

En introduktion till icke-parametrisk statistik

Anna Lindam

Statistiker, Region Jämtland Härjedalen

anna.lindam@regionjh.se

Översikt

- Vad är en icke-parametrisk metod?
- Vid vilken typ av data används det?
- Hur vet man att det är ett icke-parametrisk test vi bör använda?
- Medianer och Inter Quartile Range.
- Jämförelse mellan grupper
- Korrelationer

Vad är icke-parametrisk statistik?

Metoder där variabeln vi studerar inte antas följa någon specifik fördelning i vår population.

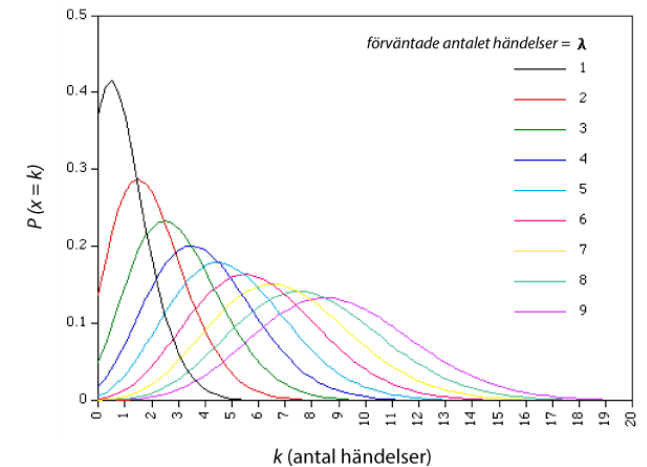
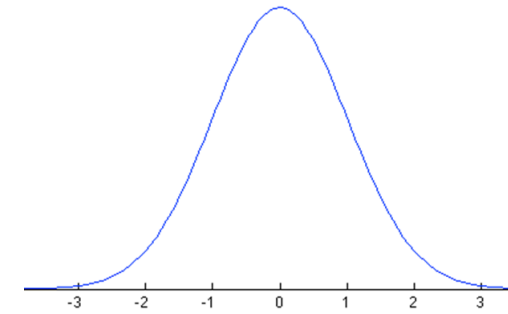


Vad är en parametrisk metod?

Metoder där man gör antaganden om variabelns fördelning i vår population.

Exempelvis t-test, ANOVA, linjär regressionsanalys, logistisk regression och poissonregression.

Normalfördelning



Poissonfördelning

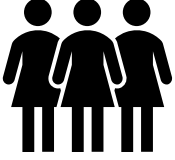
Är det någon skillnad i längd mellan personer i Västerbotten och Norrbotten?

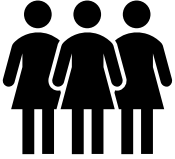
Jämföra medelvärdena för länen i längd i cm

Studiepopulation: n= 156 personer, Västerbotten n=82; Norrbotten n=74

Jämföra 2 oberoende grupper

Hypotes:

