## Excel tutorial om lineær regression

I denne tutorial skal du lære at foretage *lineær regression* i Microsoft Excel 2007. Det for­udsættes, at læseren har været igennem det indledende om *lineære funktioner*. Vi fore­stiller os at Kurt kører en tur på cykel. Han har en kilometertæller på cyklen. Des­vær­re glemte Kurt at nulstille tælleren hjemmefra. Han kører derudaf og for hvert 5. mi­nut kigger han på uret for at se, hvor lang tid han har været undervejs. Nedenstående tabel viser måleresultaterne.



|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Tid (min) | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 |
| Tæller (km) | 13,9 | 16,3 | 17,9 | 20,1 | 22,9 | 24,4 | 26,2 | 28,9 |

#### Punkt 1

I det følgende skal vi analyser Kurts måledata. Det kan vi passende gøre i regnearket Excel. Start programmet og skriv ”Lineær regression i Excel”, mens du står i celle A1. Hvis du ønsker det, kan du eventuelt klikke i en ny celle og derefter klikke tilbage i cel­le A1 for at markere cellens indhold og derefter ændre skriftens størrelse til en anden via båndet hørende til fanebladet *Startside*. I cellerne A3 og B3 skriver du henholdsvis ”Tid” og ”Tæller”, som er de to *størrelser*, der måles. Nedenfor i cellerne A4 og B4 skri­­ver du enhederne for de to størrelser, som er henholdsvis min og km. Du er nu klar til at skrive måledata ind. Da der er en fast afstand (her 5 min.) imellem tiderne. behøver du ikke skrive dem alle sammen i søjle A. Start med at skrive de to første, dvs. 5 og 10. Du har nu noget der ligner følgende:



#### Punkt 2

Marker først 5 og 10 i søjle A, mens cursoren er et tykt hvidt plustegn (venstre figur). Bemærk, at der er en lille sort firkant nede i højre hjørne af det markerede område. Bevæg cursoren hen over det og observer, at cursoren skifter til et tyndt sort plustegn (højre figur). Mens cursoren viser ser sådan ud, trækker du nedad, mens du holder venstre musetast nede – vi siger, at vi *nedkopierer*. Du får nu nye tal fordelt med den indbyrdes afstand 5, svarende til afstanden mellem de først markerede tal!

|  |  |
| --- | --- |
| Corel002b.png | Corel002c.png |

Slut af med blot at skrive de resterende data fra tabellen ind. Det er ikke systematiske tal, så du er nødt til selv at skrive dem ind. Det skulle nu gerne så sådan ud:



#### Punkt 3

Vi skal nu have lavet en *graf* for *tællerens visning som funktion af tiden*, hvormed me­nes, at førsteaksen skal indeholde tiden og andenaksen skal indeholde tællerens vis­ning. Det gøres meget nemt på følgende måde: Marker først de dele af de to søjler, som in­de­hol­der tal, dvs. området fra celle A5 til B12. Vælg dernæst fanebladet *Indsæt*. Der kom­­mer et nyt *bånd* frem, og heri klikkes på *Punktdiagram* og den øverste venstre type med punk­ter uden forbindelseslinjer vælges, som vist på figuren på næste side. Der frem­­kommer et diagramområde, som vi skal have rette yderligere til.



Bemærk, at når du markerer diagramområdet, så fremkommer der en ekstra menu helt for­oven kaldet *Diagramværktøjer* med undermenuer *Design*, *Layout* og *Formater*. Heri kan du rette diagrammet til.



Der er et væld af muligheder, og tingene kan gøres på mange måder. Det der er vigtigst for os er, at der kommer en overskrift på diagrammet, så man ved hvad det handler om, samt størrelser og enheder på akserne. Under *Diagramlayout* er der en række præ­de­fi­ne­re­de layouts. Vi vil vælge det øverst til venstre, fremkommet ved at klikke på pilen til højre for diagramlayout. Se figuren på næste side.





Klik i diagramtitlen og skriv ”Kurts cykeltur”. Klik i aksetitlerne og skriv ”Tid (min)” på førsteaksen og ”Tællerens visning (km)” på andenaksen. Klik på objektet ”Serie 1” ude til højre og slet det med delete-tasten. Det er grimt! Det begynder efter­hån­den at ligne noget, men vi savner lodrette gitterlinjer. De fås ved at vælge undermenuen *Lay­out*. Vælg *Gitterlinjer*, *Primære lodrette gitterlinjer* og *Overordnede gitterlinjer*.



