



UPPSALA
UNIVERSITET

Föreläsning 8 och 9: Regressionsanalys

Pär Nyman

par.nyman@statsvet.uu.se

1 februari 2016



Regressionsanalys

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- Vårt viktigaste verktyg för kvantitativa studier.
- Kan användas till det mesta, men svarar oftast på frågor om kausala samband.
- På kursen lär vi ut de viktigaste grunderna.
- Viktigt även för dem som inte själva vill använda det!



Disposition för hela dagen

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- 1 Repetition och passningsmått
- 2 Statistisk signifikans
- 3 Att läsa regressionstabeller
- 4 Multipel regression
- 5 Kombinationsstudier
- 6 Avancerat



Repetition och passningsmått

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Regressionsekvationen (bivariat)

$$y = a + bx + e$$

y = Beroende variabel

a = Konstant eller intercept

b = Regressionskoefficient

x = Oberoende variabel

e = Felterm eller residual



Repetition och passningsmått

Regressions- analys

Regressionssekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Regressionssekvation med indexsiffror

$$y_i = a + bx_i + e_i$$

y = Beroende variabel

a = Konstant eller intercept

b = Regressionskoefficient

x = Oberoende variabel

e = Felterm eller residual

i = Indexsiffra från observation 1 till observation n



Repetition och passningsmått

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Regressionsekvation för förväntade värden

$$\hat{y} = a + bx$$

$\hat{y}$ = Förväntat värde på den beroende variabel

a = Konstant eller intercept

b = Regressionskoefficient

x = Oberoende variabel



Multipel regressionsekvation 1

$$y = a + b_1x + b_2z + e$$

y = Beroende variabel

a = Konstant eller intercept

b_1 = Regressionskoefficient 1

b_2 = Regressionskoefficient 2

x = Oberoende variabel 1

z = Oberoende variabel 2

e = Felterm eller residual



Repetition och passningsmått

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Multipel regressionsekvation 2

$$y = a + b_1x_1 + b_2x_2 + e$$

y = Beroende variabel

a = Konstant eller intercept

b_1 = Regressionskoefficient 1

b_2 = Regressionskoefficient 2

x_1 = Oberoende variabel 1

x_2 = Oberoende variabel 2

e = Felterm eller residual



Repetition och passningsmått

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Multipel regressionsekvation 3

$$\text{Inkomst} = a + b_1 \times \text{Utbildning} + b_2 \times \text{Kvinna} + e$$

Inkomst = Beroende variabel

a = Konstant eller intercept

*b*₁ = Regressionskoefficient 1

*b*₂ = Regressionskoefficient 2

Utbildning = Oberoende variabel 1

Kvinna = Oberoende variabel 2

e = Felterm eller residual



Repetition och passningsmått

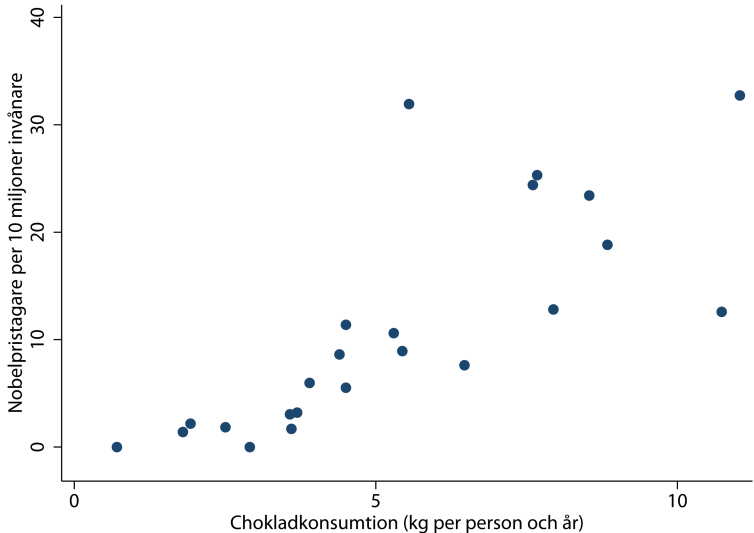
Exempel från New England Journal of Medicine: Vad förklarar variationen i länders vetenskapliga framgångar?

- Rimligtvis beror vetenskaplig framgång i någon mån på kognitiv förmåga.
- Kognitiv förmåga kan eventuellt förbättras genom kost rik på flavanoider.
- Det finns mycket flavanoider i choklad.
- Alltså bör andelen Nobelpristagare vara större i länder där man äter mycket choklad?