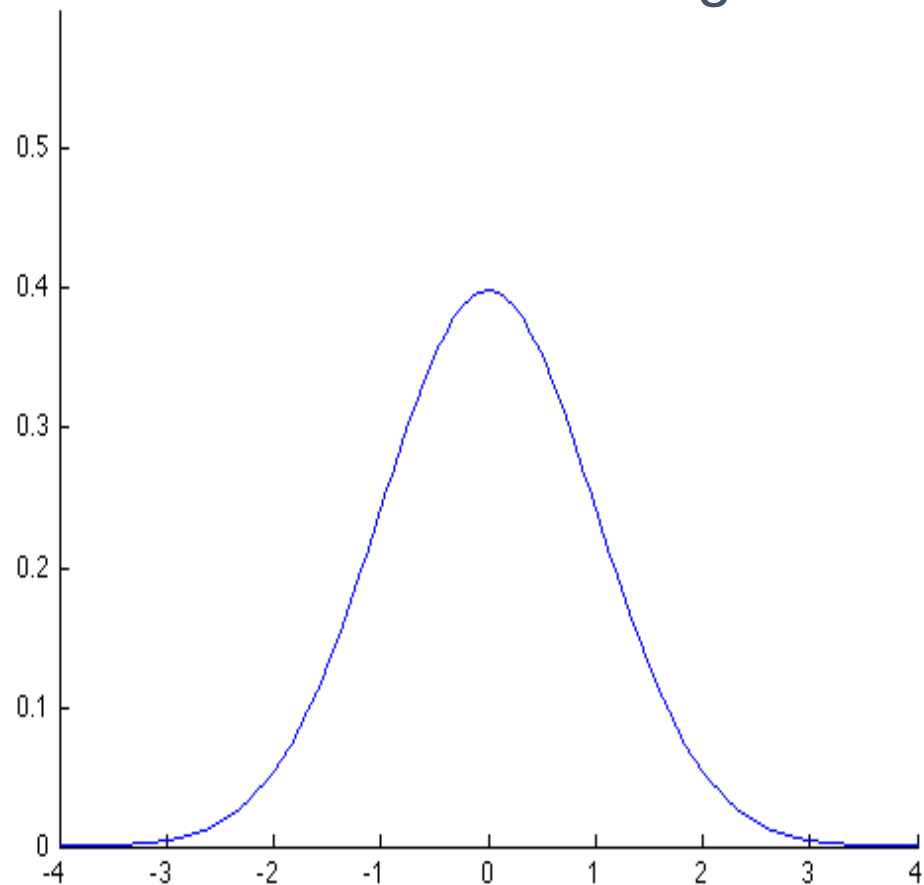
H0:  =  $H_0: \mu_{VB} = \mu_{NB}$

H1:  $\neq$  $H_1: \mu_{VB} \neq \mu_{NB}$

Metod: T-test

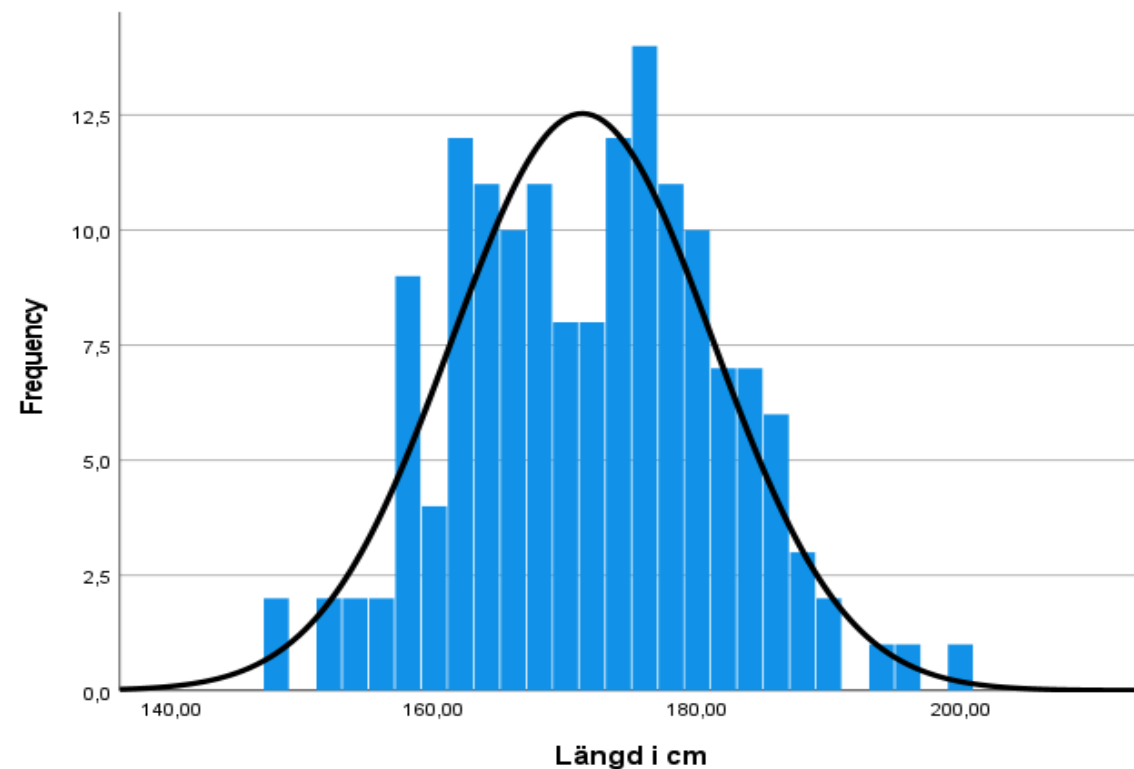
Antagande: längd är normalfördelat i bakgrundspopulationen

Standardiserad normalfördelning



Medelvärde = 0 Standardavvikelse = 1

Vårt stickprov n= 156



Test om vår variabel är normalfördelad

H0: Längd kan antas följa normalfördelningen

H1: Längd skiljer sig från normalfördelningen

Signifikansgräns $P\text{-värde} < 0.05$

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Längd i cm	,057	156	,200 [*]	,993	156	,639

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Är det någon skillnad i längd mellan personer i Västerbotten och Norrbotten?

