

KORRELATION

XY-punktdiagram

Denne diagramtype bruges til at vise, hvordan værdierne for to variable, x og y, er placeret i et koordinatsystem.

Hensigten er at undersøge, om der er en sammenhæng mellem de to variable. Vi kan f.eks. være interesseret i at undersøge, om tobaksrygning fører til lungekræft. Her har vi altså to variable: 1. Hyppighed og intensitet af rygning og 2. Antal tilfælde og intensitet af lungekræft. Hvis der efter en række observationer på data ser ud til at være sammenfald, så de, der ryger meget, har tendens til at have eller få lungekræft, taler man om, at der er korrelation mellem de to sæt af data.

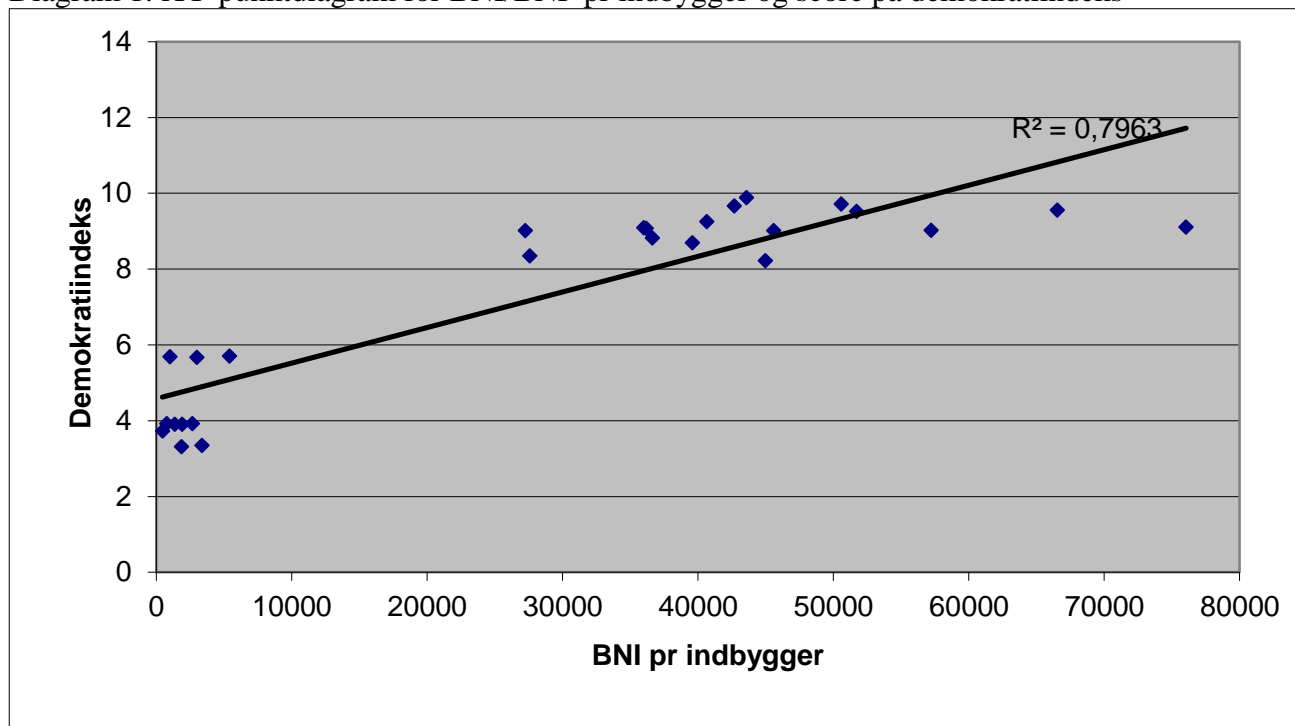
Vi har f.eks. en hypotese om, at højt BNP pr indbygger i et land vil give en tendens til, at der er demokrati i landet. Tabel 1 viser et antal observationer af dette.

Tabel 1 : Observationer af BNP pr indbygger og demokrati for en række lande

	BNP pr indbygger 2006	Demokratiindex
Sverige	43580	9,88
Island	50580	9,71
Holland	42670	9,66
Norge	66530	9,55
Danmark	51700	9,52
Finland	40650	9,25
Luxembourg	76040	9,1
Australien	35990	9,09
Canada	36170	9,07
Schweiz	57230	9,02
Irland	45580	9,01
New Zealand	27250	9,01
Tyskland	36620	8,82
Østrig	39590	8,69
Spanien	27570	8,34
USA	44970	8,22
Tyrkiet	5400	5,7
Nicaragua	1000	5,68
Thailand	2990	5,67
Pakistan	770	3,92
Jordan	2660	3,92
Marokko	1900	3,9
Egypten	1350	3,9
Burkina Faso	460	3,72
Belarus	3380	3,34
Azerbajjan	1850	3,31

Kilde: BNP pr indbygger: Verdensbanken. Demokratiindex: The Economist. Link [her](#) til regneark med korrelationen.