Repetition och passningsmått

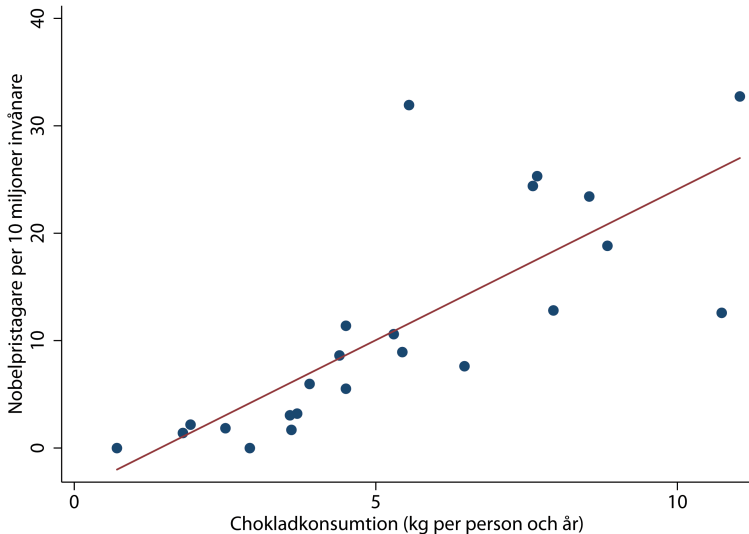
$$y = a + bx + e$$





Repetition och passningsmått

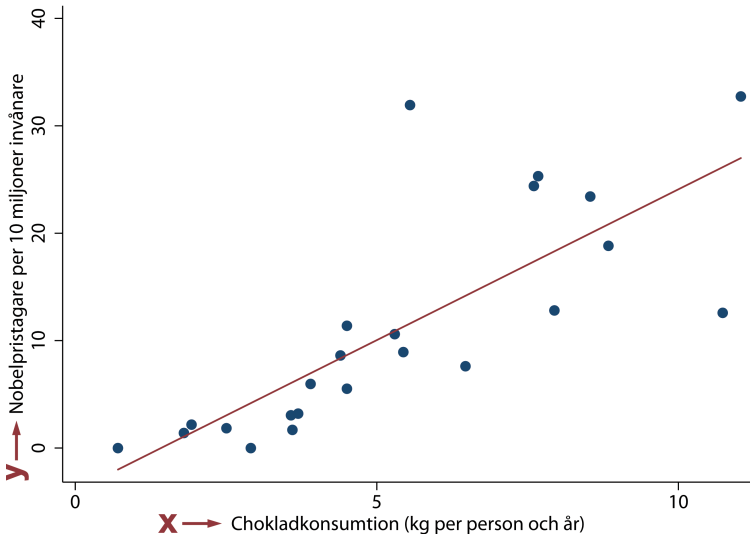
$$y = a + bx + e$$





Repetition och passningsmått

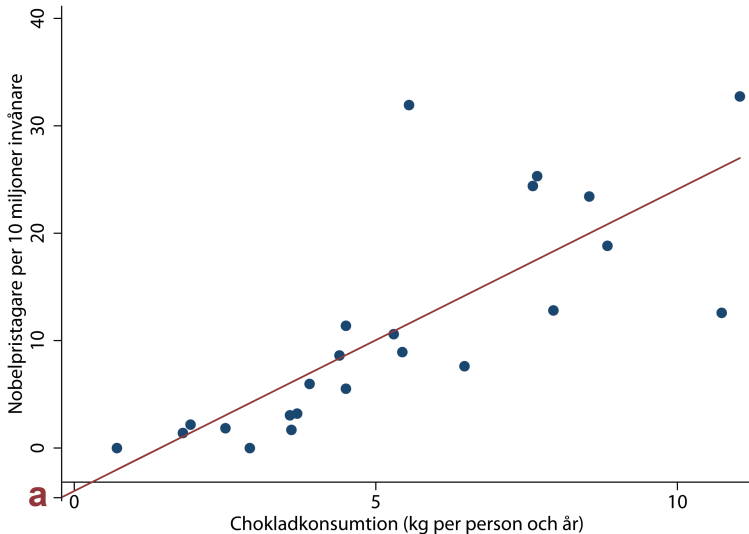
$$y = a + bx + e$$





Repetition och passningsmått

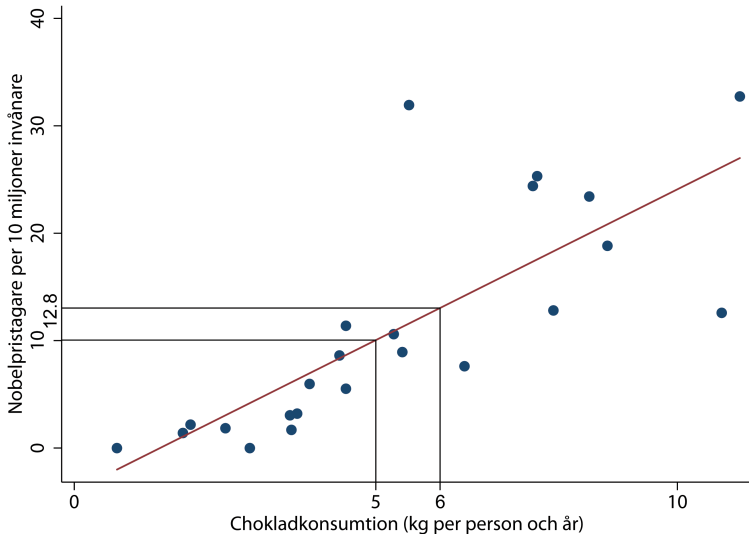
$$y = a + bx + e$$





Repetition och passningsmått

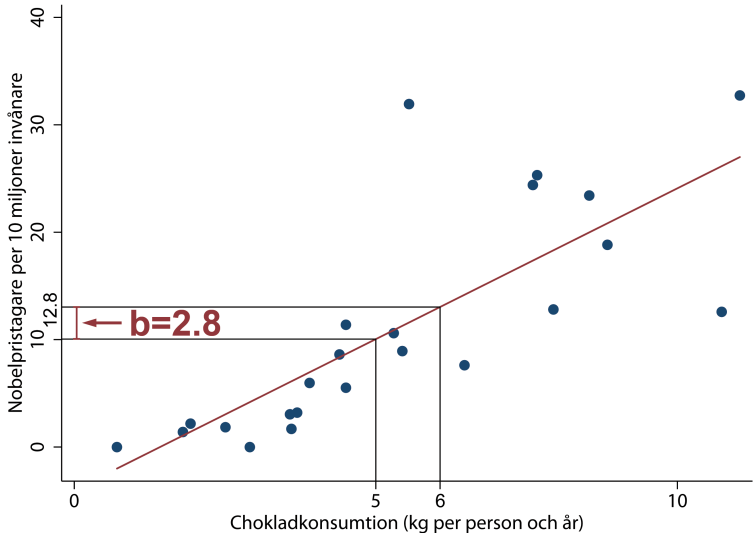
$$y = a + bx + e$$





Repetition och passningsmått

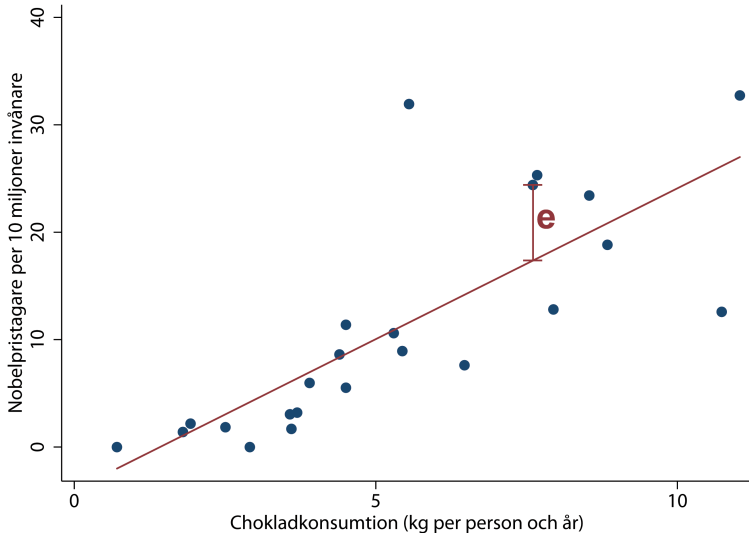
$$y = a + bx + e$$





Repetition och passningsmått

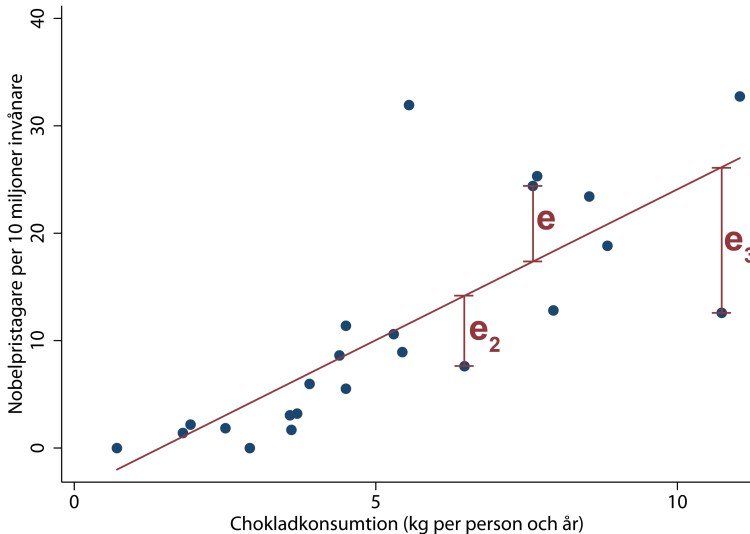
$$y = a + bx + e$$





Repetition och passningsmått

$$y = a + bx + e$$

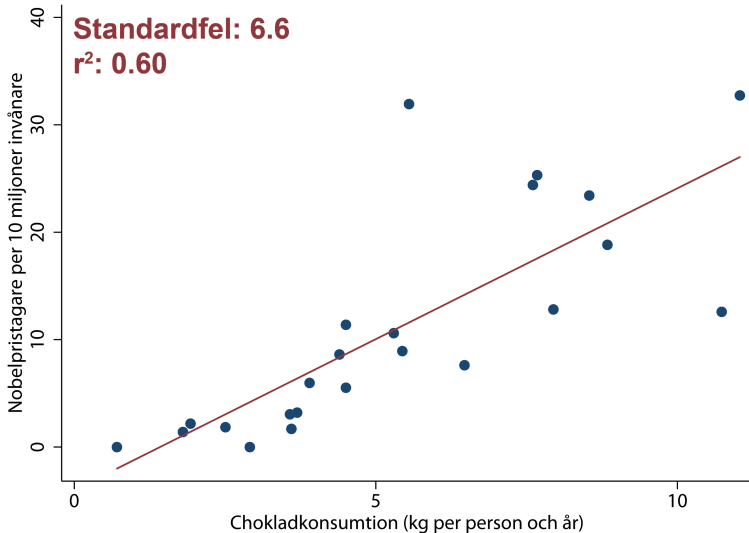




Repetition och passningsmått

$$y = a + bx + e$$

Standardfel: 6.6
 r^2 : 0.60





Repetition och passningsmått

- Passningsmått beskriver hur väl vår modell beskriver den data vi har observerat.
- De två viktigaste är regressionens standardfel och R^2 .
- Båda passningsmåten utgår ifrån storleken på residualerna, men sätter den i relation till olika saker.



Repetition och passningsmått

Regressionens standardfel

~Den genomsnittliga avvikelsen från regressionslinjen. Uttrycks i samma enheter som den beroende variabeln. Exempel: De observerade värdena avviker i genomsnitt från modellens prediktioner med 6,6 Nobelpristagare per 10 milj. invånare.

Regressions-
analys

Regressionskvationen

Passningsmått

Statistisk
signifikans

Tillämpning

Kommentarer

Regressions-
tabeller

Mjukvara

Multipel
regression

Exempel 1

Exempel 2

Exempel 3

Kombinations-
studier

Övrigt



Repetition och passningsmått

Regressionens standardfel

~Den genomsnittliga avvikelsen från regressionslinjen. Uttrycks i samma enheter som den beroende variabeln. Exempel: De observerade värdena avviker i genomsnitt från modellens prediktioner med 6,6 Nobelpristagare per 10 milj. invånare.

Överkurs

$$\text{Standardfel} = \sqrt{\frac{RSS}{n-1-k}} = \sqrt{\frac{\sum(e_i^2)}{n-1-k}} = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-1-k}}$$

RSS = Summan av de kvadrerade feltermerna (Residual Sum of Squares)

k = Antalet oberoende variabler



Repetition och passningsmått

R^2

Andelen förklarad variation i den beroende variabeln. Antar värden mellan 0 (vår modell förklarar ingenting) och 1 (vår modell förklarar 100 procent av variationen i den beroende variabeln). Exempel: Skillnader i chokladkonsumtion kan "förklara" 60 procent av variationen mellan länder i antalet Nobelpristagare.



Repetition och passningsmått

R^2

Andelen förklarad variation i den beroende variabeln. Antar värden mellan 0 (vår modell förklarar ingenting) och 1 (vår modell förklarar 100 procent av variationen i den beroende variabeln). Exempel: Skillnader i chokladkonsumtion kan "förklara" 60 procent av variationen mellan länder i antalet Nobelpristagare.

Överkurs

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2}$$

RSS = Summan av de kvadrerade feltermerna (Residual Sum of Squares)

TSS = Summan av avvikelserna från medelvärdet (Total Sum of Squares)



Repetition och passningsmått

- I grunden mäter båda måtten samma sak. Givet en viss variation i den beroende variabeln, så ökar R^2 när standardfelet sjunker, och vice versa.
- Om variationen i den beroende variabeln är stor, kan standardfelet vara stort trots högt R^2 , och vice versa.



Repetition och passningsmått

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipl regression

Exempel 1

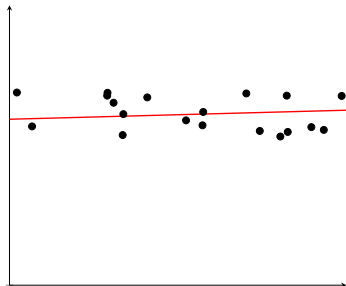
Exempel 2

Exempel 3

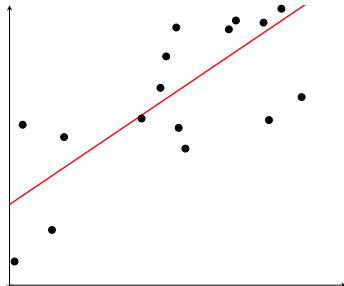
Kombinations- studier

Övrigt

Litet standardfel, lågt R^2



Stort standardfel, högt R^2





Repetition och passningsmått

Justerat R^2

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- När man adderar en variabel till en regressionsmodell kommer R^2 alltid att öka, även om den inte har något med den beroende variabeln att göra.
- För att korrigera för detta bör man i regel använda ett mått som kallas för justerat R^2 .
- Det är vanligt (och ok för er) att även justerat R^2 uttrycks som andel av variationen i den beroende variabeln som modellen förklarar.
- Mer korrekt: "justerat för antalet frihetsgrader".



Repetition och passningsmått

Allmänt om passningsmått

- Vad som är högt och lågt beror som alltid på vad vi har att jämföra med. Studenter har ofta orimligt höga förväntningar på vad våra modeller kan åstadkomma.
- Stirra er inte blinda på passningsmåten. Vårt mål är sällan att göra de bästa prediktionerna. Vanligare att vi är intresserade av ett kausalt samband.
- Då är det viktigare hur stor effekten är samt huruvida den är statistiskt signifikant, alltså om samvariationen i vårt urval kan bero på slumpen.



Korrelation är inte kausalitet

I grunden finns det fyra olika skäl till att observera en samvariation mellan två variabler x och y .

- 1** Sambandet är kausalt, så att x har en effekt på y .



Korrelation är inte kausalitet

I grunden finns det fyra olika skäl till att observera en samvariation mellan två variabler x och y .

- 1** Sambandet är kausalt, så att x har en effekt på y .
- 2** Det observerade sambandet beror på en tillfällighet.



Korrelation är inte kausalitet

I grunden finns det fyra olika skäl till att observera en samvariation mellan två variabler x och y .

- 1** Sambandet är kausalt, så att x har en effekt på y .
- 2** Det observerade sambandet beror på en tillfällighet.
- 3** Omvänd orsaksriktning: det är y som påverkar x .



Korrelation är inte kausalitet

I grunden finns det fyra olika skäl till att observera en samvariation mellan två variabler x och y .

- 1 Sambandet är kausalt, så att x har en effekt på y .
- 2 Det observerade sambandet beror på en tillfällighet.
- 3 Omvänd orsaksriktning: det är y som påverkar x .
- 4 Sambandet är *spuriöst* och beror på att en tredje variabel påverkar både x och y .



Korrelation är inte kausalitet

I grunden finns det fyra olika skäl till att observera en samvariation mellan två variabler x och y .

- 1 Sambandet är kausalt, så att x har en effekt på y .
- 2 Det observerade sambandet beror på en tillfällighet.
- 3 Omvänd orsaksriktning: det är y som påverkar x .
- 4 Sambandet är *spuriöst* och beror på att en tredje variabel påverkar både x och y .