Resultat för T-test: P-värde = 0.815

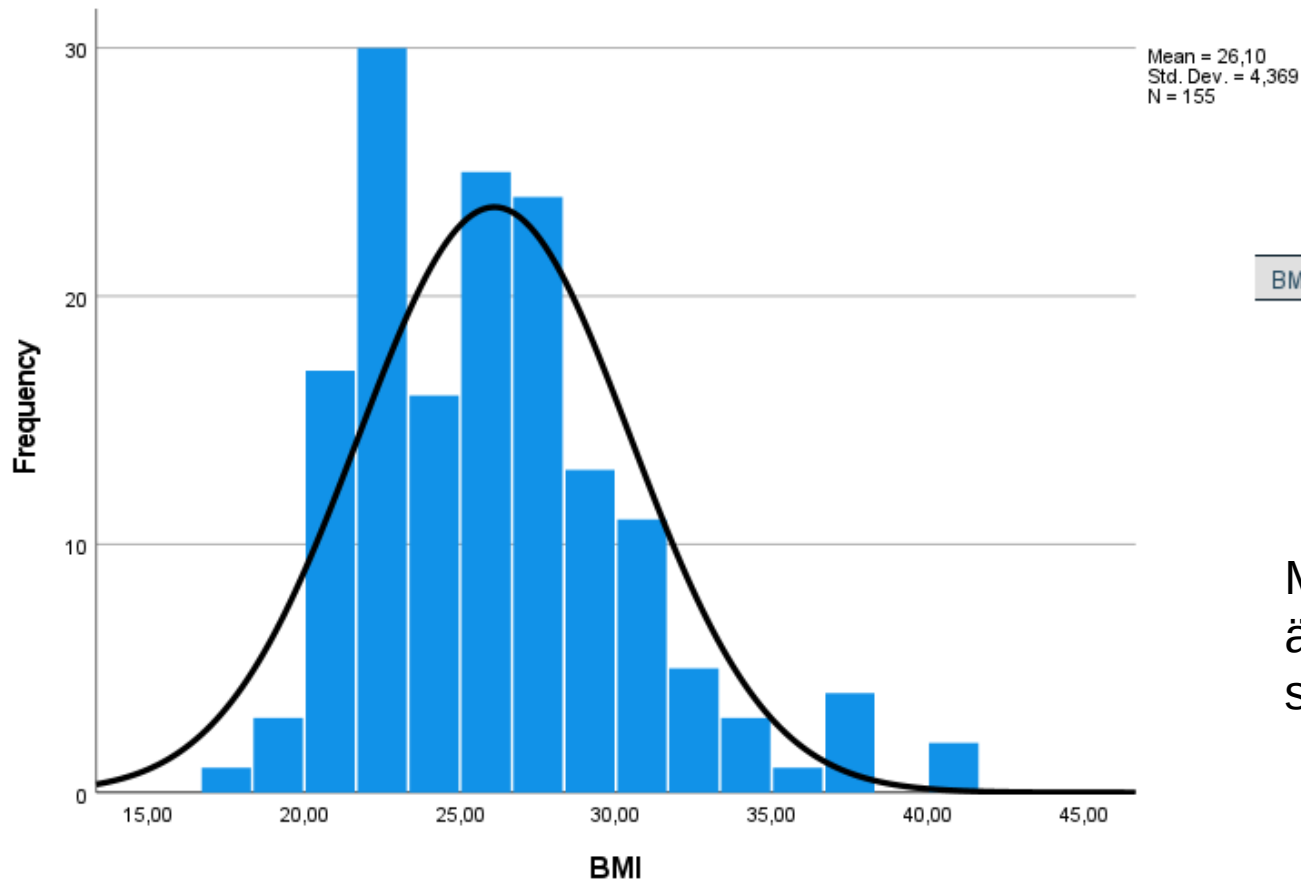
Signifikansgräns $\alpha = 0.05$

Tolkning: Nej, vi kunde inte se någon signifikant skillnad i längd mellan länen.

Finns det någon skillnad i BMI mellan personer i Västerbotten och Norrbotten?

BMI = Body Mass Index

Finns det någon skillnad i BMI mellan personer i Västerbotten och Norrbotten?



Tests of Normality							
Kolmogorov-Smirnov ^a			Shapiro-Wilk				
	Statistic	df	Sig.		Statistic	df	Sig.
BMI	,071	155	,052		,949	155	<,001

a. Lilliefors Significance Correction

Kan vi använda ett t-test test här?

Meningarna går isär, men jag rekommenderar ändå ett icke-parametrisk test eftersom data är snedfördelat.

Kan vi använda ett t-test här?

Meningarna går isär

Fördel T-test

- Robust

Fördel icke parametrisk

När använder vi icke-parametriska metoder?

När data inte är normalfördelat

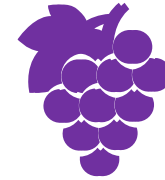
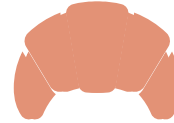
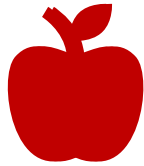
När vi har små stickprov

När vi har data på ordinalskalenivå



Datatyper

Nominalskala



Ordinalskala 1=mycket bra 2= bra 3=ok 4= dåligt 5=mycket dåligt

Intervallskala Numerisk innebörd, kan innehålla minusvärden



Kvotskala Har en absolut nollpunkt t.ex. BMI, blodtryck

Icke-parametriska metoder

Metoderna använder inte observationernas värden utan endast observationernas rangordning.

Många icke-parametriska test utnyttjar dock en parametrisk fördelning.

Skillnad i sättet *hur* testet utnyttjar informationen i variabeln.

Kan vi använda ett parametriskt test så är det bättre på att upptäcka skillnader.

Deskriptiv statistik

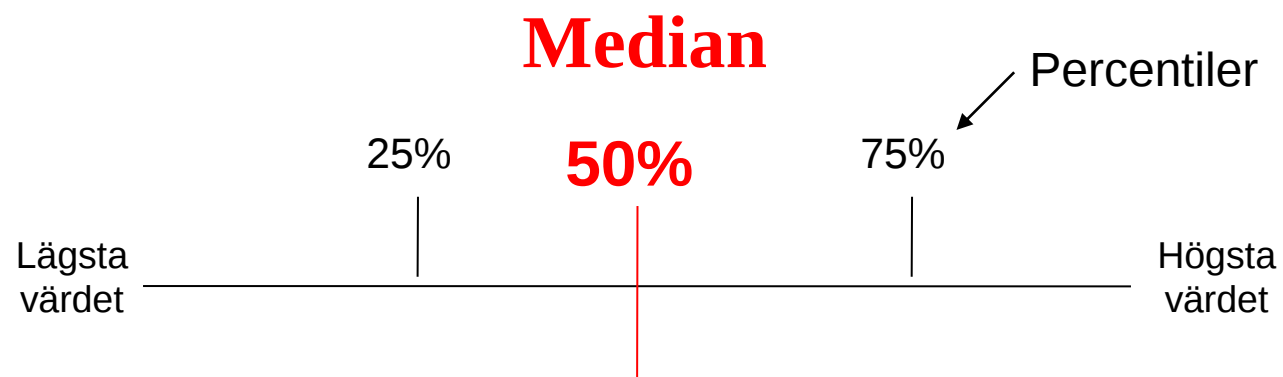
Icke-parametrisk

- Median
- Inter Quartile Range (IQR)
- Range (högsta och minsta värdet)

