig.	,003
Point Probability	,000

a. Kruskal Wallis Test

b. Grouping Variable: DRUG

NTNU

Det skapende universitet

www.ntnu.no

35

36

12.7. Multiple sammenlikninger (Dunn prosedyren)

Regn ut

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

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

Merk at dette tilsvare Bonferroni korreksjon

NTNU
Det skapende universitet

www.ntnu.no

36

37

Eks. (Forts)

Ranks

DRUG	N	Mean Rank
LIDSCORE	6	16.25
Indomethacin	6	14.17
Aspirin	6	15.25
Piroxicam	6	4.33
BW755C	6	
Total	24	

$z_{12} = 0.51, z_{13} = 0.24, z_{14} = 2.92, z_{23} = -0.27, z_{24} = 2.41, z_{34} = 2.67$

$\alpha^* = \frac{0.05}{4(4-1)} = 0.0042, z_{1-0.0042} = 2.64$

Altså:
Gruppe 1 og 4 er forskjellige
Gruppe 3 og 4 er forskjellige

NTNU
Det skapende universitet

www.ntnu.no

38

Friedman's to-veis ANOVA (Altman, 12.3.5)

- Krever ikke normalfordeling
- n subjekter og k grupper. Én observasjon per celle (subjekt og gruppe). Få eller ingen sammenfallende observasjoner.
- $H_0 (H_1)$: Det er ikke (er) forskjell på gruppene

NTNU
Det skapende universitet

www.ntnu.no

39

Table 12.9 Immersion suit leakage (g) during simulated helicopter underwater escape (Light *et al.*, 1987)
From Altman (1991)

Subject	Suit type			
	A	B	C	D
1	308	132	454	64
2	102	526	0	28
3	182	134	96	30
4	268	324	264	90
5	166	228	134	34
6	332	296	458	6
7	198	350	200	90
8	28	274	16	24
Mean	198	283	302	45.7
SD	103	127	179	31.6

NTNU
Det skapende universitet

www.ntnu.no

40

Table 12.10. Ranks of the data in Table 12.9.
From Altman (1991)

Subject	Suit type			
	A	B	C	D
1	3	2	4	1
2	3	4	1	2
3	4	3	2	1
4	3	4	2	1
5	3	4	2	1
6	3	2	4	1
7	2	4	3	1
8	3	4	1	2
Total (R)	24	27	19	10
Mean rank	3.00	3.38	2.38	1.25

NTNU
Det skapende universitet

www.ntnu.no

41

Friedman's test:

$$H = \frac{12}{nk(k+1)} \sum_{i=1}^k [R_i - n(k+1)/2]^2 = \frac{12}{nk(k+1)} \underbrace{\sum_{i=1}^k R_i^2 - 3n(k+1)}_{\text{Lettere å regne ut}}$$

er tilnærmet χ^2_{k-1} under H_0 .
(R_i er rangsum i gruppe i)

Eksempel:

$$H = \frac{12}{8 \times 4 \times 5} [24^2 + 27^2 + 19^2 + 10^2] - 3 \times 8 \times 5 = 12.45$$

p-verdi $\approx P(\chi^2_{4-1} \geq 12.45) = 0.006$

NTNU
Det skapende universitet

www.ntnu.no

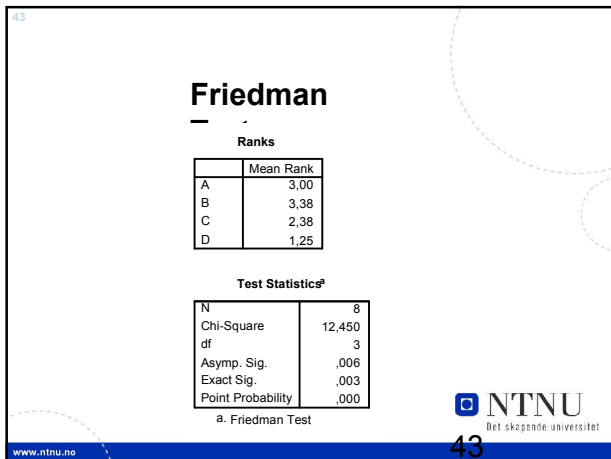
42

Friedman's test i SPSS:

- Et case pr subjekt, en variabel pr gruppe
- Analyse -> Nonparametric tests -> k related samples
- Opsjonen "Exact" gir p-verdien eksakt ("Exact Sig.") i tillegg til kji-kvadratfordelignens tilnærming ("Asymp. Sig")

NTNU
Det skapende universitet

www.ntnu.no



44

Hvilke grupper er forskjellige?