Statistisk signifikans

Vi är intresserade av den "sanna" modellen

$$y = \alpha + \beta x + e \quad (1)$$

Men observerar endast vår uppskattning av den

$$y = a + bx + e \quad (2)$$



Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

Den sanna modellen som populationens modell

- T&S: Vi är intresserade av om det finns ett samband i populationen.
- Eftersom vi endast studerar ett urval kommer urvalskoefficienten (b) på grund av slumpmässiga mätfel avvika från populationens regressionskoefficient (β).
- Vi kanske studerar ett urval av Sveriges befolkning och noterar att kvinnor tjänar mindre än män i det urvalet.
- Signifikanstest: Kan vi vara säkra på att det finns en skillnad ($\beta \neq 0$) i populationen, eller kan urvalssambandet bero på en slump?



Statistisk signifikans

Den sanna modellen som en datagenererande process

- Jag föredrar att tänka på det som att vi generaliserar från vårt urval till en datagenererande process (summan av alla de underliggande processer och strukturer som bestämmer värdet på en variabel).
- Signifikanstest: Kan vi vara säkra på att ens inkomst påverkas av vilket kön man har ($\beta \neq 0$), eller kan urvalssambandet bero på en slump?
- Ingen skillnad i hur man räknar, men det motiverar signifikanstest även vid totalundersökningar och avsaknad av population:
 - Vi studerar världens alla konflikter och konstaterar att demokratier under den studerade perioden har krigat mindre än icke-demokratier.
 - Vi genomför ett experiment där 1000 personer får sockerpiller och 1000 personer får ett läkemedel.



Statistisk signifikans

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

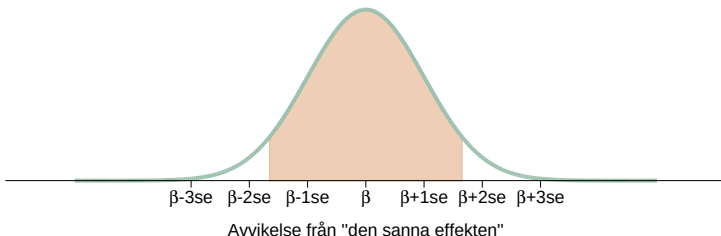
- Precis som tidigare anges den osäkerhet som beror på slumpmässiga mätfel som ett standardfel.
- Dessa beräknar datorn åt oss (alt. T&S s. 210–211).
- Om vi drog ett oändligt antal urval, skulle b "i genomsnitt" avvika ett standardfel från β .
- Ju större standardfel, desto sämre precision i vår skattning av β .
- Regressionskoefficienternas standardfel har ingenting att göra med passningsmättet regressionsstandardfel.



Statistisk signifikans

Med stora urval kan vi anta att:

- 90 procent av b-värdena avviker mindre än 1,65 standardfel från β .

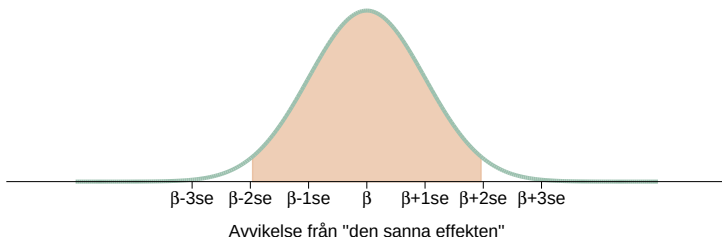




Statistisk signifikans

Med stora urval kan vi anta att:

- 90 procent av b-värdena avviker mindre än 1,65 standardfel från β .
- 95 procent av b-värdena avviker mindre än 1,96 standardfel från β .

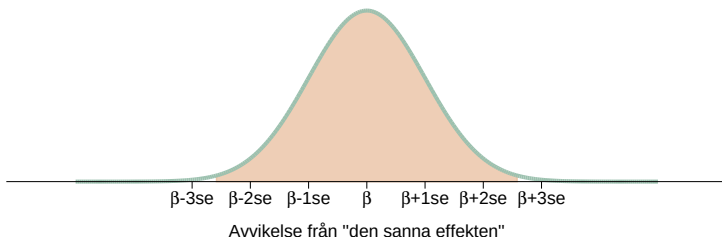




Statistisk signifikans

Med stora urval kan vi anta att:

- 90 procent av b-värdena avviker mindre än 1,65 standardfel från β .
- 95 procent av b-värdena avviker mindre än 1,96 standardfel från β .
- 99 procent av b-värdena avviker mindre än 2,58 standardfel från β .





Statistisk signifikans

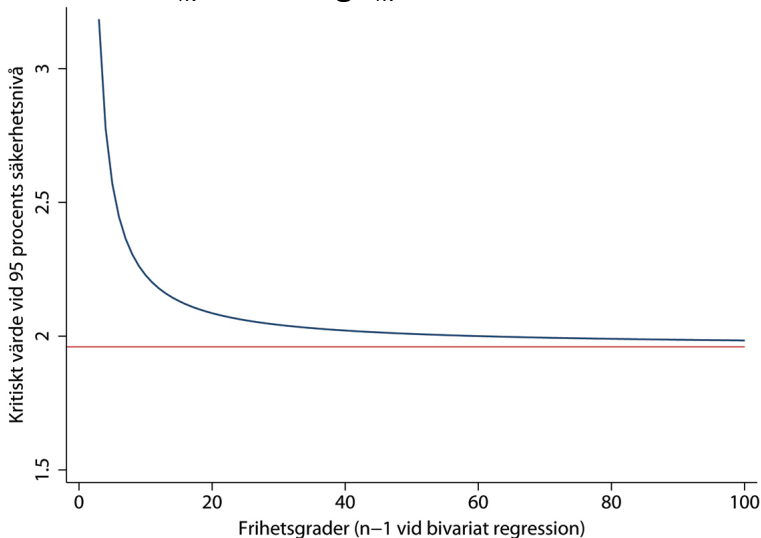
Med stora urval kan vi anta att:

- 90 procent av b-värdena avviker mindre än 1,65 standardfel från β .
- 95 procent av b-värdena avviker mindre än 1,96 standardfel från β .
- 99 procent av b-värdena avviker mindre än 2,58 standardfel från β .
- Eftersom vi använder t-fördelningen är de kritiska värdena något större än z_{kv} vid små urval.



Statistisk signifikans

t_{kv} närmar sig z_{kv} vid stora urval



Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipl regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt



Statistisk signifikans

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

- Om b avviker *mer* än 1,65 standardfel från 0 säger vi att sambandet är statistiskt signifikant vid 90 procents säkerhetsnivå.
- Sannolikheten för att detta ska inträffa är ju "bara" 10 procent om $\beta = 0$.
- Statistisk signifikans innebär att vi förkastar $\beta = 0$.
- Vi skulle också kunna beräkna ett konfidensintervall för β enligt $b \pm se$ och notera att 0 inte ligger i intervallet.



Statistisk signifikans

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

Lite mer formaliserat

- För att svara på om ett samband är statistiskt signifikant beräknar vi först ett t-värde:

$$t = \frac{b}{se_b}$$

- Om (det absoluta värdet av) t-värdet är större än det kritiska t-värdet för vår säkerhetsnivå säger vi att sambandet är statistiskt signifikant.
 - Kan också uttryckas som att sambandet är signifikant om testvärdet ligger utanför intervallet mellan $-t_{kv}$ och t_{kv} .
- Vi vet då att det finns ett samband i populationen och att den observerade samvariationen inte var en tillfällighet.



Statistisk signifikans

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Vi ska titta på tre exempel från Marcus föreläsning

- 1** Sambandet mellan kön och placering på vänster–höger-skalan.
- 2** Sambandet mellan religiositet och korruption.
- 3** Sambandet mellan tilltro till regeringen och benägenhet att ge till tiggare.



Statistisk signifikans

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipl regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Variabel	Kodning
Höger	1 (klart till vänster) - 5 (klart till höger)
Kvinna	0 (man) eller 1 (kvinna)
Månadsinkomst (brutto)	1 (<10 000 kr) till 12 (>60 000 kr)
Korruption	0 (ingen korruption) - 10 (mycket korruption)
Religiositet	1 (inte alls viktigt) till 4 (mycket viktigt)
BNP per capita	Tusentals dollar
Skänker till tiggare	1 (< 1 gång/vecka) till 4 (varje dag)
Förtroende för regeringen	1 (inget förtroende) till 4 (fullt förtroende)
Borgerlig	Skulle rösta på ett borgerligt parti

Tips: Döp gärna dikotoma variabler efter hur de är kodade.
Kvinna och höger har en underförstådd tolkning och går därför snabbare att tolka än kön och ideologisk position.



Statistisk signifikans

Exempel 1: Kön och placering på vänster–höger-skalan

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- När vi genomför den bivariata regressionsanalysen erhåller vi följande värden:

$$b = -0,125$$

$$se(b) = 0,027$$

$$t = -4,63 \ (-0,125/0,027)$$



Statistisk signifikans

Exempel 1: Kön och placering på vänster–höger-skalan

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- När vi genomför den bivariata regressionsanalysen erhåller vi följande värden:

$$b = -0,125$$

$$se(b) = 0,027$$

$$t = -4,63 \ (-0,125/0,027)$$

- Det kritiska t-värdet på 99 procents säkerhetsnivå är här 2,58 (n=7329). Eftersom 4,63 är större än 2,58 (-4,63 ligger utanför det kritiska intervallet från -2,58 till 2,58) kan vi konstatera att effekten av kön på den ideologiska positionen är statistiskt signifikant vid 99 procents säkerhetsnivå.



Statistisk signifikans

Exempel 2: Religiositet och korruption

Regressions- analys

Regressionskvantilen
Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- När vi genomför den bivariata regressionsanalysen erhåller vi följande värden:

$$b = 1,808$$

$$se(b) = 0,313$$

$$t = 5,78 \ (1,808/0,313)$$



Statistisk signifikans

Exempel 2: Religiositet och korruption

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- När vi genomför den bivariata regressionsanalysen erhåller vi följande värden:

$$b = 1,808$$

$$se(b) = 0,313$$

$$t = 5,78 \ (1,808/0,313)$$

- Det kritiska t-värdet på 99 procents säkerhetsnivå är här 2,64 (n=78). Då 5,78 är större än 2,64 (5,78 ligger utanför det kritiska intervallet från -2,64 till 2,64) kan vi slå fast att sambandet är statistiskt signifikant på 99 procents säkerhetsnivå.



Statistisk signifikans

Exempel 3: Tilltro till regeringen och benägenhet att ge till tiggare

- När vi genomför den bivariata regressionsanalysen erhåller vi följande värden:

$$b = -0,188$$

$$se(b) = 0,156$$

$$t = -1,21 \ (-0,188/0,156)$$

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Tillämpning

Kommentarer

Regressions-
tabeller

Mjukvara

Multipel
regression

Exempel 1

Exempel 2

Exempel 3

Kombinations-
studier

Övrigt



Statistisk signifikans

Exempel 3: Tilltro till regeringen och benägenhet att ge till tiggare

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- När vi genomför den bivariata regressionsanalysen erhåller vi följande värden:

$$b = -0,188$$

$$se(b) = 0,156$$

$$t = -1,21 \ (-0,188/0,156)$$

- Det kritiska t-värdet på 90 procents säkerhetsnivå är här 1,67 (n=55). Eftersom 1,21 är mindre än det kritiska värdet är sambandet inte statistiskt signifikant.