Parametrisk

- Medelvärde
- Standardavvikelse
- Standard Error

Median



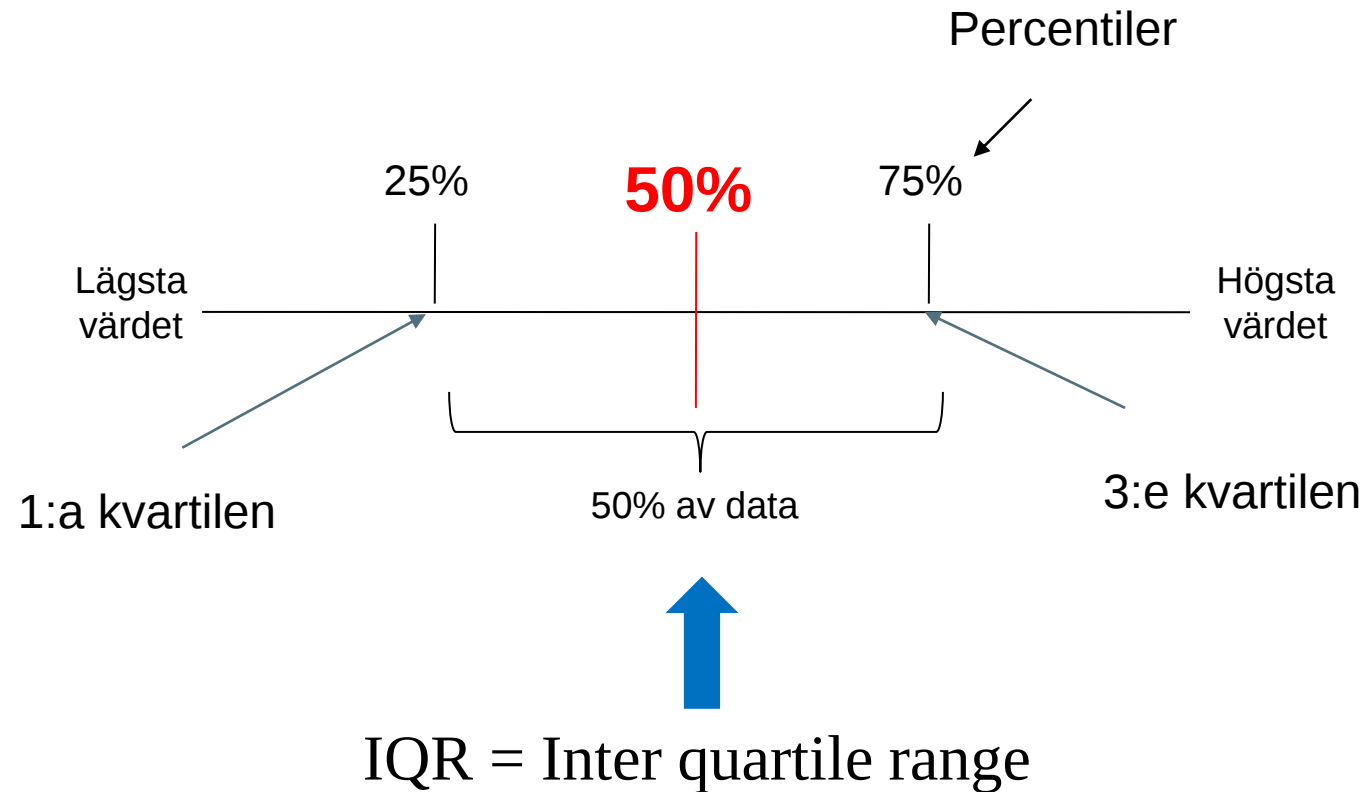
Observationerna sorteras i stigande ordning och observationen i mitten är medianen.



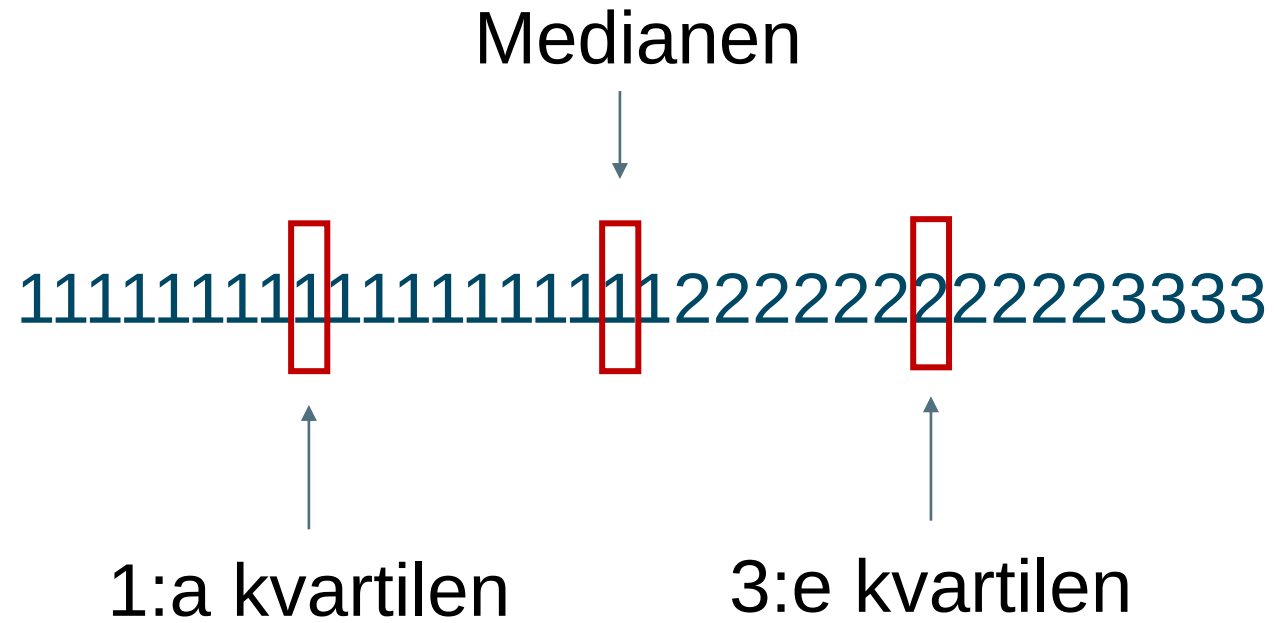
Inter Quartile Range

Skillnaden mellan 3:e och 1:a kvartilen

I artiklar ofta median (1:a kvartilen- 3:e kvartilen)



Deskriptiv statistik vid Icke-parametrisk metod



Deskriptiv statistik

Median: 25,6

IQR: 5,8 (Q1:22,5 - Q3 28.4)

