



UPPSALA
UNIVERSITET

Föreläsning 7 och 8: Regressionsanalys

Pär Nyman

par.nyman@statsvet.uu.se

3 februari 2014



Regressionsanalys

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara

- 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

Regressions- tabeller

Multivariat regression

Isolera samband

Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

- 1 Repetition och passningsmått
- 2 Statistisk signifikans – från urval till population
- 3 Att läsa regressionstabeller
- 4 Multivariat regression
 - Spuriösa samband
 - Indirekta samband
- 5 Kombinationsstudier
- 6 Några saker att se upp för
- 7 Mjukvara



Regressionsekvationen (bivariat)

$$y = a + bx + e$$

y = Beroende variabel

a = Konstant eller intercept

b = Regressionskoefficient

x = Oberoende variabel

e = Felterm eller residual



Regressionsekvation 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 = Indexsiffror från observation 1 till observation n



Repetition och passningsmått

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



Multivariat 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



Multivariat 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



Multivariat regressionsekvation 3

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

y = Beroende variabel

a = Konstant eller intercept

b_1 = Regressionskoefficient 1

b_2 = Regressionskoefficient 2

Utbildning = Oberoende variabel 1

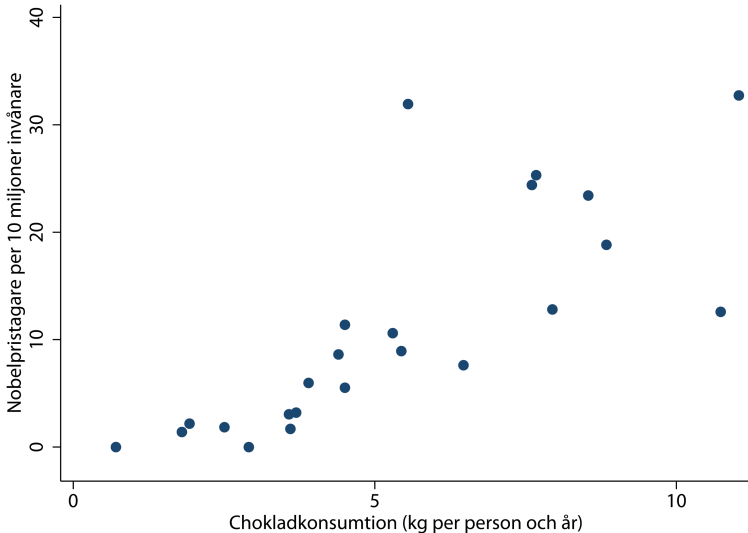
Kvinna = Oberoende variabel 2

e = Felterm eller residual



Repetition och passningsmått

$$y = a + bx + e$$



Regressions-
analys

Regressionskvationen

Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta
orsaksmekanism

Kombinations-
studier

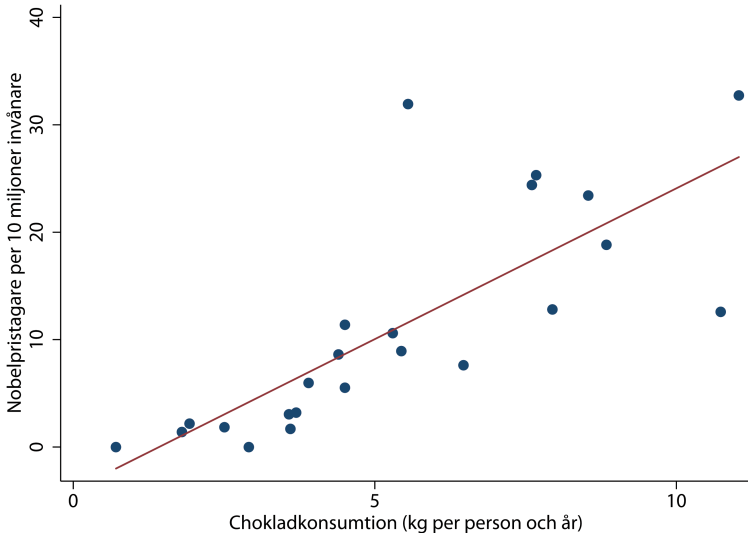
Pekpinnar

Mjukvara