Statistisk signifikans

Exempel 3: Tilltro till regeringen och benägenhet att ge till tiggare

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- När vi genomför den bivariata regressionsanalysen erhåller vi följande värden:
 $b = -0,188$
 $se(b) = 0,156$
 $t = -1,21 \ (-0,188/0,156)$
- Det kritiska t-värdet på 90 procents säkerhetsnivå är här 1,67 ($n=55$). Eftersom 1,21 är mindre än det kritiska värdet är sambandet inte statistiskt signifikant.
- Det är inte tillräckligt osannolikt att sambandet beror på slumpen.



Statistisk signifikans

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipl regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

Fyra ekvivalenta metoder. Sambandet är signifikant om:

- T-värdet är högre än det kritiska t-värdet.



Statistisk signifikans

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

Fyra ekvivalenta metoder. Sambandet är signifikant om:

- T-värdet är högre än det kritiska t-värdet.
- Konfidensintervallet runt koefficienten omsluter inte värdet 0.



Statistisk signifikans

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipl regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Fyra ekvivalenta metoder. Sambandet är signifikant om:

- T-värdet är högre än det kritiska t-värdet.
- Konfidensintervallet runt koefficienten omsluter inte värdet 0.
- Det står asterisk(er) efter regressionskoefficienten. Läs under tabellen för att se vilken säkerhetsnivå de motsvarar. Det är den vanligaste metoden när man läser en regressionstabell.



Statistisk signifikans

Fyra ekvivalenta metoder. Sambandet är signifikant om:

- T-värdet är högre än det kritiska t-värdet.
- Konfidensintervallet runt koefficienten omsluter inte värdet 0.
- Det står asterisk(er) efter regressionskoefficienten. Läs under tabellen för att se vilken säkerhetsnivå de motsvarar. Det är den vanligaste metoden när man läser en regressionstabell.
- P-värdet är mindre än risknivån ($\text{risk} = 1 - \text{säkerhetsnivå}$). Det är den vanligaste metoden när man tolkar output från ett statistikprogram.



Statistisk signifikans

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipl regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Statistisk signifikans är inte allt

- Försumbara effekter kan vara signifikanta vid stora urval.



Statistisk signifikans

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Statistisk signifikans är inte allt

- Försumbara effekter kan vara signifikanta vid stora urval.
- Sambandet kan fortfarande vara spuriöst.



Statistisk signifikans

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Statistisk signifikans är inte allt

- Försumbara effekter kan vara signifikanta vid stora urval.
- Sambandet kan fortfarande vara spuriöst.
- Orsaksriktningen kan fortfarande vara den omvända.



Statistisk signifikans

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipl regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Många signifikanta samband beror på slumpen

- Eftersom vi inte använder 100 procents säkerhetsnivå finns det alltid en risk att vi förkastar en sann nollhypotes.



Statistisk signifikans

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

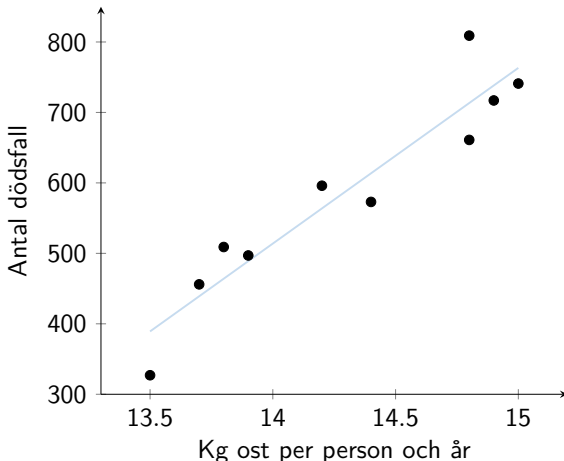
Många signifikanta samband beror på slumpen

- Eftersom vi inte använder 100 procents säkerhetsnivå finns det alltid en risk att vi förkastar en sann nollhypotes.
- Den risken ökar om vi planlöst letar efter signifikanta samband i stället för att styras av teoretiskt motiverade hypoteser.



Statistisk signifikans

Exempel: sambandet mellan ostkonsumtion och hur många som årligen avlider efter att ha fastnat i lakanen.





Regressionstabeller

	(1)	(2)
Choklad	2,81*** (0,50)	2,28*** (0,64)
BNP/capita		0,20 (0,16)
Konstant	– 3,99 (3,00)	– 8,42* (4,58)
Observationer	23	23
Standardfel	6,60	6,51
R^2	0,60	0,63

Standardfel i parenteser. * $p=10$, ** $p=05$, *** $p=01$.

En kolumn per regressionsmodell.

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Tillämpning
Kommentarer

Regressions-
tabeller

Mjukvara

Multipel
regression

Exempel 1

Exempel 2

Exempel 3

Kombinations-
studier

Övrigt



Regressionstabeller

	(1)	(2)
Choklad	2,81*** (0,50)	2,28*** (0,64)
BNP/capita		0,20 (0,16)
Konstant	− 3,99 (3,00)	− 8,42* (4,58)
Observationer	23	23
Standardfel	6,60	6,51
R^2	0,60	0,63

Standardfel i parenteser. * $p=10$, ** $p=05$, *** $p=01$.

Variabler

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt



Regressionstabeller

	(1)	(2)
Choklad	2,81*** (0,50)	2,28*** (0,64)
BNP/capita		0,20 (0,16)
Konstant	− 3,99 (3,00)	− 8,42* (4,58)
Observationer	23	23
Standardfel	6,60	6,51
R^2	0,60	0,63

Standardfel i parenteser. * $p=10$, ** $p=05$, *** $p=01$.

Regressionskoefficienter i den första regressionen.

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt



Regressionstabeller

	(1)	(2)
Choklad	2,81*** (0,50)	2,28*** (0,64)
BNP/capita		0,20 (0,16)
Konstant	— 3,99 (3,00)	— 8,42* (4,58)
Observationer	23	23
Standardfel	6,60	6,51
R^2	0,60	0,63

Standardfel i parenteser. * $p=10$, ** $p=05$, *** $p=01$.

Intercept i den första regressionen.

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Tillämpning
Kommentarer

Regressions-
tabeller

Mjukvara

Multipel
regression

Exempel 1

Exempel 2

Exempel 3

Kombinations-
studier

Övrigt



Regressionstabeller

	(1)	(2)
Choklad	2,81*** (0,50)	2,28*** (0,64)
BNP/capita		0,20 (0,16)
Konstant	− 3,99 (3,00)	− 8,42* (4,58)
Observationer	23	23
Standardfel	6,60	6,51
R^2	0,60	0,63

Standardfel i parenteser. * $p=10$, ** $p=05$, *** $p=01$.

Regressionskoefficienter i den andra regressionen.

Regressions-
analys

Regressionsekvationen
Passningsmått

Statistisk
signifikans

Tillämpning
Kommentarer

Regressions-
tabeller

Mjukvara

Multipel
regression

Exempel 1

Exempel 2

Exempel 3

Kombinations-
studier

Övrigt



Regressionstabeller

	(1)	(2)
Choklad	2,81*** (0,50)	2,28*** (0,64)
BNP/capita		0,20 (0,16)
Konstant	– 3,99 (3,00)	– 8,42* (4,58)
Observationer	23	23
Standardfel	6,60	6,51
R^2	0,60	0,63

Standardfel i parenteser. * $p=10$, ** $p=05$, *** $p=01$.

Intercept i den andra regressionen.

Regressions-
analys

Regressionsekvationen
Passningsmått

Statistisk
signifikans

Tillämpning
Kommentarer

Regressions-
tabeller

Mjukvara

Multipel
regression

Exempel 1

Exempel 2

Exempel 3

Kombinations-
studier

Övrigt



Regressionstabeller

	(1)	(2)
Choklad	2,81*** (0,50)	2,28*** (0,64)
BNP/capita		0,20 (0,16)
Konstant	− 3,99 (3,00)	− 8,42* (4,58)
Observationer	23	23
Standardfel	6,60	6,51
R^2	0,60	0,63

Standardfel i parenteser. * $p=10$, ** $p=05$, *** $p=01$.

Koefficienternas standardfel.

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt



Regressionstabeller

	(1)	(2)
Choklad	2,81*** (0,50)	2,28*** (0,64)
BNP/capita		0,20 (0,16)
Konstant	− 3,99 (3,00)	− 8,42* (4,58)
Observationer	23	23
Standardfel	6,60	6,51
R^2	0,60	0,63

Standardfel i parenteser. * $p=10$, ** $p=05$, *** $p=01$.

Passningsmått i de båda regressionerna.

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt



Hur gör man?

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- Ni behöver inte kunna räkna ut några regressionskoefficienter.
- Vi öppnar ett dataset och berättar för programmet vilken ekvation vi vill anpassa till vår data.
- Jag ska visa hur det ser ut i programmet R.



Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

- De flesta på institutionen använder **Stata**. Det är stort i forskarvärlden men inte lika vanligt utanför akademien.
- Vi tror att **SPSS** är det vanligaste programmet på svenska myndigheter. Tidigare år har vi använt gratisvarianten PSPP på kursen.
- **SAS** om ni vill arbeta på SCB med stora registerdatabaser.
- På kursen använder vi den här terminen **R**. Det är gratis och mycket avancerat.