Det skulle nu gerne se omtrent således ud:



Du kan altid senere muntre dig med at ændre skrifttyper, skriftstørrelse og skriftfarver, hvis du ønsker, eller blære dig på anden vis med fancy layout. Vi vil dog lade det være som det er her.

#### Punkt 4

Vi kan se, at punkterne omtrent ligger på en ret linje, så det er relevant at foretage lineær regression, dvs. bestemme den *bedste rette linje*, som tilpasser målepunkterne. I Excel-sprog hedder det en ”tendenslinje”. Mens diagrammet stadig er markeret, vælger du via undermenuen *Layout* til *Diagramværktøjer* følgende værktøj: *Tendenslinje*. Vælg *Flere indstillinger for tendenslinje…*. NB! Hvis du ikke kan se værktøjet *Tendenslinje*, kan det være, at det ligger under gruppen *Analyse*.





I den fremkomne dialogboks vælger du *Lineær*, fordi det er lineær regression. Dernæst væl­ger vi at ville have tegnet linjen 5 til venstre for den mindste *x*-værdi samt 5 til højre for den største *x*-værdi ved at anbringe 5 både i *Tilbage* og i *Fremad*. Endelig afmærker vi check-boksen med *Vis ligning i diagram*, fordi vi ønsker at kende ligningen for den rette linje. Afmærkes også *Vis R-kvadreret værdi i diagram*, får man et tal, som indi­ke­rer, hvor godt punkterne ligger på en ret linje. Tallet skal enten være meget tæt på 1 eller –1 for at dette er tilfældet. Vi får følgende:



Vi ser, at hældningskoefficienten (stigningstallet) er 0,421 og konstantleddet 11,854.

## Opgaver

#### Opgave 1

I ovenstående tutorial med Kurts cykeltur:

a) Hvad fortæller grafen om Kurts kørsel?

b) Hvordan kan hældningskoefficienten *a* og konstantleddet *b* tolkes med ord?

c) Hvorfor afbilder man samtlige målepunkter? Er det ikke nok at kunne tegne grafen ud fra to punkter? Hvad er filosofien bag metoden lineær regression?

#### Opgave 2

Der udføres et forsøg, hvor en sten slippes til tiden  og derefter falder frit ned fra Eiffeltårnet. Stenens hastighed måles til forskellige tidspunkter. Tabellen nedenfor viser sammenhørende værdier af tid og hastigheden *v*.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Tid (sek) | 0,5 | 1,5 | 3,0 | 4,5 | 5,0 | 5,5 | 6,5 | 7,0 |
| *v* (m/s) | 4,7 | 14,2 | 28,9 | 44,5 | 47,7 | 52,7 | 63,2 | 67,4 |

a) Benyt Excel til at lave en graf for hastigheden som funktion af tiden. Er sam­men­hæn­gen omtrent lineær?

b) Foretag lineær regression ved hjælp af Excel, idet du i dette tilfælde skal tvinge lin­jen igennem (0,0) ved i dialogboksen *Formater tendenslinje* at afkrydse feltet *Angiv skæ­ring =* og skrive 0 i feltet til højre herfor. Begrundelsen for at tvinge linjen igennem (0,0) er, at man på forhånd ved, at linjen skal gå igennem dette punkt, ef­ter­som stenens has­tighed er 0, når den slippes til tidspunktet 0.

c) Hvilken forskrift får du for hastigheden som funktion af tiden *t*? Hvad er hæld­nings­koefficienten? Hvordan kan den tolkes med ord?

#### Opgave 3

Det er ofte hensigtsmæssigt af overskuelighedsgrunde at farve visse felter i et regneark, for eksempel i overskrifter. Prøv i tutorialen at ændre til:



#### Opgave 4

Undertiden ønsker man færre eller flere tal på akserne end det antal, Excel vælger for en. Prøv i tutorialen at få tal på førsteaksen for hvert 5. sekund. *Hjælp*: Højreklik et sted tæt på et af tallene på aksen. Det vil markere aksen og en kontekstmenu vil komme frem. Vælg *Formater akse…*

#### Opgave 5

Ofte skal man foretage nogle beregninger før man har data klar til at kunne afbildes i et dia­gram. Det er oplagt at lade Excel hjælpe til her. I det følgende tænker vi os, at vi har en kogekande, som vi sætter til at varme 1 liter vand op. For hvert tyvende sekund no­terer vi temperaturen. Vi ønsker at afbilde den tilførte energi  som funktion af tem­­peraturstigningen , hvor  er starttemperaturen. Vi har givet må­lin­ger­ne i de første to søjler på figuren nedenfor. Før vi kan lave dia­gram­met, skal vi imid­ler­tid have udregnet de to sidste søjler for temperaturstigningen og den tilførte energi.