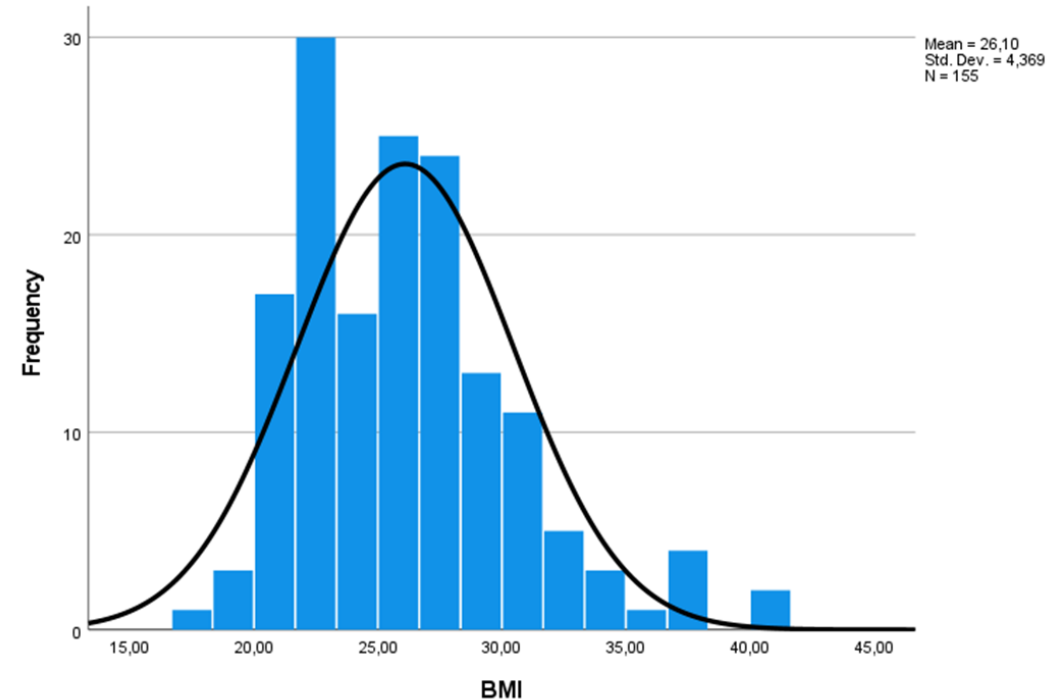
Medelvärde: 26,1

Standardavvikelse (SD): 4,4

68% inom ± 1 SD

95% inom ± 2 SD

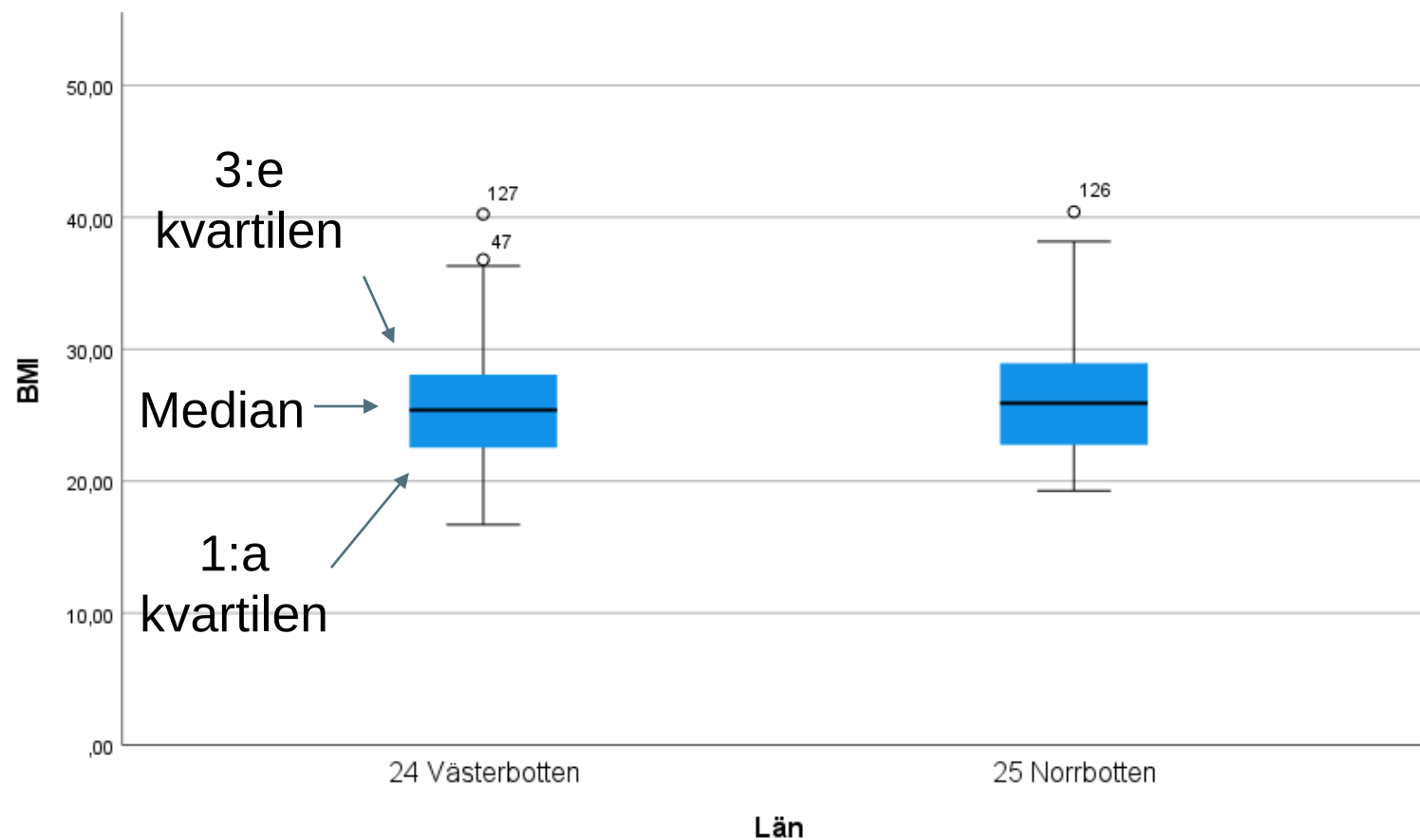
99.7% inom ± 3 SD



50% av data

IQR

Finns det någon skillnad i BMI mellan personer i Västerbotten och Norrbotten?



Mann-Whitney U (Wilcoxon's rank sum test)

Jämför om två populationers fördelning är lika.