Regressionsanalys i R

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

De flesta gör saker i R genom att skriva kod

```
mod1 <- lm(nobel ~ chocolate + gdp, data=nobel)
summary(mod1)
```



Regressionsanalys i R

Regressions- analys

Regressionslikningen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Men ni ska få klicka er fram



Regressionsanalys i R

Regressions- analys

Regressionsekvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

Oavsett metod ser outputen ut såhär

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-8.4215486	4.5811111	-1.838	0.08091	.
chocolate	2.2812114	0.6434941	3.545	0.00203	**
gdp	0.0002050	0.0001617	1.267	0.21958	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.51 on 20 degrees of freedom

Multiple R-squared: 0.6313, Adjusted R-squared: 0.5944

F-statistic: 17.12 on 2 and 20 DF, p-value: 4.643e-05



Regressionsanalys i R

Regressions- analys

Regressionsekvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

Konstant: -8.42

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-8.4215486	4.5811111	-1.838	0.08091	.
chocolate	2.2812114	0.6434941	3.545	0.00203	**
gdp	0.0002050	0.0001617	1.267	0.21958	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.51 on 20 degrees of freedom

Multiple R-squared: 0.6313, Adjusted R-squared: 0.5944

F-statistic: 17.12 on 2 and 20 DF, p-value: 4.643e-05



Regressionsanalys i R

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Regressionskoefficient för choklad: 2.28

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-8.4215486	4.5811111	-1.838	0.08091 .
chocolate	2.2812114	0.6434941	3.545	0.00203 **
gdp	0.0002050	0.0001617	1.267	0.21958

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.51 on 20 degrees of freedom

Multiple R-squared: 0.6313, Adjusted R-squared: 0.5944

F-statistic: 17.12 on 2 and 20 DF, p-value: 4.643e-05



Regressionsanalys i R

Regressions- analys

Regressionsekvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

Koefficientens standardfel: 0.64

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-8.4215486	4.5811111	-1.838	0.08091	.
chocolate	2.2812114	0.6434941	3.545	0.00203	**
gdp	0.0002050	0.0001617	1.267	0.21958	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.51 on 20 degrees of freedom

Multiple R-squared: 0.6313, Adjusted R-squared: 0.5944

F-statistic: 17.12 on 2 and 20 DF, p-value: 4.643e-05



Regressionsanalys i R

Regressions- analys

Regressionsekvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

Koefficientens t-värde: 3.55

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-8.4215486	4.5811111	-1.838	0.08091	.
chocolate	2.2812114	0.6434941	3.545	0.00203	**
gdp	0.0002050	0.0001617	1.267	0.21958	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.51 on 20 degrees of freedom

Multiple R-squared: 0.6313, Adjusted R-squared: 0.5944

F-statistic: 17.12 on 2 and 20 DF, p-value: 4.643e-05



Regressionsanalys i R

Regressions- analys

Regressionsekvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

Koefficientens p-värde: 0.00

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-8.4215486	4.5811111	-1.838	0.08091	.
chocolate	2.2812114	0.6434941	3.545	0.00203	**
gdp	0.0002050	0.0001617	1.267	0.21958	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.51 on 20 degrees of freedom

Multiple R-squared: 0.6313, Adjusted R-squared: 0.5944

F-statistic: 17.12 on 2 and 20 DF, p-value: 4.643e-05



Regressionsanalys i R

Regressions- analys

Regressionsekvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

R^2 : 0.63

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-8.4215486	4.5811111	-1.838	0.08091 .
chocolate	2.2812114	0.6434941	3.545	0.00203 **
gdp	0.0002050	0.0001617	1.267	0.21958

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.51 on 20 degrees of freedom

Multiple R-squared: 0.6313, Adjusted R-squared: 0.5944

F-statistic: 17.12 on 2 and 20 DF, p-value: 4.643e-05



Regressionsanalys i R

Regressions- analys

Regressionsekvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

Justerat R^2 : 0.59

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-8.4215486	4.5811111	-1.838	0.08091 .
chocolate	2.2812114	0.6434941	3.545	0.00203 **
gdp	0.0002050	0.0001617	1.267	0.21958

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.51 on 20 degrees of freedom

Multiple R-squared: 0.6313, **Adjusted R-squared: 0.5944**

F-statistic: 17.12 on 2 and 20 DF, p-value: 4.643e-05



Multipel regression

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- Fram tills nu har vi bara studerat samband mellan två variabler (bivariat regression). Multipel regression innebär att vi har fler än två variabler.
- Allt ni har lärt er hittills gäller även vid multipel regression, med vissa tillägg.
- Rent tekniskt lägger vi bara till fler variabler i regressionsekvationen.



Multipel regression

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- Sådär kan vi skriva en modell med två oberoende variabler:

$$y = a + b_1x_1 + b_2x_2 + e$$



Multipel regression

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- En modell med tre oberoende variabler:

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + e$$



Multipel regression

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- En modell med fyra oberoende variabler:

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + e$$



Multipel regression

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- En modell med fyra oberoende variabler:

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + e$$

- b_1 anger den förväntade ökningen i y när x_1 ökar med 1 och värdet på alla andra oberoende variabler hålls konstant.
- Uttrycks ofta som "när vi kontrollerar för (exempelvis) kön, ålder och inkomst".
- Givet värdena på x_2 , x_3 och x_4 , spelar värdet på x_1 fortfarande någon roll för y ?



Multipel regression

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- Det finns tre huvudsakliga skäl för att använda fler än en oberoende variabel:
 - 1 Förbättra förklaringen
 - 2 Hitta orsaksmekanism
 - 3 Isolera sambandet



Multipel regression

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Kom ihåg orsakskriterierna!

- Samvariation (kontrafaktiska samband) har vi visat på redan i den bivariata analysen.
- Klarlägga tidsordning är inte de kvantitativa metodernas styrka (det finns emellertid mer avancerade regressionsmodeller som försöker göra detta).
- Men genom att addera fler variabler till modellen kan vi både isolera sambandet från alternativa förklaringar samt hitta variabler som fungerar som orsaksmekanismer.



Multipel regression

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

1 Förbättra förklaringen med fler variabler

- Få samhällsfenomen kan förklaras av en enda variabel.
- Om de nya variablerna har ett samband med y kommer vi med en multipel modell förklara en större del av variationen i y och göra bättre prediktioner.
- Vi kan fortsätta att addera fler och fler variabler, men i regel väljer vi variabler som är viktiga för att isolera sambandet eller hitta orsaksmekanismen.
- Ju fler observationer vi har, desto fler variabler kan vi inkludera.



Multipel regression

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

2 Hitta orsaksmekanismen

- En variabel som påverkas av den oberoende variabeln samt påverkar den beroende variabeln kallas för mellanliggande variabel.
- Den kan peka på hur eller varför den oberoende variabeln påverkar den beroende variabeln. Kallas därför ofta för orsaksmekanism.
- Det ursprungliga sambandet mellan den oberoende och beroende variabeln är fortfarande kausalt, men indirekt.
- Att hitta mekanismer stärker trovärdigheten för våra resultat.



Multipel regression

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

3 Isolera sambandet från bakomliggande variabler

- Att ett samband mellan två variabler är statistiskt signifikant betyder inte att det måste föreligga ett kausalt samband mellan variablerna. En annan möjlighet är att samvariationen mellan de två variablerna beror på en tredje, bakomliggande variabel som påverkar de båda.
- Exempelvis kanske vi kan tänka oss att rika länder både har råd med mycket choklad och lägger mycket pengar på spetsforskning. Det skulle i så fall kunna förklara varför länder med hög chokladkonsumtion har fått fler Nobelpris. Vi skulle då kalla det ursprungliga sambandet för spuriöst eller skenbart.



Korrelation är inte kausalitet

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipl regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

I grunden finns det fyra olika skäl till att observera en samvariation mellan två variabler x och y .

- 1 Sambandet är kausalt, så att x har en effekt på y .
- 2 Det observerade sambandet beror på en tillfällighet.
- 3 Omvänd orsaksriktning: det är y som påverkar x .
- 4 Sambandet är *spuriöst* och beror på att en tredje variabel påverkar både x och y .



Multipel regression

- Om det ursprungliga sambandet kvarstår även efter att vi kontrollerat för en tredje variabel har vi isolerat sambandet (från just den variabeln).

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt



Multipel regression

- Om det ursprungliga sambandet kvarstår även efter att vi kontrollerat för en tredje variabel har vi isolerat sambandet (från just den variabeln).
- Om det ursprungliga sambandet försvann eller försvagades efter att vi kontrollerat för en *mellanliggande* variabel var sambandet indirekt. Effekten går via den nya variabeln, vilken fungerar som orsaksmekanism.

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt



Multipel regression

- Om det ursprungliga sambandet kvarstår även efter att vi kontrollerat för en tredje variabel har vi isolerat sambandet (från just den variabeln).
- Om det ursprungliga sambandet försvann eller försvagades efter att vi kontrollerat för en *mellanliggande* variabel var sambandet indirekt. Effekten går via den nya variabeln, vilken fungerar som orsaksmekanism.
- Om det ursprungliga sambandet försvann eller försvagades efter att vi kontrollerat för en *bakomliggande* variabel var det ett skensamband och spuriöst.

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt



Multipel regression

- Om det ursprungliga sambandet kvarstår även efter att vi kontrollerat för en tredje variabel har vi isolerat sambandet (från just den variabeln).
- Om det ursprungliga sambandet försvann eller försvagades efter att vi kontrollerat för en *mellanliggande* variabel var sambandet indirekt. Effekten går via den nya variabeln, vilken fungerar som orsaksmekanism.
- Om det ursprungliga sambandet försvann eller försvagades efter att vi kontrollerat för en *bakomliggande* variabel var det ett skensamband och spuriöst.
- Om det ursprungliga sambandet förstärks efter att vi kontrollerat för en tredje variabel, fungerar den nya variabeln som en suppressorvariabel.



Multipel regression

- Om det ursprungliga sambandet kvarstår även efter att vi kontrollerat för en tredje variabel har vi isolerat sambandet (från just den variabeln).
- Om det ursprungliga sambandet försvann eller försvagades efter att vi kontrollerat för en *mellanliggande* variabel var sambandet indirekt. Effekten går via den nya variabeln, vilken fungerar som orsaksmekanism.
- Om det ursprungliga sambandet försvann eller försvagades efter att vi kontrollerat för en *bakomliggande* variabel var det ett skensamband och spuriöst.
- Om det ursprungliga sambandet förstärks efter att vi kontrollerat för en tredje variabel, fungerar den nya variabeln som en suppressorvariabel.
- Orsaksriktningen mellan variabler avgörs huvudsakligen av teoretiska resonemang.



Multipel regression

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Exempel 1

- Kvinnor står längre till vänster än män.
- Vad händer med det bivariata sambandet mellan kön och ideologisk position när vi kontrollerar för ålder?
- Data från SOM-undersökningarna.



Multipel regression

Regressions- analys

Regressionssekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Beroende variabel: Höger

	(1)	(2)
Kvinna	-0,125 (-4,63)	-0,123 (-4,62)
Ålder		0,001 (0,20)
Konstant	3,074	3,066
Observationer	7329	7329

t-värden i parenteser.



Multipel regression

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- Sambandet mellan kön och ideologisk position kvarstod efter kontroll för ålder.
- Men vad händer när vi inkluderar inkomst i modellen?



Multipel regression

Regressions- analys

Regressionssekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Beroende variabel: Höger

	(1)	(2)
Kvinna	-0,125 (-4,63)	-0,024 (-0,87)
Inkomst		0,090 (17,33)
Konstant	3,074	2,636
Observationer	7329	7329

t-värden i parenteser.



Multipel regression

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

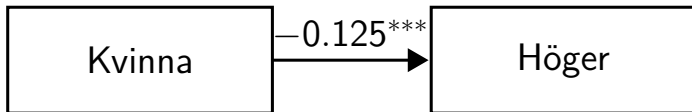
Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Ett bivariat samband...





Multipel regression

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

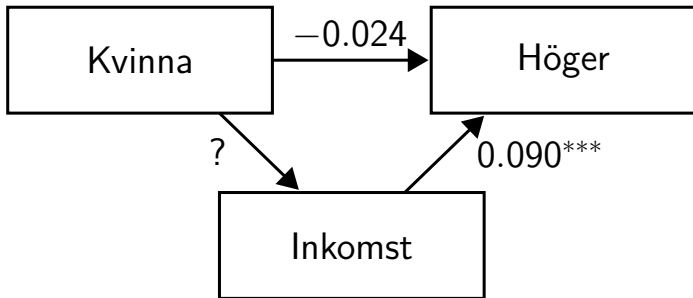
Exempel 2

Exempel 3

Kombinations- studier

Övrigt

...som var indirekt och går via inkomst...