Repetition och passningsmått

$$y = a + bx + e$$



Regressions-
analys

Regressionskvationen

Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta
orsaksmekanism

Kombinations-
studier

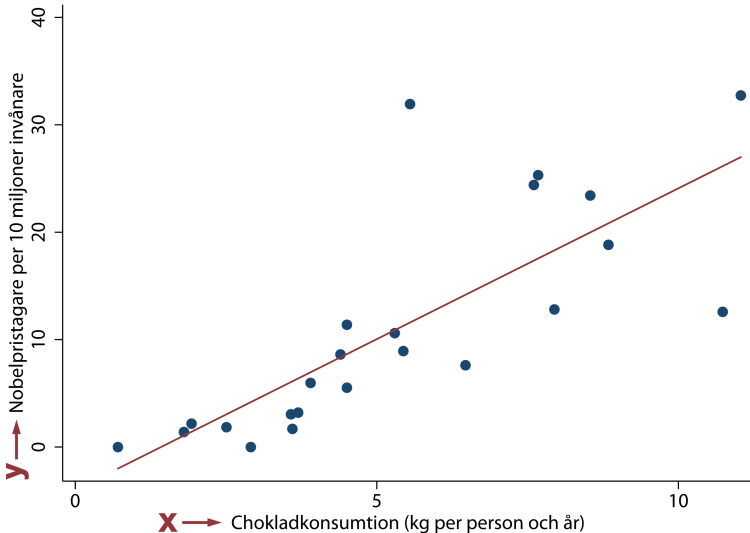
Pekpinnar

Mjukvara



Repetition och passningsmått

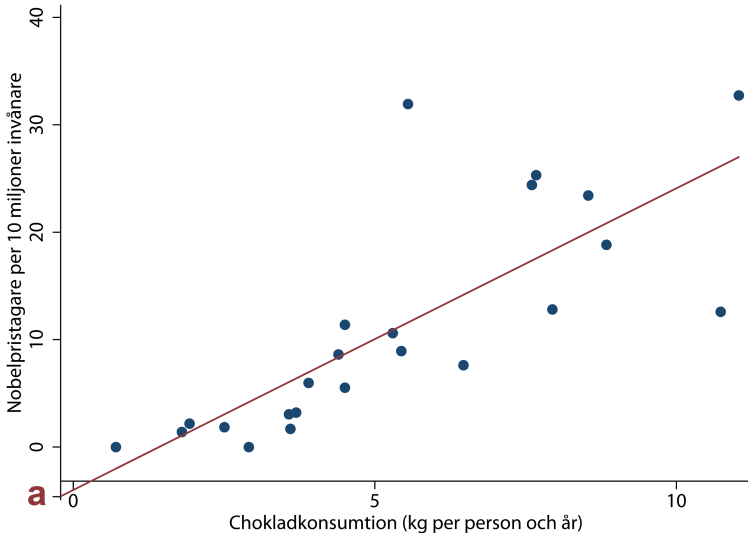
$$y = a + bx + e$$





Repetition och passningsmått

$$y = a + bx + e$$



Regressions-
analys

Regressionskvationen

Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta
orsaksmekanism

Kombinations-
studier

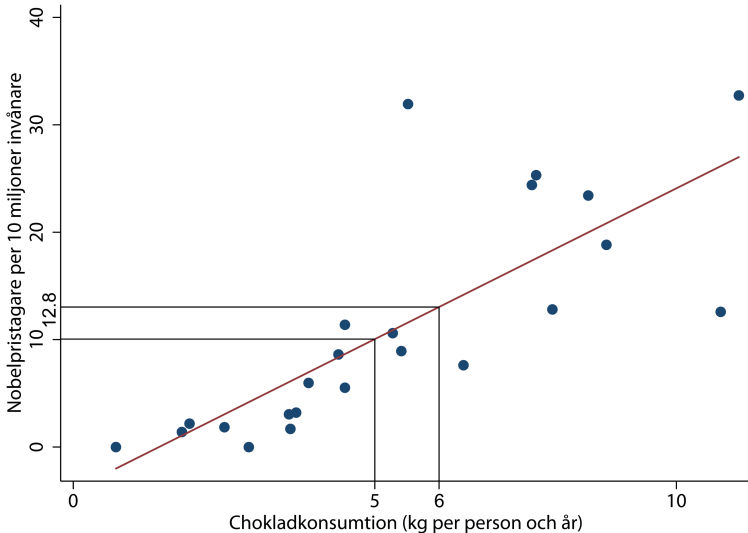
Pekpinnar

Mjukvara



Repetition och passningsmått

$$y = a + bx + e$$



Regressions-
analys

Regressionskvationen

Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta
orsaksmekanism

Kombinations-
studier

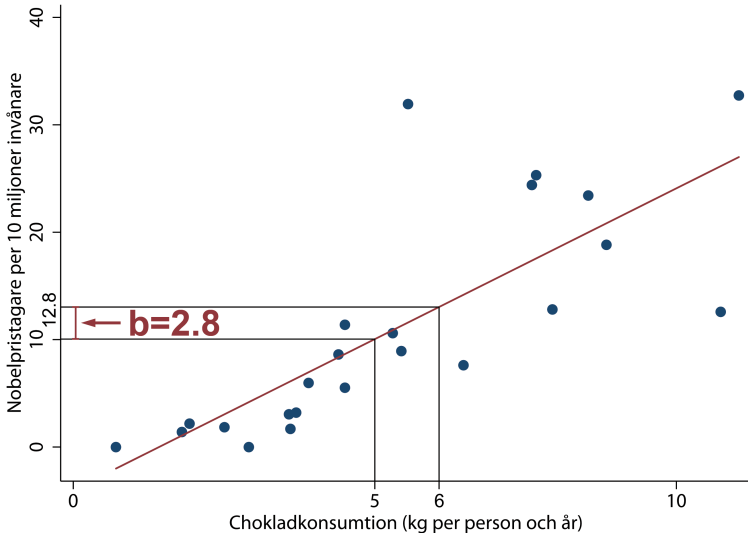
Pekpinnar

Mjukvara



Repetition och passningsmått

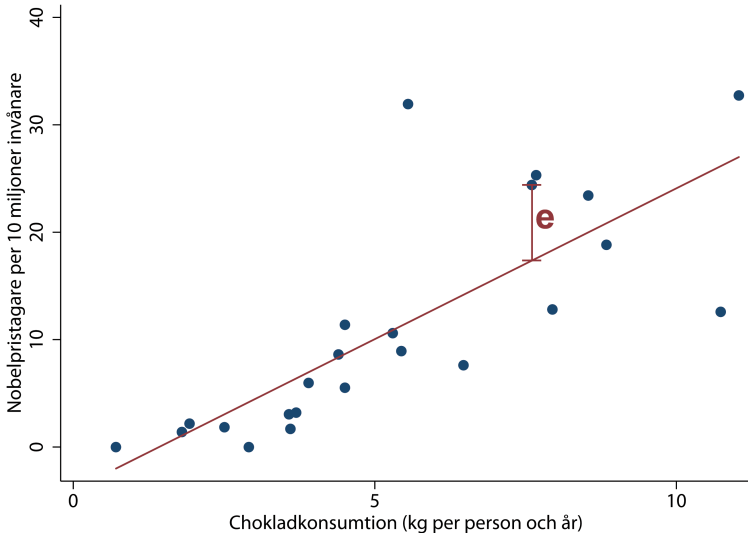
$$y = a + bx + e$$





Repetition och passningsmått

$$y = a + bx + e$$



Regressions-
analys

Regressionskvationen

Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta
orsaksmekanism

Kombinations-
studier

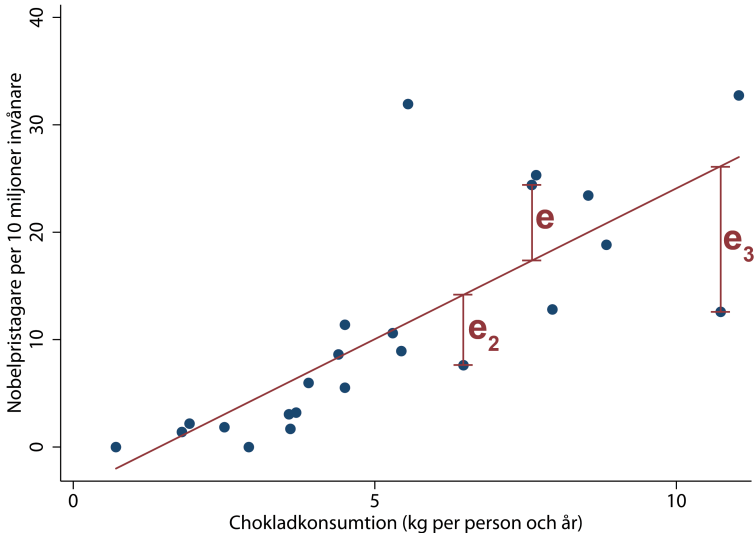
Pekpinnar

Mjukvara



Repetition och passningsmått

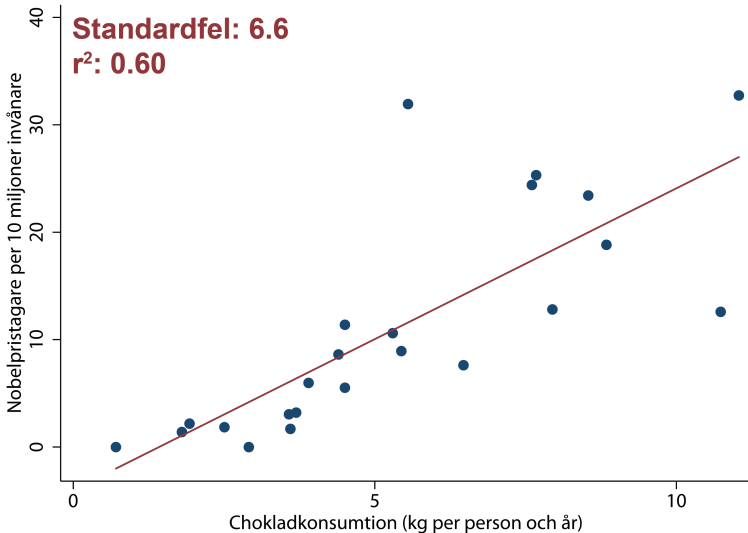
$$y = a + bx + e$$





Repetition och passningsmått

$$y = a + bx + e$$



Regressions-
analys

Regressions-
ekvationen

Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



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 avvikelser 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.



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)



Justerat R^2

- 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 när man gör en multivariat regression.
- 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.



