

Regresjonsanalyse

- Hvordan finne ut om det er en sammenheng mellom x og y?
 - Krysstabellanalyse
 - Korrelasjonsanalyse
 - Regresjonsanalyse
- Bivariat regresjonsanalyse: Tilpasse en linje til et prikkdiagram
- $Y = a + bx$
- A = konstantleddet (skjæringspunktet mellom y-aksen og regresjonslinja)
- B = stigningsforholdet til regresjonslinja



Universitetet i Tromsø
Det samfunnsvitenskapelige fakultet



Bivariat regresjonsanalyse

- Man studerer effekten av en uavhengig variabel på den avhengige
- Forutsetning: begge variablene må være på høyt målenivå
- F.eks. sammenhengen mellom utdanning og inntekt, som i utgangspunktet er en uoverkommelig oppgave for krysstabell



Universitetet i Tromsø
Det samfunnsvitenskapelige fakultet

[illegible][illegible]

Universitetet i Tromsø
Det samfunnsvitenskapelige fakultet



- Regresjonsanalyse takler dette ved å beskrive sammenhenger som en rett linje i et koordinatsystem.
- Ikke-lineære sammenhenger estimeres som om de var lineære.
- Linjas stigningsforhold angir effekten av x



Universitetet i Tromsø
Det samfunnsvitenskapelige fakultet

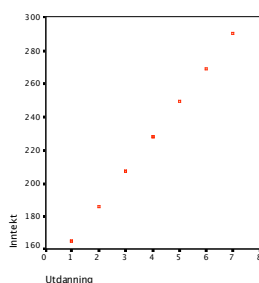


	Krysstabell	Regresjon
Fordeler	- Enkel - Oversiktlig	-Variabler på høyt målenivå kan brukes -Tillater mange uavhengige variabler
Ulemper	-Få verdier -Info-tap?	-Komplisert



Universitetet i Tromsø
Det samfunnsvitenskapelige fakultet





$$Y = 144 + 20x$$



Universitetet i Tromsø
Det samfunnsvitenskapelige fakultet

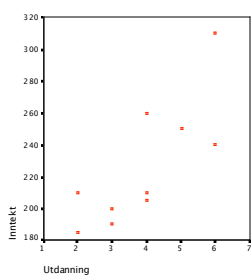


- Samfunnsvitenskap:
- $Y = a + bX + e$
- E = feilterm



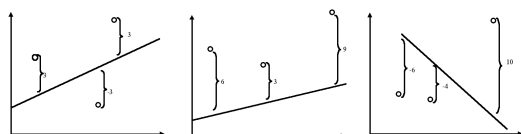
Universitetet i Tromsø
Det samfunnsvitenskapelige fakultet





Universitetet i Tromsø
Det samfunnsvitenskapelige fakultet





Universitetet i Tromsø
Det samfunnsvitenskapelige fakultet



Utrekning av kovarians mellom utdanning og inntekt

Enhet nr.	År utdanning	Inntekt (*1000)	(x-Mx)	(y-My)	(x-Mx)(y-My)	=
1	2	210	(2-3,9)	(210-226)	(-1,9) x (-16)	30,4
2	4	260	(4-3,9)	(260-226)	(0,1) x (34)	3,4
3	6	240	(6-3,9)	(240-226)	(2,1) x (14)	29,4
4	5	250	(5-3,9)	(250-226)	(1,1) x (24)	26,4
5	3	190	(3-3,9)	(190-226)	(-0,9) x (-36)	32,4
6	3	200	(3-3,9)	(200-226)	(-0,9) x (-26)	23,4
7	6	310	(6-3,9)	(310-226)	(2,1) x (84)	176,4
8	2	185	(2-3,9)	(185-226)	(-1,9) x (-41)	77,9
9	4	205	(4-3,9)	(205-226)	(0,1) x (-21)	-2,1
10	4	210	(4-3,9)	(210-226)	(0,1) x (-16)	-1,6
Kovarians(xy) = $\Sigma(x-Mx)(y-My)$						396