Værdierne i den tredje søjle fås ved fra værdierne i den anden søjle at trække start­tem­pe­raturen på de 23 grader. Det kan gøres smart ved i Excel at benytte en *formel*. Man stil­ler sig i feltet C7 og skriver formlen: . Husk lighedstegnet i starten – den fortæller Excel, at der er tale om en formel. Resten fortæller Excel, at man ønsker at tage det i celle B7 og trække det i celle C3 fra. Dollartegnene rundt om sidstnævnte sikrer, at dette felt *ikke* senere bliver ændret ved nedkopiering.



Afslut formlen ved at klikke i et andet felt. Marker dernæst feltet C7 igen. Og find den lille sorte firkant nederst i højre hjørne af det markerede felt. Kør hen over firkanten ind­til cursoren laves om til et sort plustegn, ligesom under punkt 2 i tutorialen. Træk nu ned, mens du holder venstre musetast nede. Dette kaldes at *nedkopiere*. Du har nu auto­ma­tisk nedkopieret formlen i alle de relevante felter i søjle C. Nu skal du gøre noget lig­nen­de for at beregne energierne i søjle D. Husk at . Efter nedkopiering skulle du ger­ne få værdierne vist i første figur i denne opgave.

Du kan nu endelig lave et diagram ud fra søjle C og D. Marker dem og lav diagrammet. Fore­tag lineær regression. Linjen skal tvinges igennem (0,0). Hvad siger hæld­nings­ko­effi­cien­ten noget om, tror du?

NB! Bemærk at undertiden skal man lave et diagram med to søjler, der ikke står lige op ad hinanden. I dette tilfælde markerer man først søjle 1 og holder derefter Ctrl-tasten nede mens man markerer søjle 2.

#### Opgave 6

Undertiden har man ikke søjlerne stående i den ”rigtige rækkefølge”, når man skal lave et diagram, forstået på den måde, at det kan være at de data, der skal anbringes på *y*-aksen står til venstre for de data, der skal stå på *x*-aksen. Så går det ikke bare at markere begge søjler i et hug og lave et diagram efter forskrifterne ovenfor. Vi skal se på, hvordan man så gør.

a) Start med at indtaste de to søjler, som vist på figuren:



b) Marker de to søjler med data og indsæt derefter et punktdiagram, hvor punkterne forbindes med en blød kurve: Vælg fanebladet *Indsæt* og derefter *Punktdiagram* og den undertype, som er vist på figuren:



c) I det diagram du får frem er der byttet rundt på akserne. Vi skal have fortalt Excel, hvilke data der skal afbildes på *x*-aksen og hvilke data, der skal afbildes på *y*-aksen. Højreklik med musen, mens cursoren er over et tomt område på diagrammet og vælg *Vælg Data…* i kontekstmenuen. Herved får du en dialogboks frem:



d) Marker ”Serie 1” og tryk på knappen *Rediger*. Herved fås følgende boks:



e) Slet nu alt det der står under *X-serieværdier* og gå så over og marker de rigtige *x*-vær­dier, som befinder sig i søjle 2. Straks når du har sluppet venstre musetast som afslutningen på markeringen bliver værdierne i dialogboksen ændret! Slet derefter alt under *Y-serieværdier* og gå over og marker de rigtige *y*-værdier fra søjle 1. Slut af med to gange OK, hvorefter du har den ønskede graf:



#### Opgave 7

Excel har nogle indbyggede funktioner, som kan være nyttige i forbindelse med lineær re­gression. Lad os sige, at man har *x*-værdierne stående i cellerne A5 til A12 og *y*-vær­di­erne fra B5 til B12. Så kan man få udregnet hældningskoefficienten for tendenslinjen i en celle ved i cellen at skrive =INDEKS(LINREGR(B5:B12;A5:A12);1). Skæ­ringen med *y*-aksen fås med =INDEKS(LINREGR(B5:B12;A5:A12);2). Bemærk, at y-værdi­er­ne skal stå først! Hvis man ønsker at få *y*-værdien til et punkt på ten­dens­lin­jen for en given *x*-værdi, så kan det også gøres. Lad os sige, at vi vil have *y*-værdien når *x* er lig med 22. Benyt da: =PROGNOSE(22;B5:B12;A5:A12), hvor det er antaget, at data står i tabellen som angivet tidligere. I øvrigt kan du finde en masse smarte funktioner på en liste, hvis du skriver ”statistiske funktioner” i Excels hjælpemenu.