Signifikans och generaliseringar

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

- Samma logik som när vi beräknade konfidensintervall runt proportioner.
- Kan vi generalisera resultaten till hela populationen?
- Kan sambandet bero på slumpen eller är den risken försumbar?
- I regel lika relevant vid totalundersökningar och oklara populationer.



Signifikans och generaliseringar

- Vi använder β för populationens regressionskoefficient för att skilja den från b som vi har använt för vår urvalskoefficient.



Signifikans och generaliseringar

- Vi använder β för populationens regressionskoefficient för att skilja den från b som vi har använt för vår urvalskoefficient.
- Om β är 0 finns det inget samband i populationen. Formellt kan vi ställa upp det såhär:

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$



Signifikans och generaliseringar

- Vi använder β för populationens regressionskoefficient för att skilja den från b som vi har använt för vår urvalskoefficient.
- Om β är 0 finns det inget samband i populationen. Formellt kan vi ställa upp det såhär:

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

- Vi testar dessa hypoteser genom att beräkna konfidensintervall runt b .



Signifikans och generaliseringar

- Vi använder β för populationens regressionskoefficient för att skilja den från b som vi har använt för vår urvalskoefficient.
- Om β är 0 finns det inget samband i populationen. Formellt kan vi ställa upp det såhär:

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

- Vi testar dessa hypoteser genom att beräkna konfidensintervall runt b .
- Om konfidensintervallet inte täcker in 0 kan vi förkasta nollhypotesen. Vi vågar då dra slutsatsen att det finns ett samband även i populationen och vi kallar detta för att sambandet är statistiskt signifikant.



Signifikans och generaliseringar

- Vi beräknar felmarginal och konfidensintervall på ungefär samma sätt som för proportioner.

$$\text{Konfidensintervall} = b \pm t_{kv} \times se(b)$$

Regressions-
analys

Regressionskvantilen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



Signifikans och generaliseringar

- Vi beräknar felmarginal och konfidensintervall på ungefär samma sätt som för proportioner.

$$\text{Konfidensintervall} = b \pm t_{kv} \times se(b)$$

- Statistikprogrammen beräknar ett testvärde (t-värde).

$$t = \frac{b}{se(b)}$$



Signifikans och generaliseringar

- Vi beräknar felmarginal och konfidensintervall på ungefär samma sätt som för proportioner.

$$\text{Konfidensintervall} = b \pm t_{kv} \times se(b)$$

- Statistikprogrammen beräknar ett testvärde (t-värde).

$$t = \frac{b}{se(b)}$$

- Detta t-värde anger vid vilket kritiskt värde som vårt samband upphör att vara signifikant (då den ena änden av konfidensintervallet tangerar 0).



Signifikans och generaliseringar

- Vi beräknar felmarginal och konfidensintervall på ungefär samma sätt som för proportioner.

$$\text{Konfidensintervall} = b \pm t_{kv} \times se(b)$$

- Statistikprogrammen beräknar ett testvärde (t-värde).

$$t = \frac{b}{se(b)}$$

- Detta t-värde anger vid vilket kritiskt värde som vårt samband upphör att vara signifikant (då den ena änden av konfidensintervallet tangerar 0).
- Vid stora urval är det kritiska värdena t_{90} , t_{95} och t_{99} 1,65, 1,96 och 2,58, precis som de z-värden vi tidigare använt, men vid små urval är de något större.



Signifikans och generaliseringar

- Vi beräknar felmarginal och konfidensintervall på ungefär samma sätt som för proportioner.

$$\text{Konfidensintervall} = b \pm t_{kv} \times se(b)$$

- Statistikprogrammen beräknar ett testvärde (t-värde).

$$t = \frac{b}{se(b)}$$

- Detta t-värde anger vid vilket kritiskt värde som vårt samband upphör att vara signifikant (då den ena änden av konfidensintervallet tangerar 0).
- Vid stora urval är det kritiska värdena t_{90} , t_{95} och t_{99} 1,65, 1,96 och 2,58, precis som de z-värden vi tidigare använt, men vid små urval är de något större.
- Enkel tumregel: Sambandet är signifikant vid 95 procents säkerhetsnivå om regressionskoefficienten är minst dubbelt så stor som dess standardavvikelse ($t_{kv} \sim 2,00$).



Signifikans och generaliseringar

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

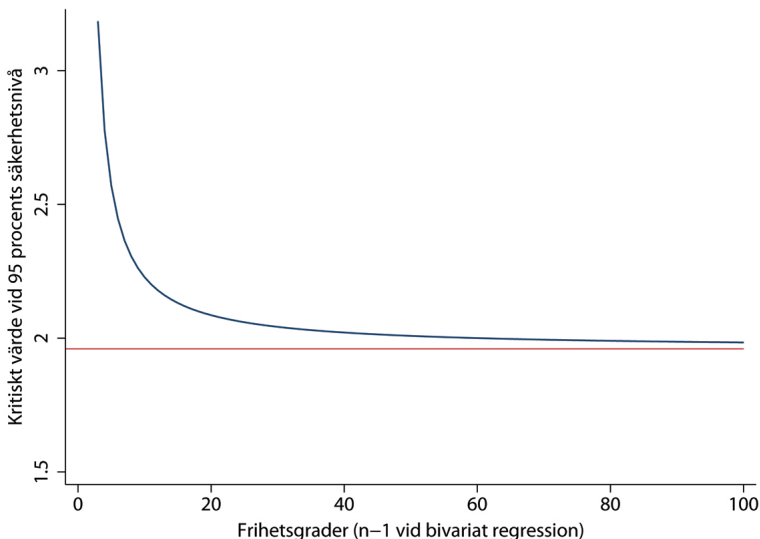
Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara





Signifikans och generaliseringar

- För att testa om ett resultat är signifikant kan vi därför jämföra detta t-värde med det kritiska värdet för den säkerhetsnivå vi har valt. Om testvärdet ligger utanför intervallet mellan $-t_{kv}$ och t_{kv} säger vi att sambandet är signifikant.



Signifikans och generaliseringar

- För att testa om ett resultat är signifikant kan vi därför jämföra detta t-värde med det kritiska värdet för den säkerhetsnivå vi har valt. Om testvärdet ligger utanför intervallet mellan $-t_{kv}$ och t_{kv} säger vi att sambandet är signifikant.
- Vi ska göra det på tre exempel från förra föreläsningen:



Signifikans och generaliseringar

- För att testa om ett resultat är signifikant kan vi därför jämföra detta t-värde med det kritiska värdet för den säkerhetsnivå vi har valt. Om testvärdet ligger utanför intervallet mellan $-t_{kv}$ och t_{kv} säger vi att sambandet är signifikant.
- Vi ska göra det på tre exempel från förra föreläsningen:
 - 1 Sambandet mellan placering på vänster-högerskalan och inställning till jämställdhet.



Signifikans och generaliseringar

- För att testa om ett resultat är signifikant kan vi därför jämföra detta t-värde med det kritiska värdet för den säkerhetsnivå vi har valt. Om testvärdet ligger utanför intervallet mellan $-t_{kv}$ och t_{kv} säger vi att sambandet är signifikant.
- Vi ska göra det på tre exempel från förra föreläsningen:
 - 1 Sambandet mellan placering på vänster-högerskalan och inställning till jämställdhet.
 - 2 Sambandet mellan utbildning och inställning till EU.