Multipel regression

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- Hur kan vi ta reda på styrkan i sambandet mellan kön och inkomst?
- Ett sätt är förstås att göra en bivariat analys med inkomst som beroende variabel och kön som oberoende.



Multipel regression

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Beroende variabel: Inkomst

	(1)
Kvinna	-1,132 (18,96)
Konstant	4,891
Observationer	7329

t-värden i parenteser.



Multipel regression

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- Den andra metoden är lite knepigare och inte nödvändigt för er att kunna, men bidrar förhoppningsvis till en djupare förståelse. Utgångspunkten är att systemet är slutet. Den totala effekten av kön på ideologisk placering består av den direkta effekten samt den del av sambandet som går via inkomst.
- Resultatet blir detsamma bortsett från avrundningsfel.



Multipel regression

Överkurs

$$\text{Total effekt} = \text{Direkt effekt} + \text{Indirekt effekt} \quad (3)$$

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt



Multipel regression

Överkurs

$$\text{Total effekt} = \text{Direkt effekt} + \text{Indirekt effekt} \quad (3)$$

Den indirekta effekten är samma sak som effekten av kön på inkomst multiplicerat med effekten av inkomst på vänster–höger-placering. Vi känner till allt förutom effekten av kön på inkomst (b).

$$-0.125 = -0.024 + b \times 0.090 \quad (4)$$

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt



Multipel regression

Överkurs

$$\text{Total effekt} = \text{Direkt effekt} + \text{Indirekt effekt} \quad (3)$$

Den indirekta effekten är samma sak som effekten av kön på inkomst multiplicerat med effekten av inkomst på vänster–höger-placering. Vi känner till allt förutom effekten av kön på inkomst (b).

$$-0.125 = -0.024 + b \times 0.090 \quad (4)$$

Om vi adderar 0,024 till båda sidorna...

$$-0.101 = b \times 0.090 \quad (5)$$



Multipel regression

Överkurs

$$Total\ effekt = Direkt\ effekt + Indirekt\ effekt \quad (3)$$

Den indirekta effekten är samma sak som effekten av kön på inkomst multiplicerat med effekten av inkomst på vänster–höger-placering. Vi känner till allt förutom effekten av kön på inkomst (b).

$$-0.125 = -0.024 + b \times 0.090 \quad (4)$$

Om vi adderar 0,024 till båda sidorna...

$$-0.101 = b \times 0.090 \quad (5)$$

...och dividerar med 0.090 kan vi lösa ut b.

$$b = -1.122 \quad (6)$$



Multipel regression

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

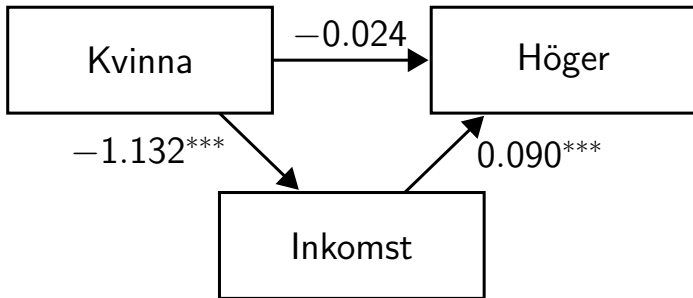
Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Vi kan nu ange effekten för varje orsakspil.





Multipel regression

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- Precis som i det bivariata fallet kan vi beräkna förväntade värden för multipla regressionsekvationer. Anta att ni blir tillfrågade om var män med en inkomst mellan 20 000 och 25 000 kr (4 på inkomstskalan) i genomsnitt befinner sig på höger-vänster-skalan.

$$\hat{y} = a + b_1x_1 + b_2x_2 \quad (7)$$



Multipel regression

Regressions- analys

Regressionssekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Beroende variabel: Höger

	(1)	(2)
Kvinna	-0,125 (-4,63)	-0,024 (-0,87)
Inkomst		0,090 (17,33)
Konstant	3,074	2,636
Observationer	7329	7329

t-värden i parenteser.



Multipel regression

Regressions- analys

Regressionsekvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- Precis som i det bivariata fallet kan vi beräkna förväntade värden för multipla regressionsekvationer. Anta att ni blir tillfrågade om var män med en inkomst mellan 20 000 och 25 000 kr (4 på inkomstskalan) i genomsnitt befinner sig på höger-vänster-skalan.

$$\hat{y} = a + b_1x_1 + b_2x_2 \quad (8)$$

$$Hoger = 2.636 + 0.090 \times 4 - 0.024 \times 0 = 2.996 \quad (9)$$

- Om en effekt har visat sig vara insignifikant kan man inte bara utesluta den ur ekvationen. Antingen låter man den vara kvar ("best guess") eller skattar man ett bivariat samband mellan de kvarvarande variablerna och använder dessa värden.



Multipel regression

Regressions- analys

Regressionsekvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Exempel 2

- Korruptionen är högre i länder där religion är en central del av många liv.
- Vad händer med det bivariata sambandet mellan religiositet och korruption när vi kontrollerar för ekonomisk utveckling?
- Data från World Value Survey och Quality of Government.



Multipel regression

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Beroende variabel: Grad av korruption

	(1)	(2)
Religiositet	1,808 (0,313)	0,239 (0,232)
BNP per capita		-0,141 (0,012)
Konstant	-0,061	6,745
Observationer	78	78

Standardfel i parenteser.



Multipel regression

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

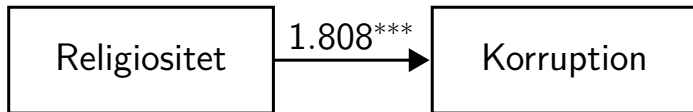
Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Ett bivariat samband...





Multipel regression

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

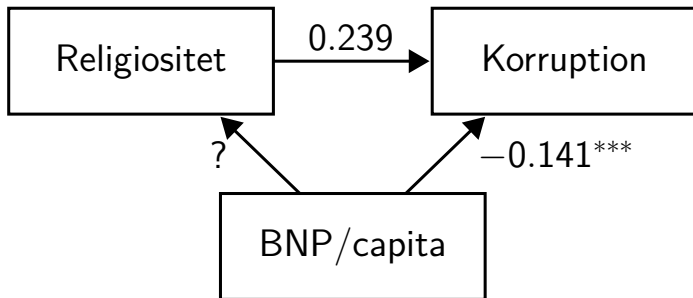
Exempel 2

Exempel 3

Kombinations- studier

Övrigt

...som var spuriöst...





Multipel regression

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

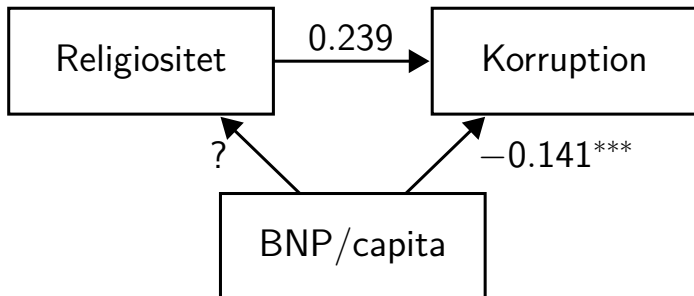
Exempel 2

Exempel 3

Kombinations- studier

Övrigt

...om vi accepterar orsaksriktningen.





Multipel regression

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Exempel 3

- Personer med stort förtroende för regeringen gav i lägre utsträckning pengar till tiggare, men sambandet var inte statistiskt signifikant.
- Vad händer med det bivariata sambandet mellan förtroende för regeringen och benägenhet att ge pengar till tiggare när vi inkluderar en dummyvariabel för borgerlig sympatisör?
- Data från Metod C-enkäten.



Multipel regression

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Beroende variabel: Ger till tiggare

	(1)	(2)
Förtroende	-0,188 (-1,20)	-0,332 (-2,15)
Borgerlig		-0,674 (-2,95)
Konstant	2,177	2,811
Observationer	55	55

t-värden i parenteser.



Multipel regression

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

- Kontrollvariabler som *stärker* sambandet kallas för suppressorvariabler. T&S skriver om detta på s. 194.



Multipel regression

Regressions- analys

Regressionsekvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

- Kontrollvariabler som *stärker* sambandet kallas för suppressorvariabler. T&S skriver om detta på s. 194.
- Order suppressor kommer förmodligen från att variabeln 'trycker tillbaka' eller konstanthåller en del av variationen i den huvudsakliga förklaringsvariabeln som har en motsatt eller svagare effekt på den beroende variabeln.



Multipel regression

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

- Kontrollvariabler som *stärker* sambandet kallas för suppressorvariabler. T&S skriver om detta på s. 194.
- Order suppressor kommer förmodligen från att variabeln 'trycker tillbaka' eller konstanthåller en del av variationen i den huvudsakliga förklaringsvariabeln som har en motsatt eller svagare effekt på den beroende variabeln.
- Tillit till regeringen sammanföll både med rödgröna sympatier (ökar benägenhet att ge) och minskat behov av privata insatser (minskar benägenhet att ge). De två effekterna tog ut varandra när vi inte kontrollerade för partisympatier.



Multipel regression

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- Idag har vi av pedagogiska skäl bara tittat på bivariata och trivariata regressioner.
- I praktiken använder vi oftast fler variabler än så, men vi tolkar resultaten på samma sätt.
- Att ändra uppsättningen variabler mellan modeller är en vanlig metod, men det behöver inte ske en variabel i taget.



Kombinationsstudier

- Det förs ständigt en diskussionen om kvantitativa respektive kvalitativa metoders användningsområden och begränsningar.

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipl regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt



Kombinationsstudier

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipl regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

- Det förs ständigt en diskussionen om kvantitativa respektive kvalitativa metoders användningsområden och begränsningar.
- De flesta är nog överens om att uppdelningen "kvantare" och "kvallare" är lite olycklig, särskilt som den leder till onödiga positioneringar och begränsar forskarens möjliga angreppssätt.



Kombinationsstudier

Regressions- analys

Regressionsekvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

- Det förs ständigt en diskussionen om kvantitativa respektive kvalitativa metoders användningsområden och begränsningar.
- De flesta är nog överens om att uppdelningen "kvantare" och "kvallare" är lite olycklig, särskilt som den leder till onödiga positioneringar och begränsar forskarens möjliga angreppssätt.
- Många framhäver intensiva och extensiva studier som komplementära och kombinationen av dem är idag ett populärt ideal.



Kombinationsstudier

Styrkor och svagheter i intensiva och extensiva studier

- Det är svårt att visa på samvariation och isolera orsakssamband i en intensiv studie. Men i gengäld kan det i dessa studier ofta vara enklare att finna belägg för tidsordning och orsaksmekanism.

