**Bent Fischer-Nielsen: Lineær regression – hvordan?**

Marker de tal (og kun selve tallene) i regnearket, som skal undersøges for lineær regression. Dvs. to rækker eller søjler med samhørende tal, fx benzinpris og årskørsel for en række år. Vælg Indsæt > Punktdiagram. Punktdiagram tegnes nu. Vælg et Diagramlayout. Du kan nu skrive Diagramtitel og Aksetitel ved at klikke på disse ord.

Excel 2010: Vælg Diagramlayout med ”fx” og du får tegnet tendenslinjen inkl. ligning og R2-værdi.
Excel 2013-: Vælg i ”Hurtigt layout” diagram med ”fx” og du får tegnet tendenslinje mm.
MAC: Gør nogenlunde dom ovenfor eller højreklik på et af punkterne i diagrammet og vælg Tilføj tendenslinje og sæt hak i Vis ligning i diagram og i Vis R-kvadreret.
Du har nu konstrueret et diagram med lineær regression!

Hvordan skal det så tolkes? ***Tendenslinjen*** er den linje, der ligger tættest på punkterne. Ligningen for linjen er af typen y = ax + b, hvor a angiver *hældningskoefficienten (= regressionskoefficienten)* for linjen: I ligningen y = -798,83x + 23270 er hældningskoefficienten -798,83: Når x (benzinpris pr. liter) stiger med 1 kr., falder y med 798,83 km (årskørsel ). Modellen udtrykker denne sammenhæng mellem x og y. Modellen er kvantitativ og lineær.

**R2** viser ***forklaringsgraden***, som udtrykker hvor stor en del af variationen i den afhængige variabel *y*, der forklares/ beskrives af den uafhængige variabel x/tendenslinjen/ligningen/modellen[[1]](#footnote-1). Lineær regression er ikke i sig selv et bevis for, at der er en årsagssammenhæng. For at kunne tale om en sammenhæng skal vi **fagligt/teoretisk kunne begrunde en sammenhæng mellem de to variable**. R2 varierer mellem 0 og 1. Hvis R2 er 1, ligger alle punkterne præcis på den rette linje, og tendenslinjen kan forklare 100 % af variationen i den afhængige variabel y. Jo tættere punkterne ligger på tendenslinjen, jo højere bliver R2. Jo mere spredt punkterne ligger i forhold til linjen, jo mere vil R2 nærme sig 0. Hvis R2 er 0,59, ligger punkterne lidt spredt i forhold til linjen, men har stadig form som ”en cigar” omkring linjen. x/tendenslinjen/ligningen/modellen om sammenhængen mellem årskørsel og benzinpris kan forklare 59 % af variationen i y (årskørsel).

**Metodediskussion:** Kan vi være sikre på at den ene variabel påvirker den anden variabel, blot fordi R2 er høj, eller kan der tænkes andre forklaringer? Der må bl.a. tages forbehold for, om der er en *tredje variabel*, der kan forklare variationen i de to faktorer. Man kan diskutere og evt. undersøge, hvilken tredje variabel det kunne være. Kunne fx øget miljøbevidsthed forklare, både at benzinprisen steg, og at årskørslen faldt?
Man kan evt. undersøge de punkter, der afviger fra linjen. Sæt markøren på et punkt og se hvilke værdier, der knytter sig til punktet. Punktet (8,71, 14700) er fra 1975 og ligger under linjen. Hvad kan forklaringen være?

**Øvelser:** Konstruer diagram med lineær regression ud fra data på de øvrige ark i filen. **Hvordan skal man tolke 1) Hældningskoefficienten (regressionskoefficienten) i ligningen, 2) R2 og 3) punkternes beliggenhed i diagrammet?**
Hvad viser lineær regression for BNP og CO2? Slet USA og Kina i tabellen, og se hvad der sker med figuren, tendenslinjens ligning og R2. Det ses, at R2 er meget påvirkelig af ekstreme observationer (outliers) som USA og Kina. Hvilke forklaringer kan der være på at Rusland, Kina, Brasilien og Frankrig afviger fra tendenslinjen?
Hvilke ”Data i EU-lande” kan det være relevant at teste for lineær regression? Hvad bliver resultatet?
Hvordan inddrager vi lineær regression i undervisningen?
Test for *andre former for regression* fx eksponentiel, hvis punkterne i diagrammet ikke ligger omkring en linje.

1. ”Vi siger derfor, at den lineære model har en forklaringsgrad på 43,8 %, fordi den kan forklare 43,8 % af datapunkternes variation, men der er altså en betydelig restvariation, som modellen ikke kan forklare. Den kan så enten skyldes skjulte variable, som f.eks. køn eller tilfældige variationer…” Matematiklærerforeningen 2010: MATHIT side 63 [↑](#footnote-ref-1)