Signifikans och generaliseringar

- För att testa om ett resultat är signifikant kan vi därför jämföra detta t-värde med det kritiska värdet för den säkerhetsnivå vi har valt. Om testvärdet ligger utanför intervallet mellan $-t_{kv}$ och t_{kv} säger vi att sambandet är signifikant.
- Vi ska göra det på tre exempel från förra föreläsningen:
 - 1 Sambandet mellan placering på vänster-högerskalan och inställning till jämställdhet.
 - 2 Sambandet mellan utbildning och inställning till EU.
 - 3 Sambandet mellan regeringssätt och korruption.



Signifikans och generaliseringar

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

Variabel	Kodning
Kön	0 (man) eller 1 (kvinna)
Utbildning	Antal år
Inkomst	Månadslön i tusentals kr
Vänster-högerskala	0 (mkt långt t. höger) - 10 (mkt långt t. vänster)
Ökad jämställdhet	0 (mycket litet behov) - 10 (mycket stort behov)
Inställning till EU	0 (mycket negativ) - 10 (mycket positiv)
Regeringssätt	0 (parlamentarism) eller 1 (presidentialism)
Korruption	0 (ingen korruption) - 10 (mycket korruption)
Brittisk koloni	1 (brittisk koloni) eller 0 (ej brittisk koloni)

Tips: Döp gärna dikotoma variabler efter hur de är kodade. Kvinna, presidentialism och brittisk koloni har en underförstådd kodning medan kön, regeringssätt och kolonial bakgrund kan vara kodade hur som helst.



Signifikans och generaliseringar

Exempel 1: Placering på vänster–höger-skalan och inställning till jämställdhet

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

$$b = 0,48$$

$$se(b) = 0,16$$

$$t = 3 (0,48/0,16)$$



Signifikans och generaliseringar

Exempel 1: Placering på vänster–höger-skalan och inställning till jämställdhet

- När vi genomför den bivariata regressionsanalysen erhåller vi följande värden:
 $b = 0,48$
 $se(b) = 0,16$
 $t = 3 (0,48/0,16)$
- Det kritiska t-värdet på 95 procents säkerhetsnivå är här 1,96 (n=2000). Eftersom t-värdet ligger utanför det kritiska intervallet kan vi förkasta nollhypotesen (3 är större än 1,96).



Signifikans och generaliseringar

Exempel 1: Placering på vänster–höger-skalan och inställning till jämställdhet

Regressions- analys

Regressions-
ekvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

- När vi genomför den bivariata regressionsanalysen erhåller vi följande värden:
 $b = 0,48$
 $se(b) = 0,16$
 $t = 3 (0,48/0,16)$
- Det kritiska t-värdet på 95 procents säkerhetsnivå är här 1,96 (n=2000). Eftersom t-värdet ligger utanför det kritiska intervallet kan vi förkasta nollhypotesen (3 är större än 1,96).
- Effekten av ideologisk placering på inställningen till behovet av ökad jämställdhet mellan könen är således statistiskt signifikant på 95-procents säkerhetsnivå.



Signifikans och generaliseringar

Exempel 2: Utbildning och inställning till EU

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara

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

$$b = 0,25,$$

$$se(b) = 0,11$$

$$t = 2,27 (0,25/0,11)$$



Signifikans och generaliseringar

Exempel 2: Utbildning och inställning till EU

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara

- När vi genomför den bivariata regressionsanalysen erhåller vi följande värden:
 $b = 0,25,$
 $se(b) = 0,11$
 $t = 2,27 (0,25/0,11)$
- Det kritiska t-värdet på 95 procents säkerhetsnivå är här 1,96 ($n=2000$). Då t-värdet ligger utanför det kritiska intervallet kan vi förkasta nollhypotesen (2,27 är större än 1,96).



Signifikans och generaliseringar

Exempel 2: Utbildning och inställning till EU

Regressions- analys

Regressions-
ekvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

- När vi genomför den bivariata regressionsanalysen erhåller vi följande värden:
 $b = 0,25,$
 $se(b) = 0,11$
 $t = 2,27 (0,25/0,11)$
- Det kritiska t-värdet på 95 procents säkerhetsnivå är här 1,96 ($n=2000$). Då t-värdet ligger utanför det kritiska intervallet kan vi förkasta nollhypotesen (2,27 är större än 1,96).
- Det kritiska t-värdet på 95 procents säkerhetsnivå är här 2,58. Om vi höjer säkerhetsnivån till 99 procent är sambandet inte längre signifikant eftersom t-värdet då ligger inom det kritiska intervallet ($2,27 < 2,58$).



Signifikans och generaliseringar

Exempel 3: Regeringssätt och korruption

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

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

$$b = -1,15$$

$$se(b) = 0,98$$

$$t = -1,17 \quad (-1,15/0,98)$$



Signifikans och generaliseringar

Exempel 3: Regeringssätt och korruption

Regressions- analys

Regressionskvantiteten
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

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

$$b = -1,15$$

$$se(b) = 0,98$$

$$t = -1,17 \quad (-1,15/0,98)$$

- Det kritiska t-värdet på 95 procents säkerhetsnivå är här 2,00 (n=60). Vi finner därmed att vi inte kan förkasta nollhypotesen att det inte finns något samband i populationen då t-värdet (-1,17) ligger inom det kritiska intervallet (-1,17 är större än -2 men mindre än 2).



Signifikans och generaliseringar

Exempel 3: Regeringssätt och korruption

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

- När vi genomför den bivariata regressionsanalysen erhåller vi följande värden:
 $b = -1,15$
 $se(b) = 0,98$
 $t = -1,17 (-1,15/0,98)$
- Det kritiska t-värdet på 95 procents säkerhetsnivå är här 2,00 ($n=60$). Vi finner därmed att vi inte kan förkasta nollhypotesen att det inte finns något samband i populationen då t-värdet (-1,17) ligger inom det kritiska intervallet (-1,17 är större än -2 men mindre än 2).
- Effekten av regeringssätt på korruptionsgraden är således inte statistisk signifikant.



Signifikans och generaliseringar

Fem ekvivalenta metoder. Sambandet är signifikant om:

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



Signifikans och generaliseringar

Fem ekvivalenta metoder. Sambandet är signifikant om:

- T-värdet är högre än det kritiska t-värdet.
- Regressionskoefficienten är mer än t_{kv} gånger så stor som koefficientens standardfel. Givet approximationen $t_{kv} = 2$, är det samma sak som att koefficienten är minst dubbelt så stor som dess standardfel.



Signifikans och generaliseringar

Fem ekvivalenta metoder. Sambandet är signifikant om:

- T-värdet är högre än det kritiska t-värdet.
- Regressionskoefficienten är mer än t_{kv} gånger så stor som koefficientens standardfel. Givet approximationen $t_{kv} = 2$, är det samma sak som att koefficienten är minst dubbelt så stor som dess standardfel.
- Konfidensintervallet runt koefficienten omsluter inte värdet 0.



Signifikans och generaliseringar

Fem ekvivalenta metoder. Sambandet är signifikant om:

- T-värdet är högre än det kritiska t-värdet.
- Regressionskoefficienten är mer än t_{kv} gånger så stor som koefficientens standardfel. Givet approximationen $t_{kv} = 2$, är det samma sak som att koefficienten är minst dubbelt så stor som dess standardfel.
- Konfidensintervallet runt koefficienten omsluter inte värdet 0.
- Det står asterisk(er) efter efter regressionskoefficienten. Läs under tabellen för att se vilken säkerhetsnivå de motsvarar. Det är den vanligaste metoden när man läser en regressionstabell.



