

F16

Korrelation och
regression
Tidsserier
Exponentiella trender

Regression

- Om sambandet mellan två variabler (på minst intervall-nivå) verkar linjärt kan vi anpassa den "bästa" räta linjen genom punktsvärmen med MK-metoden
- Vi minimerar summan av de kvadrerade avvikelserna från linjen

$$\sum_{i=1}^n e_i^2$$

- där e_i är residualen för observation i , $i=1,2,\dots,n$
- Residualen beräknas som

$$e_i = y - \hat{y} = y - (a + bx)$$

Regression

- Regressionsekvationen

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

- Regressionskoefficienten b , dvs lutningen på linjen, eller riktningskoefficienten kan beräknas från formler (se KD F14-F16). Tolkningen på b är den genomsnittliga ökningen i Y då X ökar en enhet.
- Interceptet a , eller "konstanten" är det skattade värdet på Y då X antar värdet 0, dvs där linjen skär Y -axeln. Kan också beräknas från formler

Att rita in den skattade linjen

- När vi har skattat a och b kan vi rita in den räta linjen i spridningsdiagrammet. Sätt då in två värden på X i ekvationen, erhåll två skattade Y -värden, samt rita in linjen

Korrelation

- Korrelationskoefficienten r_{XY} mäter styrkan på det linjära sambandet.
- Regressionskoefficienten b mäter endast lutning – ej styrka
- Vid enkel regression gäller att $r_{XY}^2 = R^2$
- Kan vi beräkna korrelationskoefficienten genom att veta determinationskoefficienten?

Residualspridning

- Vi anpassar regressionslinjen så att de kvadrerade residualerna minimeras
- När vi har anpassat linjen kan vi även beräkna residualspridningen
 - Ett mått på hur mycket observationerna avviker från linjen
 - Standardavvikelsen för residualerna

$$s = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n-2}}$$

- Kvadraten på residualspridningen är residualvariansen
- Residualspridningen anges som s i Minitabutskrift

Ex Minitabutskrift (Pris i tusen kr för villor och fritidshus) se Övn R4

```
■ The regression equation is
■ Pris på fritidshus = - 39,9 + 0,582 Pris på villor

■ Predictor      Coef    SE Coef    T    P
■ Constant      -39,94   55,15   -0,72  0,477
■ Pris på villor  0,58220  0,09590  6,07  0,000

■ S = 52,9523   R-Sq = 63,7%   R-Sq(adj) = 62,0%

■ Analysis of Variance

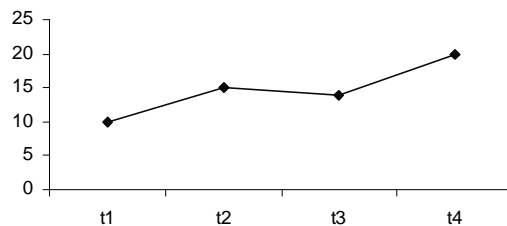
■ Source        DF      SS      MS      F      P
■ Regression     1  103333  103333  36,85  0,000
■ Residual Error 21   58883   2804
■ Total         22  162216

■ Unusual Observations

■      Pris på      Pris på
■ Obs  villor      fritidshus  Fit  SE Fit  Residual  St Resid
■ 9      878      479,0  471,2  32,1     7,8     0,18 X
■ 22     687      461,0  360,0  16,2    101,0   2,00R
■ 23     479      375,0  238,9  13,7    136,1   2,66R
```

Tidsserier

- En variabel som varierar över tiden
- Vi studerar den, vanligen med jämna tidsintervall
 - Varje år, halvår, tertial (4 mån), kvartal, månad, vecka, dag
- Linjediagram, stapeldiagram



Variationer över tiden

- Trenden – den långsiktiga utvecklingen
- Konjunkturvariationer
 - Cykliska variationer
- Säsongsvariationer
- Slumpvariationer

Trenden

1. Antagande om att en viss *modell* gäller för trenden
2. Inget modellantagande

Modeller

- Linjär trend
 - $Y = a + bt$
- Exponentiell trend
 - $Y = ab^t$
- Andragradskurva
 - $Y = a + bt + ct^2$

Linjär trend

- Tiden som oberoende variabel
 - $\hat{Y} = a + b \times t$
- Den absoluta förändringen av Y då t ökar en enhet är densamma för alla värden på t
- Minsta kvadratmetoden
 - Samma som vid enkel regression

$$\hat{Y} = a + b \times X$$

Skattning av a och b

- Om vi centrerar t så blir beräkningarna enklare

År (x)	1998	1999	2000	2001	2002
t	-2	-1	0	1	2

År (x)	1997	1998	1999	2000	2001	2002
t	-5	-3	-1	1	3	5

eller

t	-2,5	-1,5	-0,5	0,5	1,5	2,5
---	------	------	------	-----	-----	-----

Se K&W s186

Formler för a och b då t är centrerad

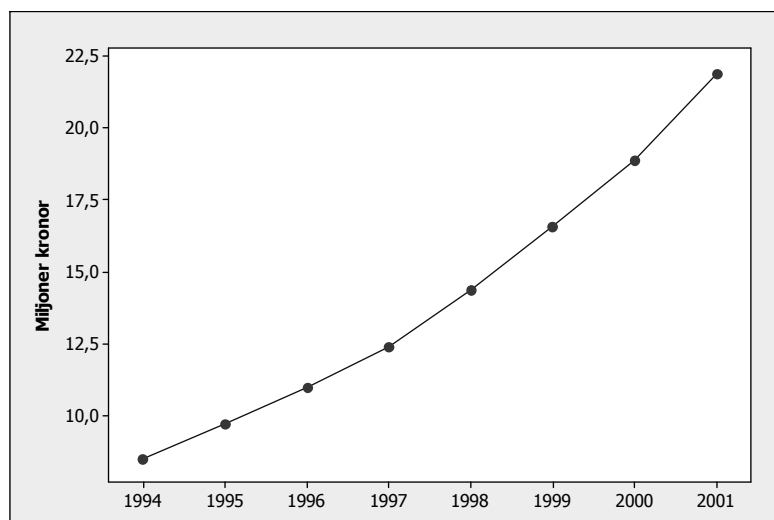
$$b = \frac{\sum_{i=1}^n y_i t_i}{\sum_{i=1}^n t_i^2}$$

$$a = \bar{y}$$

Prognoser

- Behövs som underlag för beslut
- Vi kan göra prognoser enligt den skattade linjen
 - Förutsättningar:
 - Utvecklingen i framtiden följer den anpassade trendfunktionen
 - Den slumpmässiga avvikelsen blir liten

Exponentiell trend (se K&W s191)



Exponentiell trend

- Exponentialfunktion $y = a \times b^t$
- Vi kan göra om denna funktion till en linjär funktion genom att logaritmera

$$\log y = \log(a \times b^t)$$

- Då får vi

$$\log y = \log a + t \times \log b$$

- Vi döper om termerna:

$$\log y = y' \quad \log a = a' \quad \log b = b'$$

- Då har vi fått en "vanlig" rät linje!

$$y' = a' + b' \times t$$

Exponentiell trend

- Nu kan vi använda MK-metoden för att skatta regressionslinjen
- Istället för att använda y -värdena använder vi $\log y$ -värden
- Lösningarna ger $\log b$ och $\log a$ så vi får sedan ta antilogen för att få a och b .