Hypotes:

H0: BMI har samma fördelning i länen

H1: BMI har ej samma fördelning i länen

Kan användas:

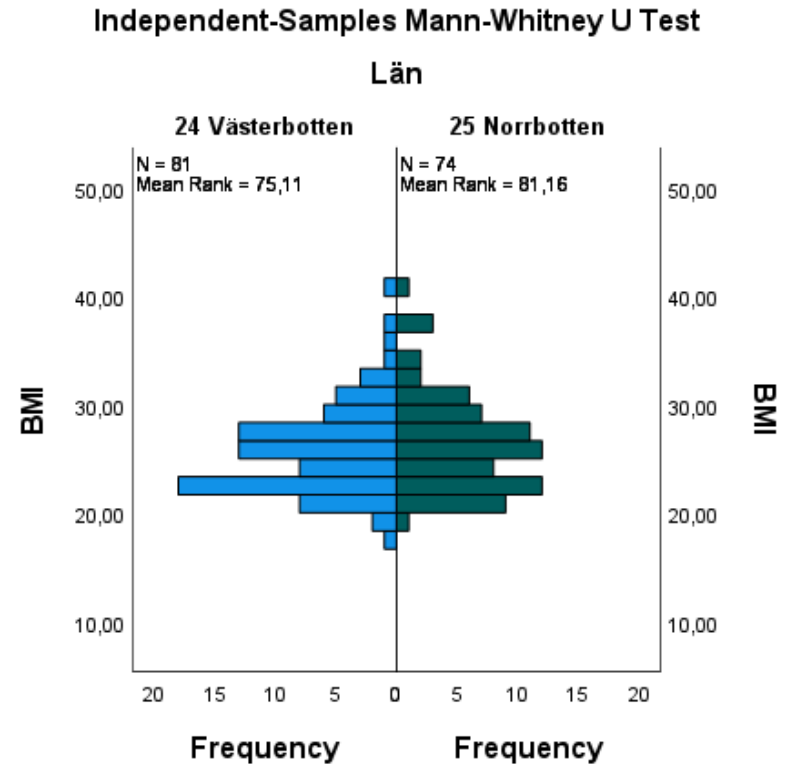
Vid oberoende grupper

Minst ordinalskalenivå

Mann-Whitney U-test

1. Alla observationer i grupperna rangordnas gemensamt.
2. Rangerna summeras för respektive grupp.
3. Rangsumma och medelrang beräknas för respektive grupp.
4. Test-variabel bildas.
 1. Summan av det ena stickprovets ranger jämförs med ett förväntat värde.

$$U = W_1 - \left(\frac{n_1(n_1+1)}{2} \right)$$
 där W_1 = summan av Västerbottens ranger och n_1 = antal personer i Västerbotten.



Mann- Whitney U test

- Sätter signifikansgräns (0,05)
- Signifikansgränser enligt tabell

Resultat: Mann - Whitney U test

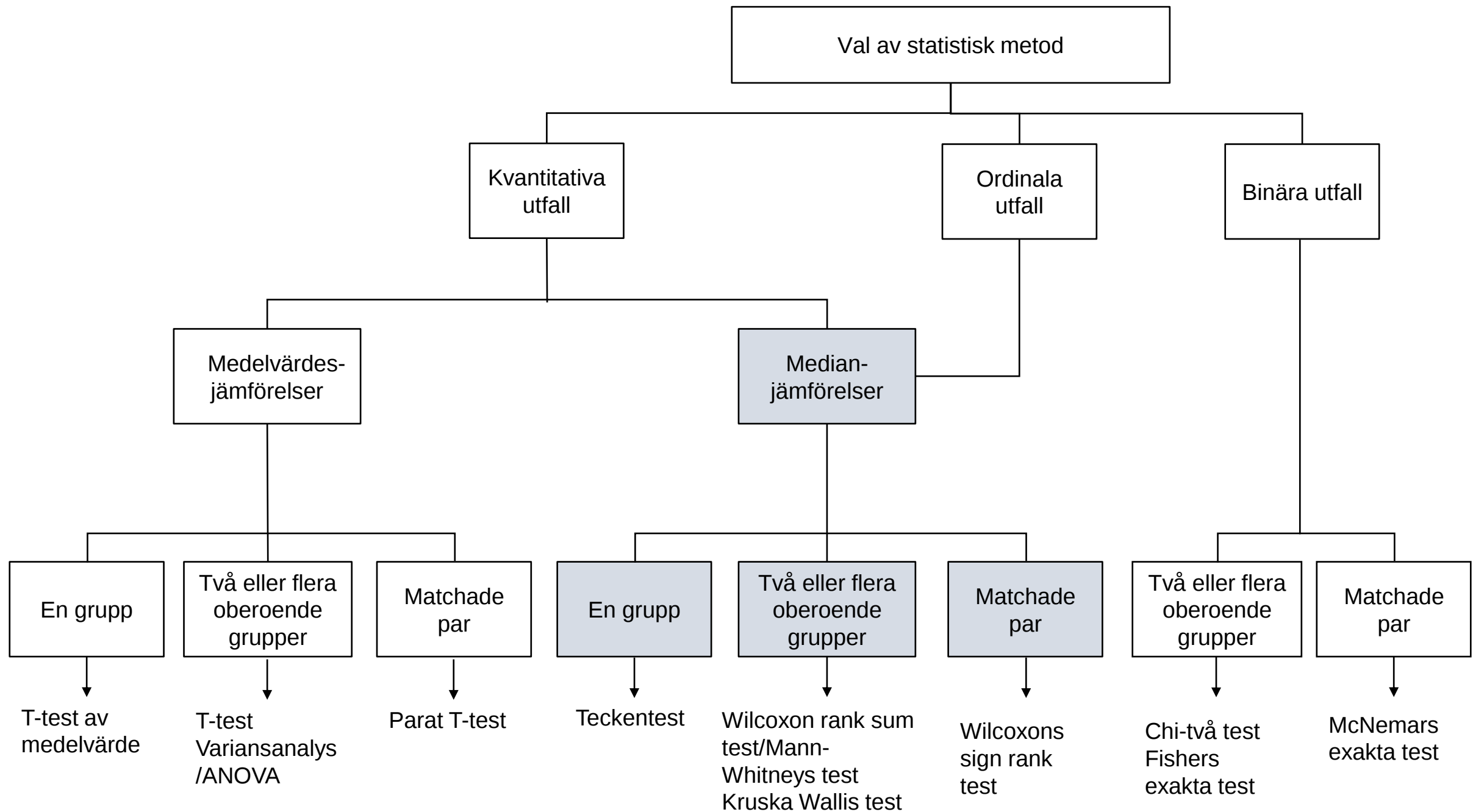
Hypothesis Test Summary				
	Null Hypothesis	Test	Sig. ^{a,b}	Decision
1	The distribution of BMI is the same across categories of Län.	Independent-Samples Mann-Whitney U Test	,402	Retain the null hypothesis.

a. The significance level is ,050.

b. Asymptotic significance is displayed.