Signifikans och generaliseringar

Fem ekvivalenta metoder. Sambandet är signifikant om:

- T-värdet är högre än det kritiska t-värdet.
- Regressionskoefficienten är mer än t_{kv} gånger så stor som koefficientens standardfel. Givet approximationen $t_{kv} = 2$, är det samma sak som att koefficienten är minst dubbelt så stor som dess standardfel.
- Konfidensintervallet runt koefficienten omsluter inte värdet 0.
- Det står asterisk(er) efter 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. Heter Significance i PSPP.



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

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



Regressionstabeller

	(1)	(2)
Choklad	2,81***	2,28***
	(0,50)	(0,64)
BNP/capita		0,20
		(0,16)
Konstant	- 3,99	- 8,42*
	(3,00)	(4,58)
Observationer	23	23
Standardfel	6,60	6,51
R²	0,60	0,63

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

Variabler

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



Regressionstabeller

	(1)	(2)
Choklad	2,81***	2,28***
	(0,50)	(0,64)
BNP/capita		0,20
		(0,16)
Konstant	- 3,99	- 8,42*
	(3,00)	(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

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



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

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



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

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



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*
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

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



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
signifikansRegressions-
tabellerMultivariat
regression

Isolera samband

Hitta
orsaksmekanismKombinations-
studier

Pekpinnar

Mjukvara



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.



Output från PSPP

Model Summary

<i>R</i>	<i>R Square</i>	<i>Adjusted R Square</i>	<i>Std. Error of the Estimate</i>
.79	.63	.61	6.51

Regressions-
analys

Regressionskvationen

Passningsmått

Statistisk
signifikans

ANOVA

	<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>Significance</i>
<i>Regression</i>	1451.20	2	725.60	17.12	.00
<i>Residual</i>	847.57	20	42.38		
<i>Total</i>	2298.77	22			

Regressions-
tabellerMultivariat
regression

Isolera samband

Hitta
orsaksmekanismKombinations-
studier

Pekpinnar

Mjukvara

Coefficients

	<i>B</i>	<i>Std. Error</i>	<i>Beta</i>	<i>t</i>	<i>Significance</i>
<i>(Constant)</i>	-8.42	4.58	.00	-1.84	.08
chocolate	2.28	.64	.63	3.55	.00
gdp	.20	.16	.23	1.27	.22



Multivariat regression

- Fram tills nu har vi bara studerat samband mellan två variabler (bivariat regression). Multivariat regression innebär att vi har fler än två variabler.
- Det finns tre huvudsakliga skäl för att använda fler än en oberoende variabel:
 - 1 Förbättra förklaringen
 - 2 Isolera sambandet
 - 3 Hitta orsaksmekanism



Multivariat regression

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

**Multivariat
regression**

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara

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.



Multivariat regression

1 Förbättra förklaringen med fler variabler

- Få samhällsfenomen kan förklaras av en enda variabel. Såhär kan vi skriva en modell med två oberoende variabler:

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

- Om även x_2 har ett samband med y kommer vi med denna modell att kunna förklara en större del av variationen i y och göra bättre prediktioner. Avgörande för att förbättra passningen.
- 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.



2 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.



Multivariat regression

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

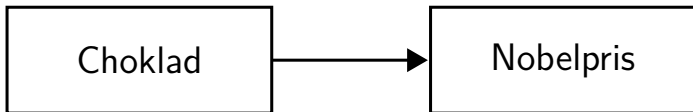
Hitta
orsaksmechanism

Kombinations-
studier

Pekpinnar

Mjukvara

Ett bivariat samband...





Multivariat regression

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

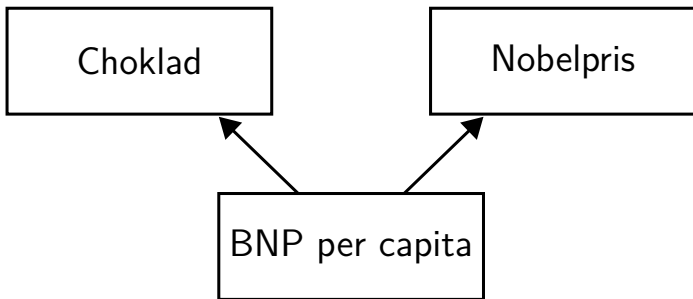
Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara

...som i själva verket är spuriöst?





Multivariat regression

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband

Hitta
orsaksmechanism

Kombinations- studier

Pekpinnar

Mjukvara

- Vad händer med det bivariata sambandet mellan ideologisk position och inställning till jämställdhet när vi kontrollerar för kön?



Multivariat regression

Regressions-
analys

Regressionskvationen

Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara

Beroende variabel: Inställning till jämställdhet

	(1)	(2)
Vänster-höger	0,48 (3,00)	0,03 (0,78)
Kön		1,78 (3,18)
Konstant	2,97	5,14
Observationer	2000	2000

t-värden i parenteser.



Multivariat regression

Beroende variabel: Inställning till jämställdhet

	(1)	(2)
Vänster-höger	0,48*** (3,00)	0,03 (0,78)
Kön		1,78*** (3,18)
Konstant	2,97***	5,14***
Observationer	2000	2000

t-värden i parenteser.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$



Multivariat regression

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband

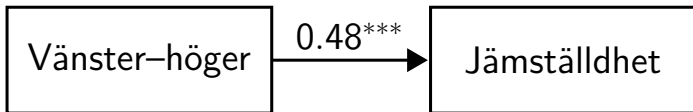
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

Ett bivariat samband...





Multivariat regression

Regressions-
analys

Regressions-
ekvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

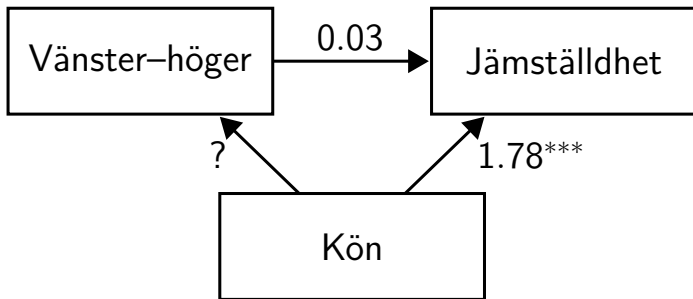
Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara

...som var spuriöst.





Multivariat regression

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta
orsaksmechanism

Kombinations-
studier

Pekpinnar

Mjukvara

Beroende variabel: Placering på vänster-höger-skalan

	(1)
Kön	1,27 (2,94)
Konstant	4,14
Observationer	2000

t-värden i parenteser.



Multivariat regression

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

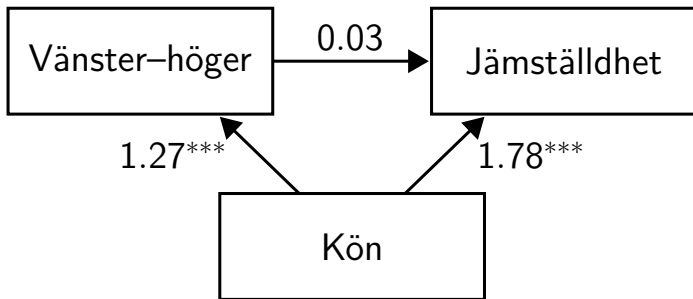
Isolera samband
Hitta
orsaksmechanism

Kombinations-
studier

Pekpinnar

Mjukvara

Vi kan nu ange effekten för varje orsakspil.





