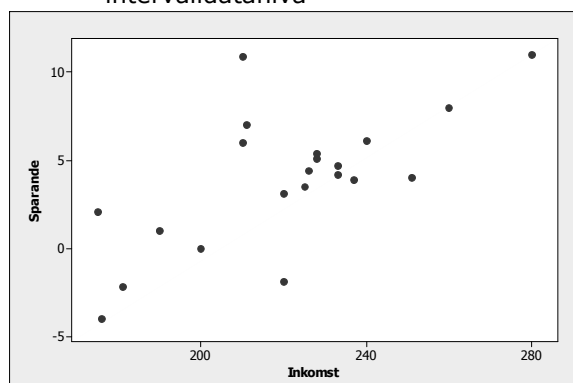


F14

Enkel linjär regression

Samband mellan X och Y

- Ett spridningsdiagram ger en överskådlig första bild av ett eventuellt samband
 - Variablerna måste vara mätta på minst intervalldatanivå

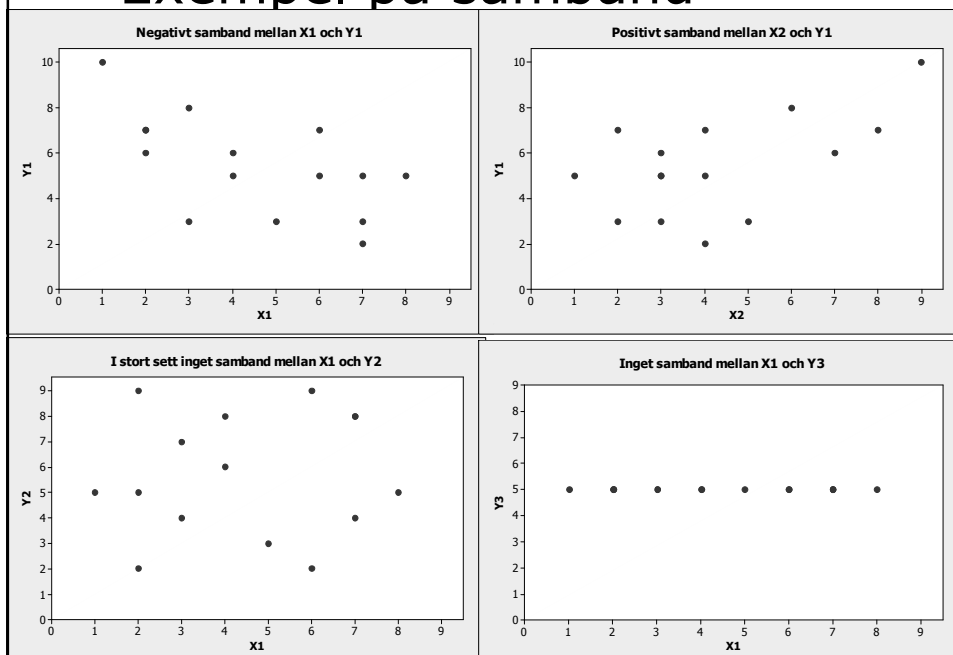


Inkomst	Sparande
175	2,1
176	-4,0
181	-2,2
190	1,0
200	0,0
210	6,0
210	10,9
211	7,0
220	3,1
220	-1,9
225	3,5
226	4,4
228	5,1
228	5,4
233	4,2
233	4,7
237	3,9
240	6,1
251	4,0
260	8,0
280	11,0

Samband mellan X och Y

- Samband mellan X och Y
 - X: Bakgrundsvariabel/oberoende variabel
 - Y: Undersökningsvariabel/beroende variabel
 - Om det finns ett samband (samvariation) mellan två variabler X och Y kan man i ett spridningsdiagram se hur Y i genomsnitt förändras när X ändras
 - Numeriskt samband – behöver ej vara kausalt
 - *Linjärt* samband
 - Den genomsnittliga förändringen i Y är lika stor för en given förändring i X oavsett vilket X-värde vi utgår från

Exempel på samband



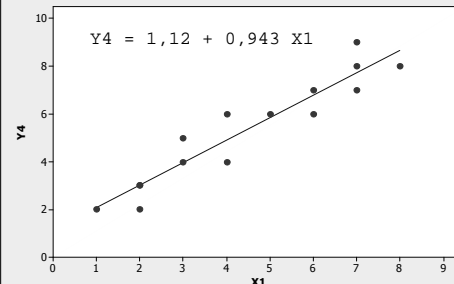
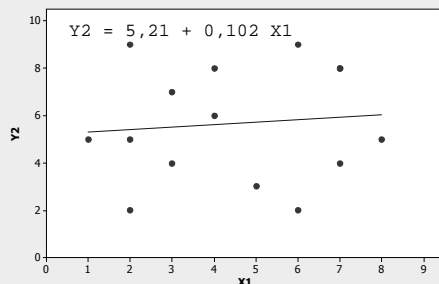
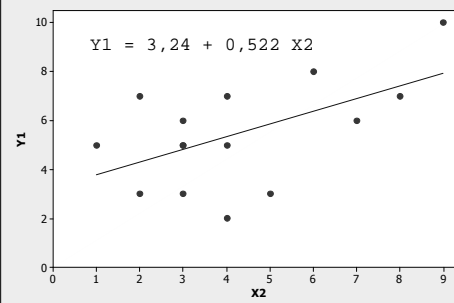
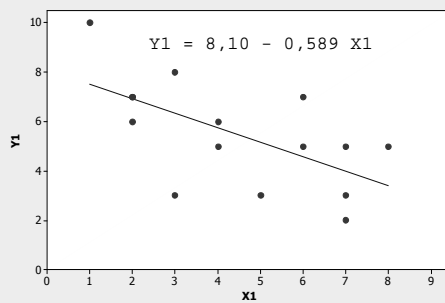
Linjärt samband: regressionslinje