Universitetet i Tromsø
Det samfunnsvitenskapelige fakultet



Utrekning av varians i X

Enhet nr.	År utdanning	(x-Mx)	(x-Mx) ²	=
1	2	(2-3,9) = -1,9	(-1,9) ²	3,61
2	4	(4-3,9) = 0,1	(0,1) ²	0,01
3	6	(6-3,9) = 2,1	(2,1) ²	4,41
4	5	(5-3,9) = 1,1	(1,1) ²	1,21
5	3	(3-3,9) = -0,9	(-0,9) ²	0,81
6	3	(3-3,9) = -0,9	(-0,9) ²	0,81
7	6	(6-3,9) = 2,1	(2,1) ²	4,41
8	2	(2-3,9) = -1,9	(-1,9) ²	3,61
9	4	(4-3,9) = 0,1	(0,1) ²	0,01
10	4	(4-3,9) = 0,1	(0,1) ²	0,01
Varians (x) = $\Sigma(x-Mx)^2$				19,08



Universitetet i Tromsø
Det samfunnsvitenskapelige fakultet



Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Utdanning ^a	.	Enter

- a. All requested variables entered.
b. Dependent Variable: Inntekt

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.781 ^a	.611	.562	25,72173

- a. Predictors: (Constant), Utdanning

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8297,14	1	8297,143	12,541	,008 ^a
	Residual	5292,86	8	661,607		
	Total	13590,0	9			

- a. Predictors: (Constant), Utdanning
b. Dependent Variable: Inntekt

Coefficients^a

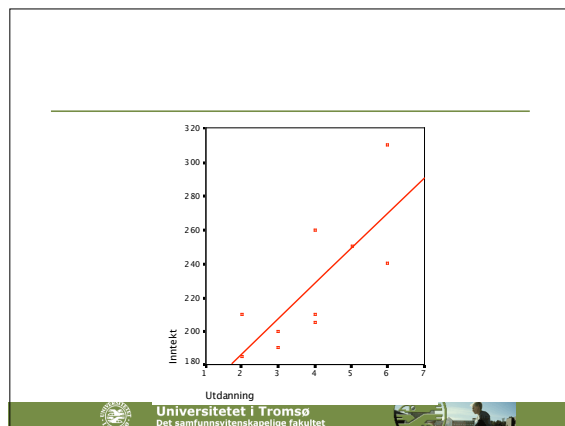
Model		Unstandardized Coefficients	Standardized Coefficients	t	Sig.
1	(Constant)	144,29		24,466	,000
	Utdanning	20,952	,781	3,541	,008

- a. Dependent Variable: Inntekt



Universitetet i Tromsø
Det samfunnsvitenskapelige fakultet





Grunneffekten

- Grunneffekten (a-koeffisienten/konstanten) angir verdien til Y når $x = 0$.
- Substansiell fortolkning ikke alltid mulig

B-koeffisienten

- B-koeffisienten angir regresjonslinjas stigningsforhold.
- Fortolkning: Den forventede gjennomsnittlige endringa på den avhengige variabelen som følge av en enhets endring på den uavhengige variabelen.

Ustandardisert mål (dvs. at den vil variere mye avhengig av hvilken målestokk variabelen er i (kroner, prosentpoeng, dummyvariabel, holdningsskala, m.m)

Universitetet i Tromsø
Det samfunnsvitenskapelige fakultet

Korrelasjon

- Pearson's r
- Varierer mellom -1 og 1, der ytterverdiene indikere meget sterke sammenhenger mens 0 indikerer ingen sammenheng.
- Standardisert, symmetrisk mål, og gjelder bare for bivariate sammenhenger
- Lineært mål. Sammenhengen kan være veldig sterk og kurvlineær, men likevel ikke slå ut med Pearson's r .