Multivariat regression

- Precis som i det bivariata fallet kan vi beräkna förväntade värden för multivariata regressionslikningar. Anta att ni blir tillfrågade om vilken inställning till jämställdhet ni tror att en kvinna med ideologisk placering 5 har.

$$\hat{y} = a + b_1x_1 + b_2x_2$$

$$\text{Jämställdhet} = 5,14 + 0,03 \times 5 + 1,78 \times 1 = 7,07$$

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



Multivariat regression

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband

Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

- Vad händer med det bivariata sambandet mellan utbildning och inställning till EU när vi kontrollerar för kön?



Multivariat regression

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara

Beroende variabel: Inställning till EU

	(1)	(2)
Utbildning	0,25 (2,27)	0,24 (2,28)
Kön		-0,85 (-2,25)
Konstant	2,02	2,14
Observationer	2000	2000

t-värden i parenteser.



Multivariat regression

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

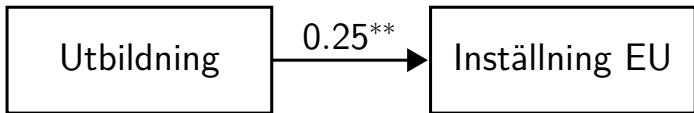
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara

Ett bivariat samband...





Multivariat regression

Regressions-
analys

Regressions-
ekvationen
Passningsmätt

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

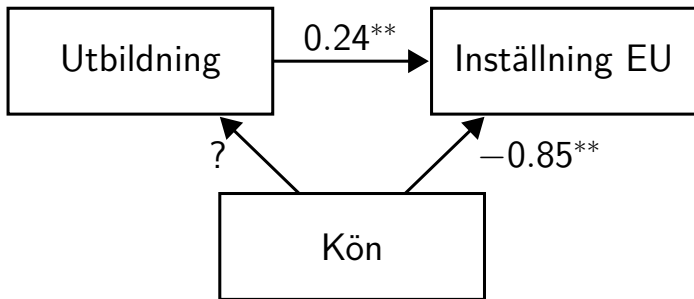
Isolera samband
Hitta
orsaksmechanism

Kombinations-
studier

Pekpinnar

Mjukvara

...som kvarstår vid kontroll för kön.





Multivariat regression

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband

Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

- Vad händer med det bivariata sambandet mellan presidentialism (regeringssätt) och korruption när vi kontrollerar för om landet har varit en brittisk koloni?



Multivariat regression

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara

Beroende variabel: Korrupsion

	(1)	(2)
Presidentialism	-1,15 (0,98)	-1,89 (0,72)
Brittisk koloni		-1,10 (-0,43)
Konstant	7,18	8,10
Observationer	60	60

Standardfel i parenteser.



Multivariat regression

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

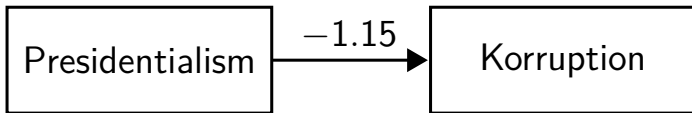
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara

Ett bivariat samband...





Multivariat regression

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

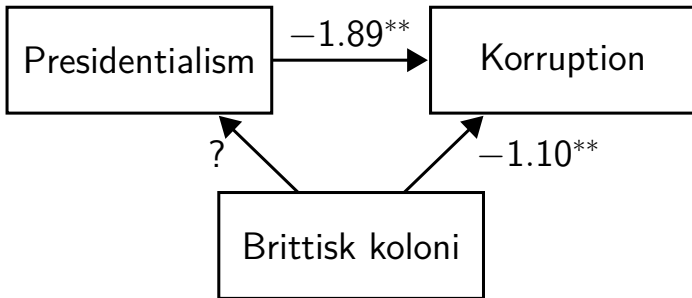
Hitta
orsaksmechanism

Kombinations-
studier

Pekpinnar

Mjukvara

...som förstärks när vi adderar en tredje variabel..





Multivariat regression

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband

Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

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



Multivariat regression

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband

Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

- 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.



Multivariat regression

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband

Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

- 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.
- Tidigare brittiska kolonier har i regel parlamentariska system men har tack vare sin bakgrund andra institutioner som motverkar korruption. När vi kontrollerar för brittisk kolonialbakgrund försvinner denna del från sambandet mellan presidentialism och korruption.



Multivariat regression

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara

3 Hitta orsaksmekanismen

- En variabel som påverkas av den oberoende variabeln samt påverkar den beroende variabeln kallas för mellanliggande variabel.



3 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.



3 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.



Multivariat regression

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara

3 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.



Multivariat regression

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband

Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

- Vad händer med det bivariata sambandet mellan kön och ideologisk position när vi adderar inkomst till modellen?
- Hint: Det är inte sannolikt att kön påverkas av en bakomliggande variabel. Vi är alltså inte ute efter att isolera sambandet, utan att hitta en orsaksmekanism. Varför står kvinnor till vänster om män?



Multivariat regression

Regressions-
analys

Regressionskvationen

Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara

Beroende variabel: Placering på vänster-höger-skalan

	(1)	(2)
Kön	1,27 (2,94)	-0,05 (-0,83)
Inkomst		-0,23 (-4,27)
Konstant	4,14	8,14
Observationer	2000	2000

t-värden i parenteser.



Multivariat regression

Regressions-
analys

Regressionskvationen

Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

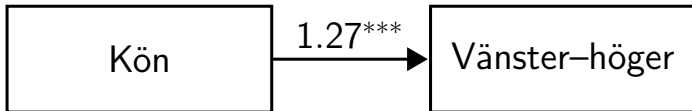
Hitta
orsaksmechanism

Kombinations-
studier

Pekpinnar

Mjukvara

Ett bivariat samband...





Multivariat regression

Regressions-
analys

Regressionskvationen

Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

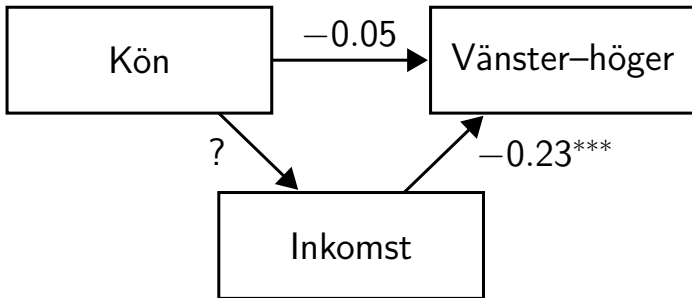
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara

...som visar sig vara indirekt.





Multivariat regression

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband

Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

- 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.
- 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.