- Friedman's test forteller om minst to grupper er forskjellige. (I eksempelet har D åpenbart lavere verdier)
- Par av grupper kan sammenliknes vha Wilcoxons test for matchede par. Juster for multiple sammenlikninger.
- Friedman's test for 2 grupper tilsvarer tegn-testen!

NTNU
Det skapende universitet

www.ntnu.no

44

45

Rang-korrelasjon, Spearman's r_s

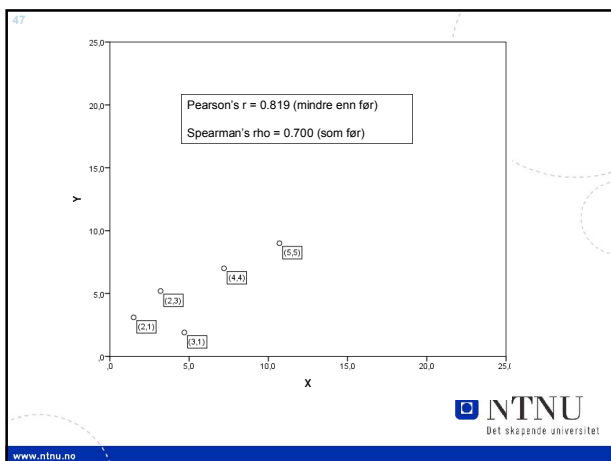
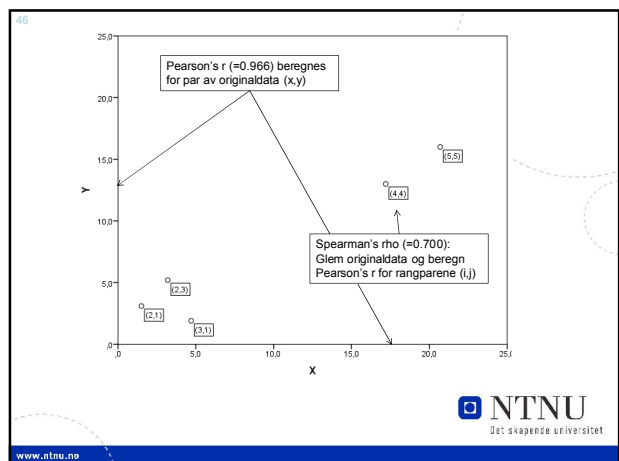
- Observasjonene ordnes i stigende rekkefølge
- Regn ut Pearson's korrelasjonskoeffisient for rangene (i)

Transformasjon:

$$x_i \xrightarrow{\text{ordnes}} x_{(i)}$$

NTNU
Det skapende universitet

www.ntnu.no



48

Ikke-paramterisk korrelasjonskoeffisient

- Kan være å foretrekke når:
 - Sammenhengen er ikke-lineær
 - Data er ordinale (kvalitative)
 - Ved avvik fra normalfordelingen(?)
- Alternativer:
 - Spearman's ρ
 - Kendall's tau
 - Liten forskjell på dem, men kanskje en viss preferanse for Kendall's tau.

NTNU
Det skapende universitet

www.ntnu.no

49

"Our results suggest that Kendall's tau, has many advantages over Pearson's and Spearman's r; when applied to psychiatric data, tau, maintained adequate control of type I errors, was nearly as powerful as Pearson's r, provided much tighter confidence intervals and had a clear interpretation."

Arndt S, Turvey C, Andreasen NC: Correlating and predicting psychiatric symptom ratings: Spearman's r versus Kendall's tau correlation. JOURNAL OF PSYCHIATRIC RESEARCH Volume: 33 Issue: 2 Pages: 97-104, 1999

NTNU
Det skapende universitet

www.ntnu.no

50

Denne, derimot, skriver at begge er ganske bra:

Kraemer HC: Correlation coefficients in medical research: from product moment correlation to the odds ratio. STATISTICAL METHODS IN MEDICAL RESEARCH Volume: 15 Issue: 6 Pages: 525-545, 2006

NTNU
Det skapende universitet

www.ntnu.no

51

Ikke-parametriske tester

- Basert på rekkefølgen (rangordningen) av data, ikke de faktiske verdiene
- Få forutsetninger om fordelingen(e). Wilcoxon's test for pardata (ettutvalgstest) forutsetter at d-ene er symmetrisk fordelt under H_0 .
- Tåler ekstreme verdier / avvik fra normalfordelingen
- Hvis data virkelig er normalfordelt: Ved "store" utvalg er Wilcoxon-testene nesten like sterke som t-testene. Ved små utvalg: Betydelig svakere enn parametriske tester.
- Ulempe: Gir bare p-verdi, ikke estimat og konfidensintervall for effekt!

NTNU
Det skapende universitet

www.ntnu.no