Diagram 1: XY-punktdiagram for BNI/BNP pr indbygger og score på demokratiindeks

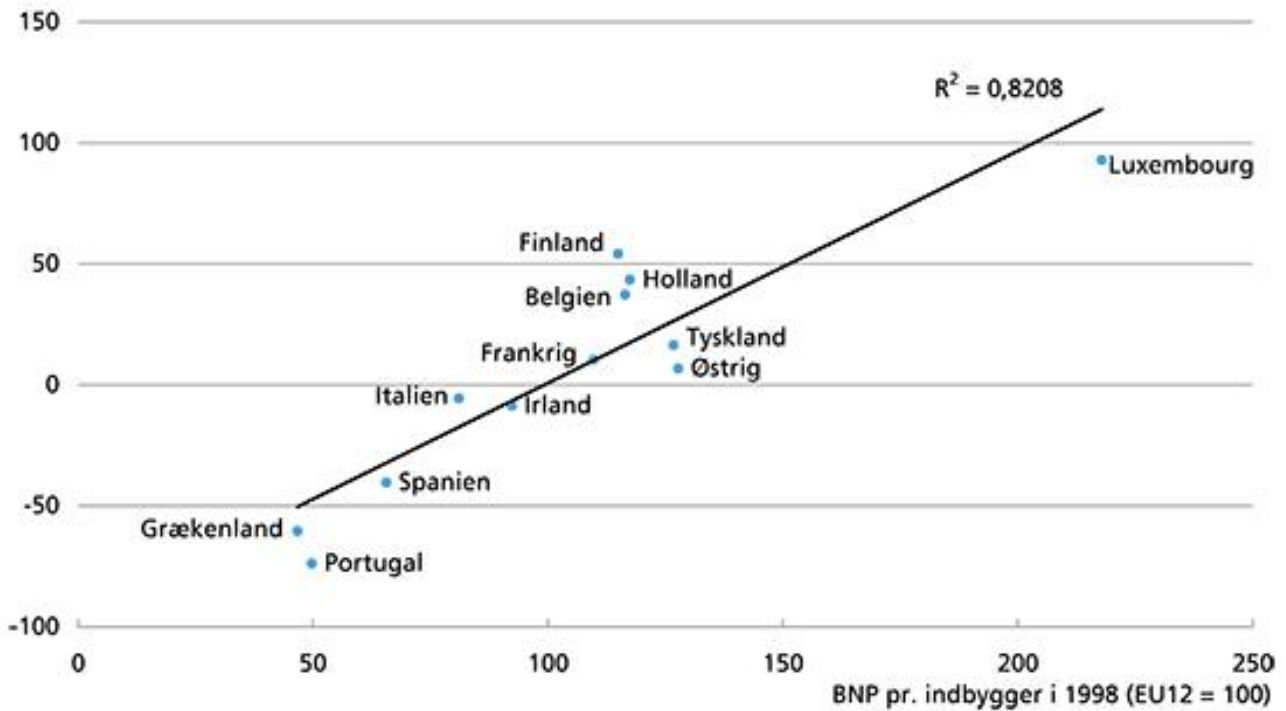


Når cursoren stilles på et af punkterne og der højreklikkes med musen, er der mulighed for at tilføje tendenslinje. Der vælges lineær tendenslinje, og der kan tilføjes ligning for *R squared* (R^2), der viser graden af korrelation (samvariation) mellem de to variable. Jo tættere på 1, desto større er korrelationen.

Et andet eksempel:

Figur 2: Korrelation mellem overskud på betalingsbalancen og BNP pr indbygger

Akkumuleret overskud på betalingsbalancen 1998-2007, pct. af BNP



Kilde: Nationalbanken. Kvartalsberetning 4, 2010. Link [her](#) til demokratiex på korr.

Ovenover ses et eksempel på, hvordan to variable er forsøgt korreleret i en artikel om EU-økonomi. Den ene variabel, y-variablen, er akkumuleret (ophobet) overskud på landes betalingsbalancer i en årrække. Den anden variabel, x-variablen er BNP pr indbygger i 1998 i landene. Man kan nu se, om der er sammenhæng mellem variablene. Det måles ved R squared (kvadrat af R), i dette tilfælde udregnet af et regneark. Ved en værdi på 1 er der 100 pct sammenhæng mellem variablene. Ved en værdi på 0 ingen sammenhæng overhovedet.

Selv om der er en høj værdi af R, kan man ikke uden videre udlede, at der er en kausalrelation mellem variablene, altså at de hænger sammen i et årsags-virkningsforhold. Men ofte vil man kunne opstille en hypotese, f.eks. ovenover, at et højt BNP pr indbygger tenderer imod at give overskud på betalingsbalance.