Slutsats: Vi hittade ingen signifikant skillnad för fördelning av BMI i Västerbotten och Norrbotten.

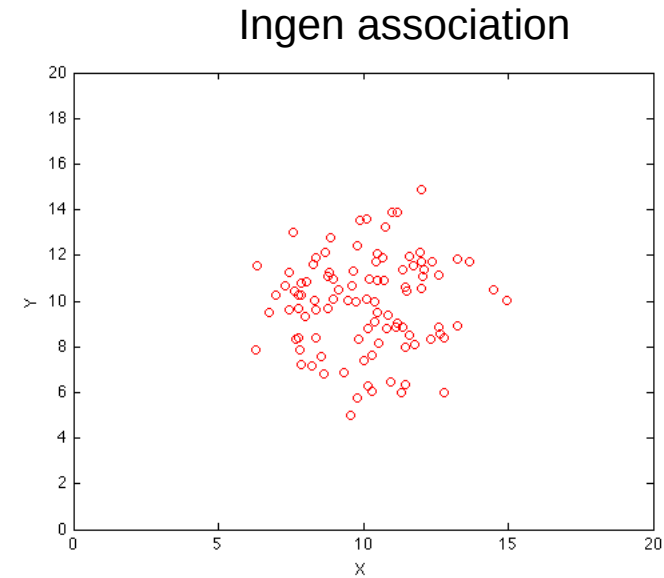
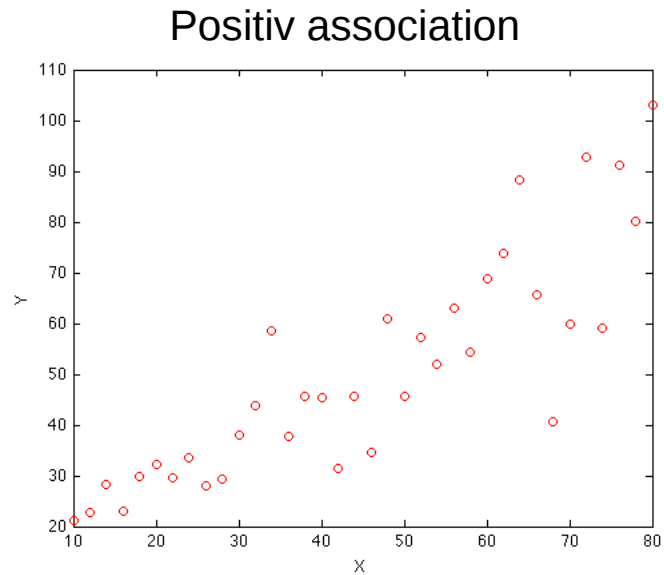
Fler test då vi jämför 2 eller flera grupper



Vilka test när?

	Parametriska test	Icke-parametriska test
Vid två oberoende grupper	T-test	Mann Whitney U / Wilcoxons rank sum test
Vid 3 eller fler oberoende grupper	ANOVA	Kruska Vallis test
Matchade par	Parat T-test	Wilcoxons sign rank test