Multivariat regression

Överkurs

$$\textit{Total effekt} = \textit{Direkt effekt} + \textit{Indirekt effekt}$$

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



Multivariat regression

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara

Överkurs

$$\text{Total effekt} = \text{Direkt effekt} + \text{Indirekt effekt}$$

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).

$$1,27 = -0,05 + b \times -0,23$$



Multivariat regression

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara

Överkurs

$$\text{Total effekt} = \text{Direkt effekt} + \text{Indirekt effekt}$$

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).

$$1,27 = -0,05 + b \times -0,23$$

Om vi adderar 0,05 till båda sidorna...

$$1,32 = b \times -0,23$$



Multivariat regression

Regressions-
analys

Regressions-
ekvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta
orsaksmeکانism

Kombinations-
studier

Pekpinnar

Mjukvara

Överkurs

$$\text{Total effekt} = \text{Direkt effekt} + \text{Indirekt effekt}$$

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).

$$1,27 = -0,05 + b \times -0,23$$

Om vi adderar 0,05 till båda sidorna...

$$1,32 = b \times -0,23$$

...och dividerar med -0,23 kan vi lösa ut b .

$$b = -5,74$$



Multivariat regression

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

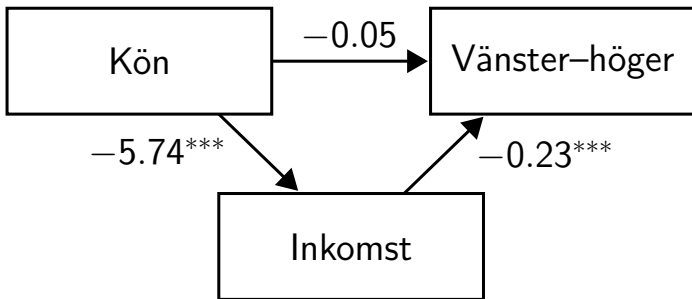
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara

Vi kan nu ange effekten för samtliga orsakspilar.





Multivariat regression

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband

Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

- Vi fann tidigare att sambandet mellan utbildning och inställning till EU var signifikant även när vi kontrollerade för den bakomliggande variabeln kön. Men vad kan förklara att högutbildade är mer positiva till EU än högutbildade?
- Vi testar att lägga till inkomst respektive ideologisk position till modellen.



Multivariat regression

Beroende variabel: Inställning till EU

	(1)	(2)	(3)
Utbildning	0,25 (0,11)	0,20 (0,10)	0,08 (0,09)
Inkomst		0,11 (0,06)	
Vänster-höger			-1,14 (-0,27)
Konstant	2,02	2,14	8,14
Observationer	2000	2000	2000

Standardfel i parenteser.



Multivariat regression

Regressions-
analys

Regressions-
ekvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

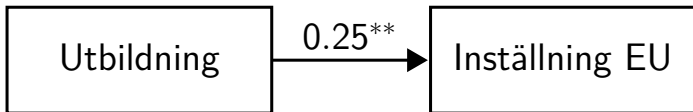
Hitta
orsaksme-
kanism

Kombinations-
studier

Pekpinnar

Mjukvara

Ett bivariat samband...





Multivariat regression

Regressions-
analys

Regressionskvationen

Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

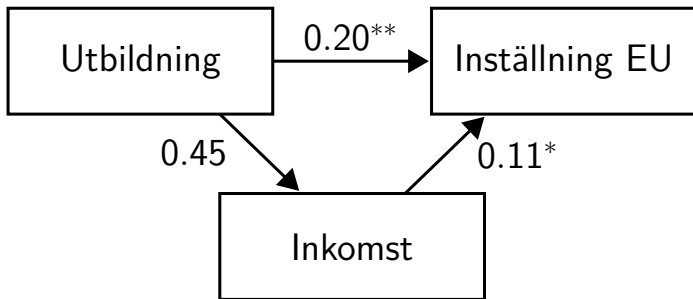
Isolera samband

Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara





Multivariat regression

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

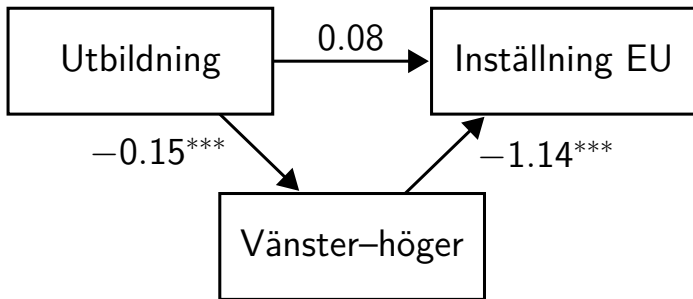
Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara





Multivariat regression

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

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



Multivariat regression

- Om det ursprungliga sambandet är signifikant även efter att vi kontrollerat för en tredje variabel har vi isolerat sambandet (från den variabeln).
- Om det ursprungliga sambandet försvann eller försvagades efter att vi kontrollerat för en *mellanliggande* variabel var det indirekt. Vi har specificerat en orsaksmekanism och hypotesen får stöd.

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



Multivariat regression

- Om det ursprungliga sambandet är signifikant även efter att vi kontrollerat för en tredje variabel har vi isolerat sambandet (från den variabeln).
- Om det ursprungliga sambandet försvann eller försvagades efter att vi kontrollerat för en *mellanliggande* variabel var det indirekt. Vi har specificerat en orsaksmekanism och hypotesen får stöd.
- Om det ursprungliga sambandet försvann eller försvagades efter att vi kontrollerat för en *bakomliggande* variabel var det ett skensamband och spuriöst. Hypotesen får *ej* stöd.

Regressions-
analys

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



Multivariat regression

- Om det ursprungliga sambandet är signifikant även efter att vi kontrollerat för en tredje variabel har vi isolerat sambandet (från den variabeln).
- Om det ursprungliga sambandet försvann eller försvagades efter att vi kontrollerat för en *mellanliggande* variabel var det indirekt. Vi har specificerat en orsaksmekanism och hypotesen får stöd.
- Om det ursprungliga sambandet försvann eller försvagades efter att vi kontrollerat för en *bakomliggande* variabel var det ett skensamband och spuriöst. Hypotesen får *ej* stöd.
- Om en variabel är bakomliggande eller mellanliggande, alltså orsaksriktningen på ett samband, är ofta svårare att avgöra än i dagens exempel.



Multivariat regression

Regressions- analys

Regressions-
ekvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

- 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

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta

orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



Kombinationsstudier

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

- 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

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

- 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

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



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

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



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

Regressionskvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



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

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



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).

Regressions-
analys

Regressions-
ekvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



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

Regressions- tabeller

Multivariat regression

Isolera samband

Hitta

orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara



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

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara



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

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



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

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



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

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



Kombinationsstudier

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

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara



Kombinationsstudier

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

- 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

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

- 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

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



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

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



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

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



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.