Hvis man vil regne ut Pearson's r mellom to variabler x og y , kan man bruke følgende formel:

$$\text{Pearson's } r = \frac{\text{kov}(xy)}{S_x S_y}$$

Universitetet i Tromsø
Det samfunnsvitenskapelige fakultet

T-verdi

Formel:

$$\frac{b}{Sb}$$

- Brukes til *signifikanstesting*.
- Standardfeilen: Gjennomsnittsmål for avstanden fra hver predikert verdi og de faktiske, observerte verdiene
- Tommelfingerregelen: b-koeffisienten gjelder for populasjonen hvis den er mer enn dobbelt så stor som sin standardfeil.

$$\frac{b}{Sb} > 2$$

Sb



Universitetet i Tromsø
Det samfunnsvitenskapelige fakultet



Signifikans

- Sannsynligheten for at sammenhengen mellom de to variablene er lik null i virkeligheten, dvs. at regresjonskoeffisienten (b) er lik null i populasjonen, gitt utvalgsresultatet.

Forklart varians

- Angir hvor stor del av variansen i den avhengige variabelen som forklares av den uavhengige. Målet for forklart varians er R^2 , og den varierer mellom 0 og 1.
- Jo høyere andel forklart varians, jo lavere andel uforklart varians. Hvis R^2 er lav betyr det at den uavhengige variabelen ikke kan forklare variasjonen i den avhengige variabelen særlig godt. Uforklart varians blir dermed tilsvarende høy.

I en bivariat regresjonsanalyse kan vi kvadrere

Pearson's r for å finne R^2



Universitetet i Tromsø
Det samfunnsvitenskapelige fakultet



Multivariat regresjonsanalyse

- Forklare Y bedre
- Øke modellens forklaringskraft (R^2)
- Få kontrollerte mål for hver uavhengige variabel
- Ligning:
- $Y = a + bx_1 + bx_2 + bx_3 + bx_4 + \dots + e$



Universitetet i Tromsø
Det samfunnsvitenskapelige fakultet



Repondent	År utdanning	Ansiennitet	Inntekt (x1000)
1	2	12	210
2	4	10	260
3	6	3	240
4	5	6	250
5	3	14	190
6	3	12	200
7	6	20	310
8	2	8	185
9	4	7	205
10	4	9	210



Universitetet i Tromsø
Det samfunnsvitenskapelige fakultet



Variables Entered / Removed ^a			
Model	Variables Entered	Variables Removed	Method
1	Ansiennitet Utdanning	.	Enter

a. All requested variables entered.
b. Dependent Variable: Inntekt

Model Summary			
Model	R	R Square	Std. Error of the Estimate
1	.877 ^a	.769	704

a. Predictors: (Constant), Ansiennitet, Utdanning

ANOVA ^a					
Model	Sum of Squares	df	Mean Square	F	
1	Regression	10437.2	2	5218.615	
	Residual	11121.7	7	1588.814	11.687
	Total	21558.9	9		

a. Predictors: (Constant), Ansiennitet, Utdanning
b. Dependent Variable: Inntekt

Coefficients ^a				
Model	Unstandardized Coefficients		Standardized Coefficients	
	B	Std. Error	Beta	Sig.
1	(Constant)	108.63		4.202
	Utdanning	21.627	.4876	.807
	Ansiennitet	1.260	1.488	.399

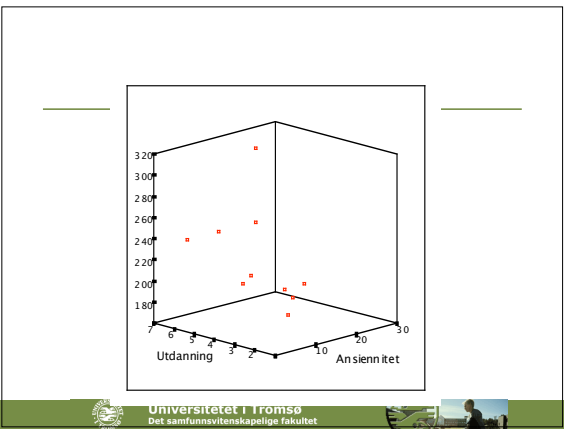
a. Dependent Variable: Inntekt

Universitet i Tromsø
Det samfunnsvitenskapelige fakultet



Universitetet i Tromsø
Det samfunnsvitenskapelige fakultet





Universitetet i Tromsø
Det samfunnsvitenskapelige fakultet