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt



Kombinationsstudier

Styrkor och svagheter i intensiva och extensiva studier

- Det är svårt att visa på samvariation och isolera orsakssamband i en intensiv studie. Men i gengäld kan det i dessa studier ofta vara enklare att finna belägg för tidsordning och orsaksmekanism.
- På motsatt vis är extensiva studier bra på samvariation och isolering men brister ofta i att belägga tidsordning och spåra orsaksmekanismer.

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt



Kombinationsstudier

Kontrafaktisk skillnad och isolering

- *Method of Agreement* innebär att vi väljer fall som har samma (likartade) utfall på den beroende variabeln men är så olika som möjligt i alla andra relevanta avseenden. Fortfarande inget belägg för kontrafaktisk skillnad!

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt



Kombinationsstudier

Kontrafaktisk skillnad och isolering

- *Method of Agreement* innebär att vi väljer fall som har samma (likartade) utfall på den beroende variabeln men är så olika som möjligt i alla andra relevanta avseenden. Fortfarande inget belägg för kontrafaktisk skillnad!
- Endast i *Method of Difference* har vi belägg för kontrafaktisk skillnad, men de intensiva metodernas akilleshäl kvarstår: Vi vet fortfarande inte om sambandet är systematiskt eller slumpartat.

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt



Kombinationsstudier

Kontrafaktisk skillnad och isolering

- *Method of Agreement* innebär att vi väljer fall som har samma (likartade) utfall på den beroende variabeln men är så olika som möjligt i alla andra relevanta avseenden. Fortfarande inget belägg för kontrafaktisk skillnad!
- Endast i *Method of Difference* har vi belägg för kontrafaktisk skillnad, men de intensiva metodernas akilleshäl kvarstår: Vi vet fortfarande inte om sambandet är systematiskt eller slumpartat.
- Extensiva upplägg är därför bättre på att ge belägg för samvariation eller kontrafaktisk skillnad samt isolera andra förklaringar. De har också fördelen att de kan hantera probabilistiska samband bättre än fåfallsstudier (se Teorell & Svensson s.241).



Kombinationsstudier

Tidsordning och orsaksmekanism

- Hela dagen har vi gjort antaganden om orsaksriktningar. Det brukar vara mer problematiskt än vad det har varit i våra exempel.

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt



Kombinationsstudier

Tidsordning och orsaksmekanism

- Hela dagen har vi gjort antaganden om orsaksriktningar. Det brukar vara mer problematiskt än vad det har varit i våra exempel.
- De kvantitativa metoder som finns för att belägga orsaksriktning kräver bra data med tidsvariation samt ofta andra antaganden, exempelvis om effektens fördröjning.

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt



Kombinationsstudier

Tidsordning och orsaksmekanism

- Hela dagen har vi gjort antaganden om orsaksriktningar. Det brukar vara mer problematiskt än vad det har varit i våra exempel.
- De kvantitativa metoder som finns för att belägga orsaksriktning kräver bra data med tidsvariation samt ofta andra antaganden, exempelvis om effektens fördröjning.
- Även om extensiva studier kan visa på en orsaksmekanism kan de inte följa en process lika nära som en intensiv studie kan göra.

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt



Kombinationsstudier

Tidsordning och orsaksmekanism

- Hela dagen har vi gjort antaganden om orsaksriktningar. Det brukar vara mer problematiskt än vad det har varit i våra exempel.
- De kvantitativa metoder som finns för att belägga orsaksriktning kräver bra data med tidsvariation samt ofta andra antaganden, exempelvis om effektens fördröjning.
- Även om extensiva studier kan visa på en orsaksmekanism kan de inte följa en process lika nära som en intensiv studie kan göra.
- *Utöver* detta fyller intensiva studier även viktiga teoriutvecklande eller hypotesgenererande funktioner.

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt



Kombinationsstudier

Kombinationsstudier: Det bästa av två världar?

- Vill vi hitta belägg för alla orsakskriterier bör vi således kombinera extensiva och intensiva ansatser. De är komplementära snarare än att stå i konflikt med varandra.
- Allt behöver inte göras i samma studie eller av samma forskare!

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipl regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt



Kombinationsstudier

- De vanligaste ansatserna är förmodligen att att välja fall på basis av en extensiv studie.

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt



Kombinationsstudier

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipl regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- De vanligaste ansatserna är förmodligen att att välja fall på basis av en extensiv studie.
- Om syftet är att belägga orsaksriktning och/eller orsaksmekanism bör vi välja fall som passar in i huvudmönstret, det vill säga som är representativa för det samband som vi har funnit.



Kombinationsstudier

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- De vanligaste ansatserna är förmodligen att att välja fall på basis av en extensiv studie.
- Om syftet är att belägga orsaksriktning och/eller orsaksmekanism bör vi välja fall som passar in i huvudmönstret, det vill säga som är representativa för det samband som vi har funnit.
- En annan möjlighet är att använda den intensiva studien till att generera nya (konkurrerande eller komplementterande) hypoteser om vad som kan förklara ett visst fenomen. Vi väljer då länder som ligger långt ifrån regressionslinjen.



Kombinationsstudier

Typiska fall

- Fall som är representativa med avseende på ett givet samband. Små residualer. Ligger "på linjen".

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipl regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt



Kombinationsstudier

Typiska fall

- Fall som är representativa med avseende på ett givet samband. Små residualer. Ligger "på linjen".
- Behöver inte vara representativa i de enskilda variablerna (kan ha extrema värden på de oberoende eller beroende variablerna).

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipl regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt



Kombinationsstudier

Typiska fall

- Fall som är representativa med avseende på ett givet samband. Små residualer. Ligger "på linjen".
- Behöver inte vara representativa i de enskilda variablerna (kan ha extrema värden på de oberoende eller beroende variablerna).
- Hypotesprövande.
 - Utforska kausala mekanismer.
 - Utesluta omvänd orsaksriktning och spuriösa samband.

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt



Kombinationsstudier

Typiska fall

- Fall som är representativa med avseende på ett givet samband. Små residualer. Ligger "på linjen".
- Behöver inte vara representativa i de enskilda variablerna (kan ha extrema värden på de oberoende eller beroende variablerna).
- Hypotesprövande.
 - Utforska kausala mekanismer.
 - Utesluta omvänd orsaksriktning och spuriösa samband.
- Kallas även illustrativa fall eftersom de kan ge en djupare eller mer konkret förståelse för ett samband.

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt



Kombinationsstudier

Typiska fall

- Fall som är representativa med avseende på ett givet samband. Små residualer. Ligger "på linjen".
- Behöver inte vara representativa i de enskilda variablerna (kan ha extrema värden på de oberoende eller beroende variablerna).
- Hypotesprövande.
 - Utforska kausala mekanismer.
 - Utesluta omvänd orsaksriktning och spuriösa samband.
- Kallas även illustrativa fall eftersom de kan ge en djupare eller mer konkret förståelse för ett samband.
- Ibland används begreppet även om det vi har kallat för representativa fall. Då skiljer man på typiskt för en univariat fördelning eller typiskt för ett samband.

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt



Kombinationsstudier

Avvikande fall

- Avviker från vad vi skulle förvänta oss utifrån ett visst samband. Förklaras inte av modellen.

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipl regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt



Kombinationsstudier

Avvikande fall

- Avviker från vad vi skulle förvänta oss utifrån ett visst samband. Förklaras inte av modellen.
- Stora residualer. Långt ifrån regressionslinjen.

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipl regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt



Kombinationsstudier

Avvikande fall

- Avviker från vad vi skulle förvänta oss utifrån ett visst samband. Förklaras inte av modellen.
- Stora residualer. Långt ifrån regressionslinjen.
- Hypotesgenererande.
 - Letar efter variabler som förklarar det avvikande fallet.
 - Dessa förklaringar ska helst kunna appliceras även på andra fall.

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt



Kombinationsstudier

Avvikande fall

- Avviker från vad vi skulle förvänta oss utifrån ett visst samband. Förklaras inte av modellen.
- Stora residualer. Långt ifrån regressionslinjen.
- Hypotesgenererande.
 - Letar efter variabler som förklarar det avvikande fallet.
 - Dessa förklaringar ska helst kunna appliceras även på andra fall.
- Hittar vi nya möjliga förklaringar kan vi addera dem till våra modeller och göra nya regressionsanalyser där fallen förhoppningsvis ligger närmare linjen.



Kombinationsstudier

Regressions- analys

Regressionsekvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

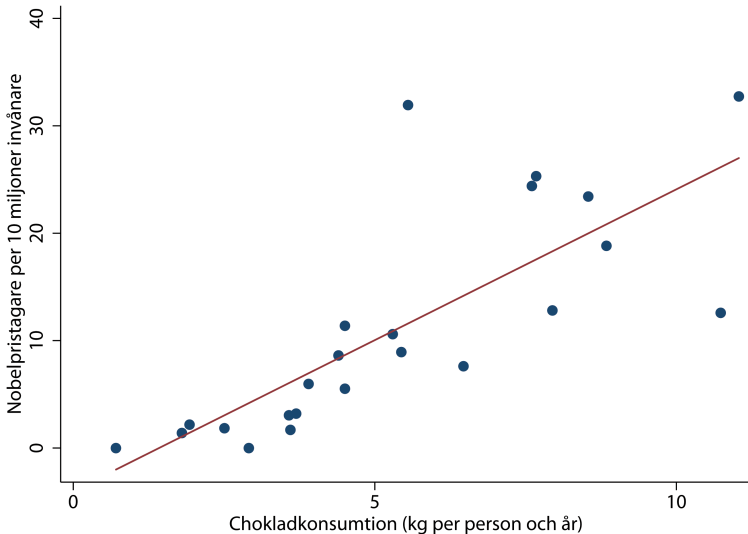
Mjukvara

Multipl regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt





UPPSALA
UNIVERSITET

Kombinationsstudier

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipl regression

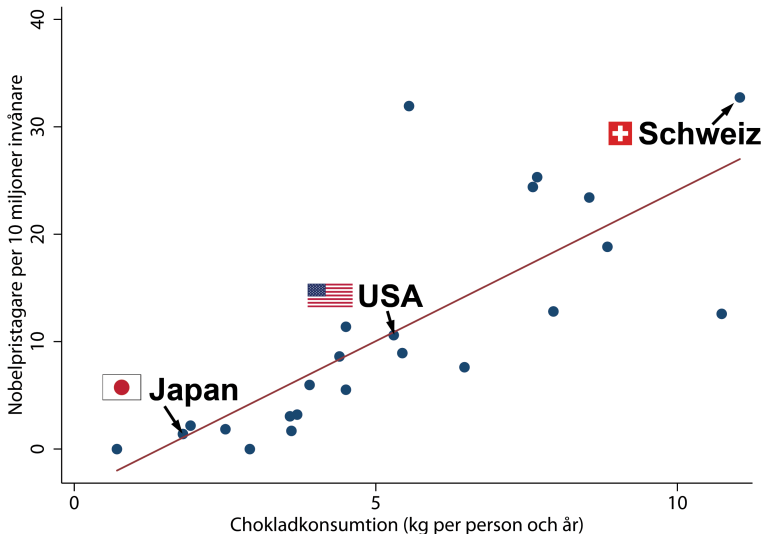
Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt