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

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



Kombinationsstudier

Avvikande fall

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

Regressions-
analys

Regressionskvationen

Passningsmått

Statistisk
signifikans

Regressions-
tabeller

Multivariat
regression

Isolera samband

Hitta

orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



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

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



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

Regressions-
tabeller

Multivariat
regression

Isolera samband
Hitta
orsaksmekanism

Kombinations-
studier

Pekpinnar

Mjukvara



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

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

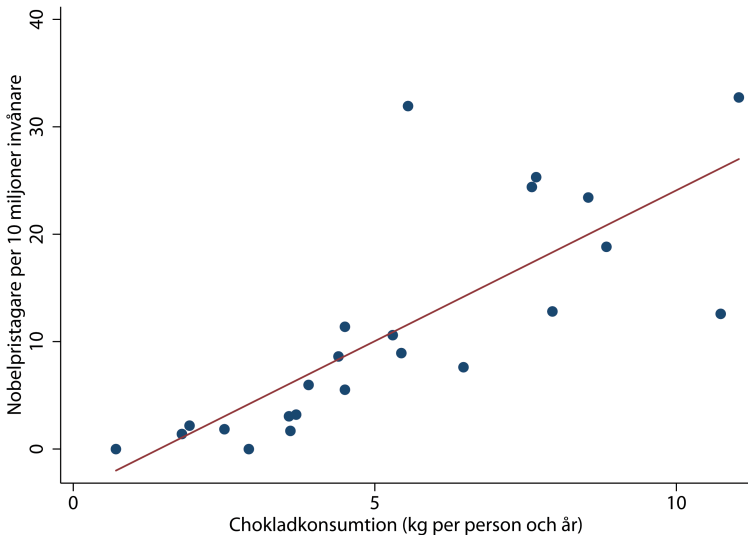
Multivariat regression

Isolera samband
Hitta
orsaksmechanism

Kombinations- studier

Pekpinnar

Mjukvara





Kombinationsstudier

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

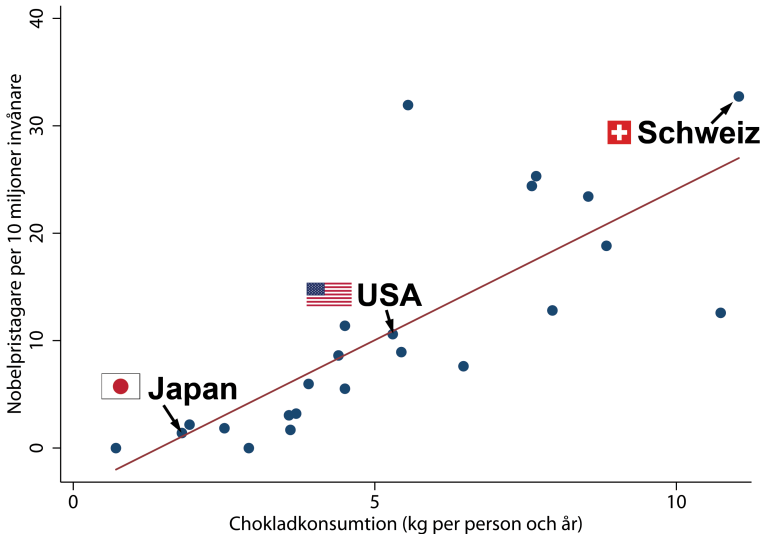
Multivariat regression

Isolera samband
Hitta
orsaksmechanism

Kombinations- studier

Pekpinnar

Mjukvara





Kombinationsstudier

Regressions-
analys

Regressions-
ekvationen
Passningsmått

Statistisk
signifikans

Regressions-
tabeller

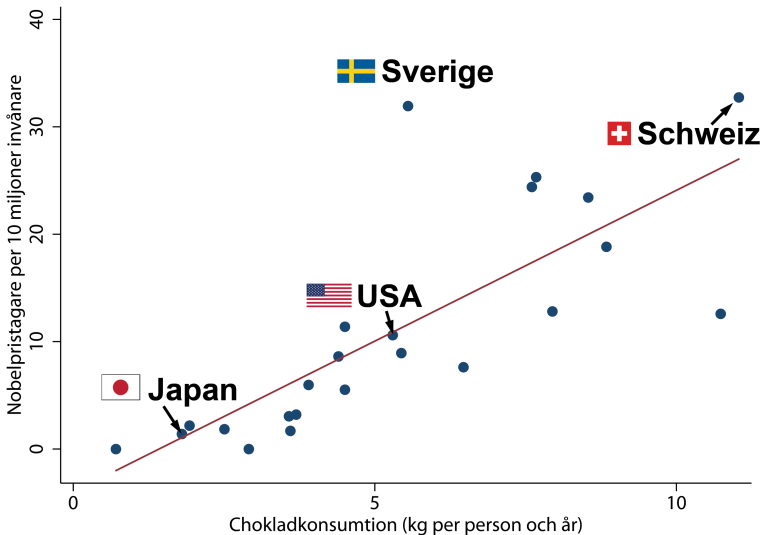
Multivariat
regression

Isolera samband
Hitta
orsaksmechanism

Kombinations-
studier

Pekpinnar

Mjukvara





Några saker att vara uppmärksam på

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

Det är svårt att jämföra koefficienter

- Variabler är ofta mätta på olika skalor.



Några saker att vara uppmärksam på

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

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.



Några saker att vara uppmärksam på

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

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.



Några saker att vara uppmärksam på

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

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!



Några saker att vara uppmärksam på

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

Regressionsanalysen testar inte orsaksriktningen

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



Några saker att vara uppmärksam på

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband

Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

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.



Några saker att vara uppmärksam på

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband

Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

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.



Några saker att vara uppmärksam på

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband

Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

Har vi observationer från olika tidpunkter?

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



Några saker att vara uppmärksam på

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

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?).



Några saker att vara uppmärksam på

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

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).



Några saker att vara uppmärksam på

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

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.



Några saker att vara uppmärksam på

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

Är den beroende variabeln dikotom?

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



Några saker att vara uppmärksam på

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

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).



Några saker att vara uppmärksam på

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

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.



Några saker att vara uppmärksam på

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband

Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

Det är svårt att isolera för alla tänkbara förklaringar

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



Några saker att vara uppmärksam på

Regressions- analys

Regressionskvationen

Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband

Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

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.



Några saker att vara uppmärksam på

Regressions- analys

Regressionsekvationen

Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband

Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

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.



Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

- De flesta på institutionen använder **Stata**. Det är stort i forskarvärlden men inte lika vanligt utanför akademien.



Mjukvara

Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

- De flesta på institutionen använder **Stata**. Det är stort i forskarvärlden men inte lika vanligt utanför akademien.
- **R** är ett avancerat gratisprogram som växer snabbt.



- De flesta på institutionen använder **Stata**. Det är stort i forskarvärlden men inte lika vanligt utanför akademien.
- **R** är ett avancerat gratisprogram som växer snabbt.
- Vi tror att **SPSS** är det vanligaste programmet på svenska myndigheter.



Regressions- analys

Regressionskvationen
Passningsmått

Statistisk signifikans

Regressions- tabeller

Multivariat regression

Isolera samband
Hitta
orsaksmekanism

Kombinations- studier

Pekpinnar

Mjukvara

- De flesta på institutionen använder **Stata**. Det är stort i forskarvärlden men inte lika vanligt utanför akademien.
- **R** är ett avancerat gratisprogram som växer snabbt.
- Vi tror att **SPSS** är det vanligaste programmet på svenska myndigheter.
- På kursen använder vi **PSPP**. Det påminner mycket om SPSS, men är lite mer begränsat.



- De flesta på institutionen använder **Stata**. Det är stort i forskarvärlden men inte lika vanligt utanför akademien.
- **R** är ett avancerat gratisprogram som växer snabbt.
- Vi tror att **SPSS** är det vanligaste programmet på svenska myndigheter.
- På kursen använder vi **PSPP**. Det påminner mycket om SPSS, men är lite mer begränsat.
 - Gratis!



- De flesta på institutionen använder **Stata**. Det är stort i forskarvärlden men inte lika vanligt utanför akademien.
- **R** är ett avancerat gratisprogram som växer snabbt.
- Vi tror att **SPSS** är det vanligaste programmet på svenska myndigheter.
- På kursen använder vi **PSPP**. Det påminner mycket om SPSS, men är lite mer begränsat.
 - Gratis!
 - Man kan skriva upp SPSS på sitt CV.



- De flesta på institutionen använder **Stata**. Det är stort i forskarvärlden men inte lika vanligt utanför akademien.
- **R** är ett avancerat gratisprogram som växer snabbt.
- Vi tror att **SPSS** är det vanligaste programmet på svenska myndigheter.
- På kursen använder vi **PSPP**. Det påminner mycket om SPSS, men är lite mer begränsat.
 - Gratis!
 - Man kan skriva upp SPSS på sitt CV.
- **SAS** om ni vill arbeta på SCB med stora registerdatabaser.