Korrelation - association mellan två variabler



Pearson korrelation

Mäter ett linjärt
samband

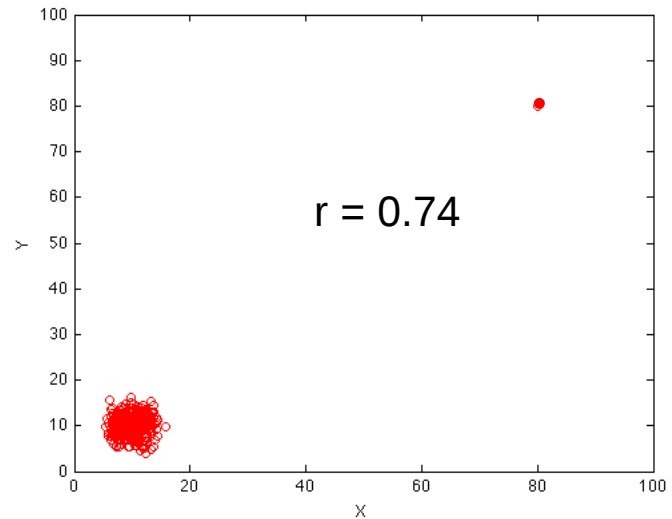
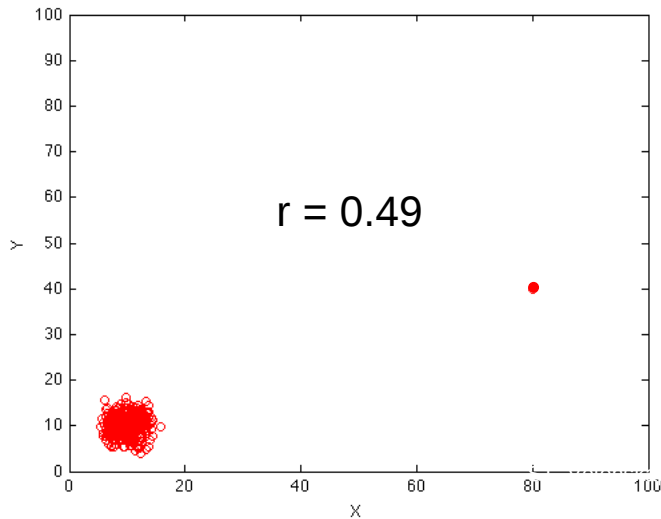
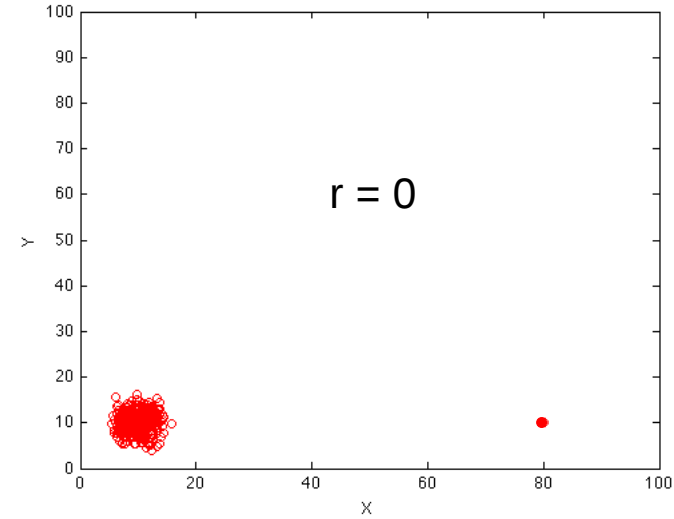
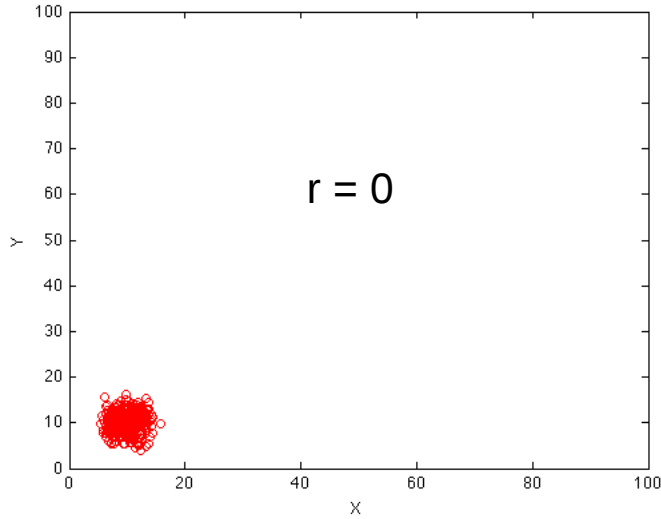
Används vid:

- Normalfördelade variabler
- Kvotskala

Känslig för outliers!!!



Korrelationskoefficienten, r , är extremt känslig för outliers



Spearman's rang – korrelation

Mäter förhållandet
mellan två variabler

Används vid:

- Vid ej normalfördelade data
- Ordinalskala
- Små dataset

Icke-parametrisk motsvarighet till Pearson

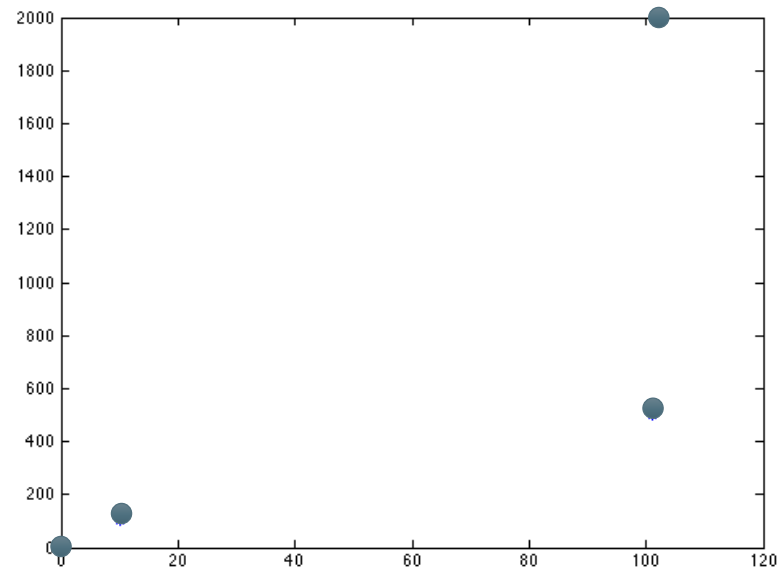
Krävs 3 kategorier eller fler.....



Spearman – exempel

Antag följande data

X	Y
0	1
10	100
101	500
102	2000



Vi har en perfekt rangkorrelation, vilket ger Spearman = 1

Pearson ger däremot korrelationskoefficienten $r = 0.7544$.





Kliniska Studier Sverige Forum Norr

- så förbättrar vi möjligheterna att bedriva kliniska studier i Norrland





Kliniska Studier Sverige – Forum Norr

- Plattform för forskningsstödjande infrastruktur och en bro mellan hälso- och sjukvård, näringsliv och akademi i Norra sjukvårdsregionen
- Samarbete mellan:
 - Region Jämtland Härjedalen
 - Region Västernorrland
 - Region Västerbotten
 - Region Norrbotten
 - Umeå universitet
- Regional nod för Norra sjukvårdsregionen

Kliniska forskningscentrum

Forum Norr vill sänka trösklar för klinisk forskning i Norra sjukvårdsregionen

Lokaler i **Sunderbyn, Sundsvall, Umeå** och **Östersund**

Samverkan som möjliggör inkludering av studiedeltagare från en större del av norra sjukvårdsregionen



Foto: Lisette Marjavaara



Foto: Sara Rönnerberg



Foto: Beatrice Backman Lönn



Foto: Erik Holmstedt

Kika gärna in på vår CANVAS-sida

<https://www.canvas.umu.se/courses/2600>

1/6 kl 12-13 Poweranalys/stickprovdimensionering Simon Wallin , Region Västerbotten / institutionen för folkhälsa och klinisk medicin, Umeå universitet

...Och så klart kommer vi att fortsätta till hösten. Schema skickas ut under våren.

Tack för mig!

Finns det frågor?