Kombinationsstudier

Regressions- analys

Regressionsekvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

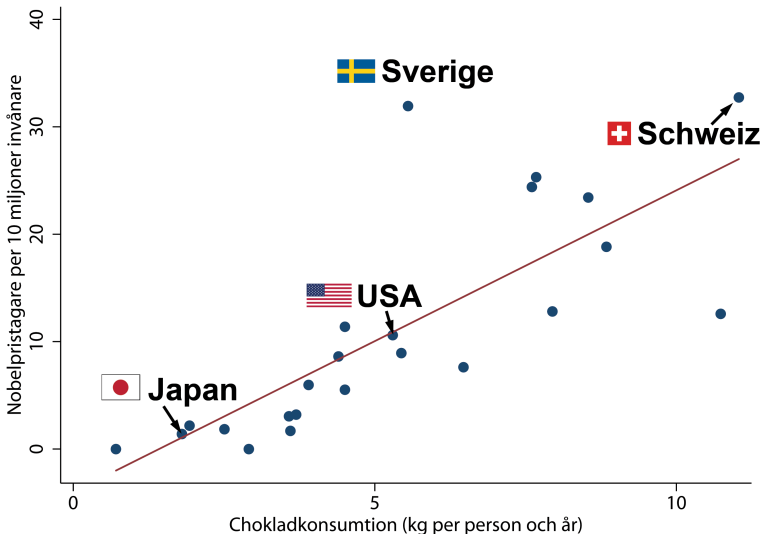
Mjukvara

Multipl regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt





Icke-linjära samband

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

- Hittills har vi antagit att alla samband är linjära, så att en ökning i x alltid ger samma ökning i $\hat{y}$.



Icke-linjära samband

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

- Hittills har vi antagit att alla samband är linjära, så att en ökning i x alltid ger samma ökning i $\hat{y}$.
- Men alla känner vi till samband där det antagandet inte stämmer.



Icke-linjära samband

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipl regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

- Hittills har vi antagit att alla samband är linjära, så att en ökning i x alltid ger samma ökning i $\hat{y}$.
- Men alla känner vi till samband där det antagandet inte stämmer.
- En person utvecklas mer mellan 2 och 4 år än mellan 30 och 32.



Icke-linjära samband

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

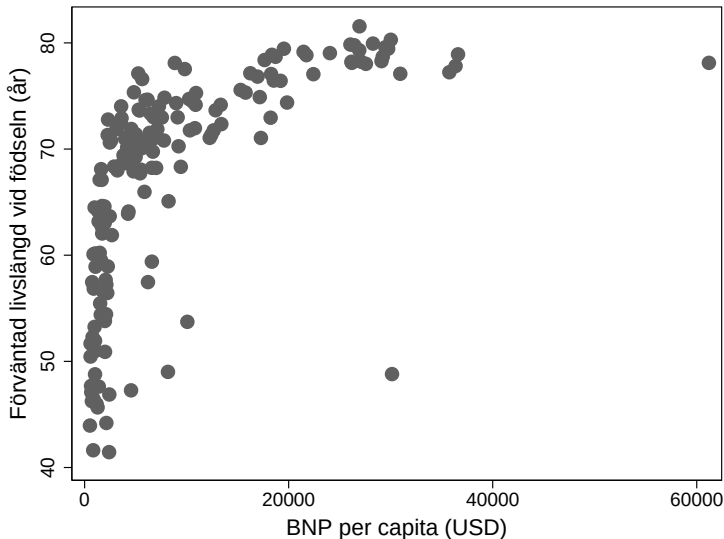
Kombinations- studier

Övrigt

- Hittills har vi antagit att alla samband är linjära, så att en ökning i x alltid ger samma ökning i $\hat{y}$.
- Men alla känner vi till samband där det antagandet inte stämmer.
- En person utvecklas mer mellan 2 och 4 år än mellan 30 och 32.
- En ökning av inkomsten med 1 000 kr betyder i de flesta avseenden mer för en student än för en höginkomsttagare.



Avtagande samband



Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipl regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt



Outliers

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

- Observationer med extrema variabelvärden kallas för outliers.
- Sådana observationer kan ha en avgörande effekt på regressionsresultaten.
- Anledning till det är att vår metod för att bestämma regressionslinjen – OLS – ger avvikande observationer stor betydelse.



Outliers

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

- Om outliers är mätfel. Ta bort dem och berätta om det.
- Om vi inte är säkra på att de är mätfel, och det är tydligt vilka observationer som är outliers, redovisa resultat för regressioner både med och utan outliers. Diskutera hur resultaten skiljer sig åt.
- Om det inte är möjligt, använd något annat än OLS (långt bortanför vår kurs).



Andra saker att vara uppmärksam på

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Det är svårt att jämföra koefficienter

- Variabler är ofta mätta på olika skalor.



Andra saker att vara uppmärksam på

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Det är svårt att jämföra koefficienter

- Variabler är ofta mätta på olika skalor.
- Även om de är mätta på samma skala, kan spridningen skilja sig åt.



Andra saker att vara uppmärksam på

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Det är svårt att jämföra koefficienter

- Variabler är ofta mätta på olika skalor.
- Även om de är mätta på samma skala, kan spridningen skilja sig åt.
- Oavsett skala och spridning är det aldrig okomplicerat att jämföra effekter av vitt skilda saker.



Andra saker att vara uppmärksam på

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Det är svårt att jämföra koefficienter

- Variabler är ofta mätta på olika skalor.
- Även om de är mätta på samma skala, kan spridningen skilja sig åt.
- Oavsett skala och spridning är det aldrig okomplicerat att jämföra effekter av vitt skilda saker.
- Det betyder inte att vi ska låta bli att jämföra!



Andra saker att vara uppmärksam på

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipl regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Regressionsanalysen testar inte orsaksriktningen

- Inom samhällsvetenskapen är det ofta långt ifrån självklart i vilken riktning som ett samband går.



Andra saker att vara uppmärksam på

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Regressionsanalysen testar inte orsaksriktningen

- Inom samhällsvetenskapen är det ofta långt ifrån självklart i vilken riktning som ett samband går.
- När vi gör våra regressioner antar vi en orsaksriktning, men att sambandet är signifikant betyder inte att antagandet var korrekt.



Andra saker att vara uppmärksam på

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Regressionsanalysen testar inte orsaksriktningen

- Inom samhällsvetenskapen är det ofta långt ifrån självklart i vilken riktning som ett samband går.
- När vi gör våra regressioner antar vi en orsaksriktning, men att sambandet är signifikant betyder inte att antagandet var korrekt.
- Fallstudier, teoretiska resonemang och experiment kan hjälpa oss att reda ut orsaksriktningen.



Andra saker att vara uppmärksam på

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Har vi observationer från olika tidpunkter?

- När vi använder data från olika tidpunkter skapar det både möjligheter och problem.



Andra saker att vara uppmärksam på

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Har vi observationer från olika tidpunkter?

- När vi använder data från olika tidpunkter skapar det både möjligheter och problem.
- Å ena sidan ger det oss vissa möjligheter att studera orsaksriktningen (vad kommer först?).



Andra saker att vara uppmärksam på

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Har vi observationer från olika tidpunkter?

- När vi använder data från olika tidpunkter skapar det både möjligheter och problem.
- Å ena sidan ger det oss vissa möjligheter att studera orsaksriktningen (vad kommer först?).
- Vi kan också välja att endast studera variation över tid (t.ex. genom en dummyvariabel per land).



Andra saker att vara uppmärksam på

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Tillämpning
Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1
Exempel 2
Exempel 3

Kombinations- studier

Övrigt

Har vi observationer från olika tidpunkter?

- När vi använder data från olika tidpunkter skapar det både möjligheter och problem.
- Å ena sidan ger det oss vissa möjligheter att studera orsaksriktningen (vad kommer först?).
- Vi kan också välja att endast studera variation över tid (t.ex. genom en dummyvariabel per land).
- Å andra sidan framstår icke-signifikanta samband ofta som signifikanta om vi inte tar hänsyn till att Sverige 2012 och Sverige 2011 knappast är oberoende av varandra.



Andra saker att vara uppmärksam på

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipl regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Är den beroende variabeln dikotom?

- Vanlig linjär regression är dåligt lämpad för variabler som bara kan anta värdena 0 eller 1.



Andra saker att vara uppmärksam på

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Är den beroende variabeln dikotom?

- Vanlig linjär regression är dåligt lämpad för variabler som bara kan anta värdena 0 eller 1.
- Kan ge orimliga prediktioner ($\hat{y} < 0$ eller $\hat{y} > 1$) samt bryter mot en del andra antaganden (t.ex. homoskedasticitet).



Andra saker att vara uppmärksam på

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Är den beroende variabeln dikotom?

- Vanlig linjär regression är dåligt lämpad för variabler som bara kan anta värdena 0 eller 1.
- Kan ge orimliga prediktioner ($\hat{y} < 0$ eller $\hat{y} > 1$) samt bryter mot en del andra antaganden (t.ex. homoskedasticitet).
- Det vanligaste sättet att hantera detta är genom logistisk regression, men det är inte självklart bättre än OLS.



Andra saker att vara uppmärksam på

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Det är svårt att isolera för alla tänkbara förklaringar

- Vi vet inte vilka alla de möjliga bakomliggande förklaringarna är.



Andra saker att vara uppmärksam på

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Det är svårt att isolera för alla tänkbara förklaringar

- Vi vet inte vilka alla de möjliga bakomliggande förklaringarna är.
- Även om vi visste det, är det inte självklart hur vi ska mäta dem eller att det är praktiskt möjligt.



Andra saker att vara uppmärksam på

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Det är svårt att isolera för alla tänkbara förklaringar

- Vi vet inte vilka alla de möjliga bakomliggande förklaringarna är.
- Även om vi visste det, är det inte självklart hur vi ska mäta dem eller att det är praktiskt möjligt.
- Även om vi kände till och kunde mäta alla bakomliggande variabler, vet vi inte hur vi ska kontrollera för dem. Den linjära och additiva regressionsekvationen är bara en möjlighet.



Andra saker att vara uppmärksam på

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Det är svårt att isolera för alla tänkbara förklaringar

- Vi vet inte vilka alla de möjliga bakomliggande förklaringarna är.
- Även om vi visste det, är det inte självklart hur vi ska mäta dem eller att det är praktiskt möjligt.
- Även om vi kände till och kunde mäta alla bakomliggande variabler, vet vi inte hur vi ska kontrollera för dem. Den linjära och additiva regressionsekvationen är bara en möjlighet.
- Lösningen på det problemet får ni på fredag (studentföreläsning och experiment).



Avslutning

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Tillämpning

Kommentarer

Regressions- tabeller

Mjukvara

Multipel regression

Exempel 1

Exempel 2

Exempel 3

Kombinations- studier

Övrigt

Tack för att ni lyssnat!