- Om det verkar vara ett linjärt samband mellan X och Y kan vi sammanfatta sambandet med en rät linje
- Den linje som "bäst" sammanfattar en linjär punktsvärm av n talpar (x_i, y_i) , $i=1,2,\dots,n$ kallas *regressionslinje*:

$$\hat{Y} = a + bX$$

- a är linjens intercept: där linjen skär Y-axeln
 - Det förväntade värdet för Y om X är 0
- b är linjens lutning
 - Den genomsnittliga förändringen i Y när X ökar med en enhet
- X är oberoende variabeln
- \hat{Y} är det skattade värdet på Y (den beroende variabeln) enligt linjen
 - Ex. Vad skulle vi gissa på att Y är om vi vet att X antar ett specifikt värde?

Linjärt samband: regressionslinjer



Regression

- Bestämning av regressionskoefficienterna a och b
 - Linjen bestäms så att summan av de kvadrerade avvikelserna från linjen blir så små som möjligt
 - Minsta kvadratmetoden
 - Se föreläsningssanteckningar (KD) F14-F16

Regression Analysis: Sparande versus Inkomst

The regression equation is
Sparande = - 17,3 + 0,0961 Inkomst

Predictor	Coef	SE Coef	T	P
Constant	-17,280	5,492	-3,15	0,005
Inkomst	0,09607	0,02472	3,89	0,001

S = 2,96905 R-Sq = 44,3% R-Sq(adj) = 41,4%

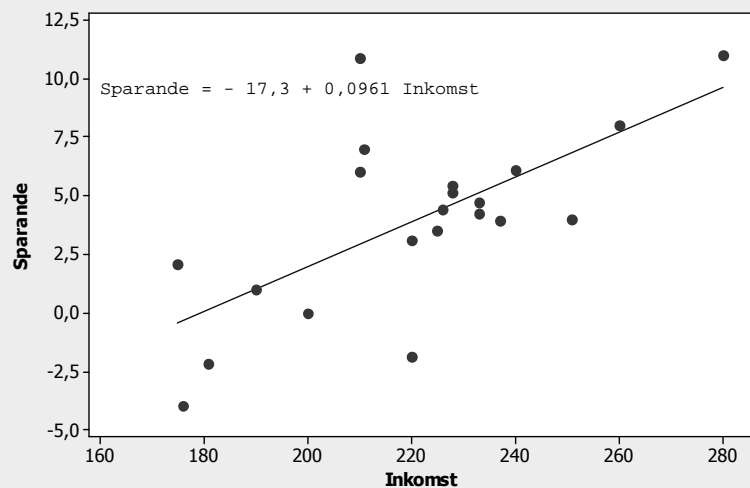
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	133,18	133,18	15,11	0,001
Residual Error	19	167,49	8,82		
Total	20	300,67			

Unusual Observations

Obs	Inkomst	Sparande	Fit	SE Fit	Residual	St Resid
7	210	10,900	2,894	0,699	8,006	2,77R
21	280	11,000	9,619	1,603	1,381	0,55 X

Regressionslinje



Regressionslinje

$$\text{Sparande} = - 17,3 + 0,0961 \text{ Inkomst}$$

■ $a = -17,3$

- Interceptet, linjens skärning av Y-axeln
- Det förväntade värdet på sparande för en person med en inkomst på 0 kronor: Om en person har en inkomst på 0 kronor förväntar vi oss att sparandet är -17 300 kronor (dvs negativt sparande)
- Obs! Är detta rimligt? Om värdet 0 ej finns i vårt undersökta material är tolkningen ej rimlig. Riskabelt att extrapolera utanför det undersökta området

■ $b = 0,096$

- Lutningen
- Om inkomsten ökar med 1000 kronor ökar i genomsnitt sparandet med 96 kronor
- Obs! Sambandet behöver ej vara kausalt!

Uppskatta en regressionslinje

- Antag följande talpar $(x_i; y_i)$: (1; 4)
(2; 6) (3; 5)
 1. Rita in punkterna i ett spridningsdiagram
 2. Rita in linjen med "ögonmått"
 3. Uppskatta intercept och lutning, dvs uppskatta linjens ekvation

Regression

- Kom ihåg:
 - Regressionslinjen sammanfattar endast det *linjära* sambandet – det kan finnas icke-linjära samband
 - Ett eventuellt linjärt samband behöver ej vara kausalt
 - Lutningskoefficienten är ej ett mått på *styrkan* på det linjära sambandet