Morven Muilwijk
E-mail: morven.muilwijk@student.uib.no

UNIVERSITY OF BERGEN
GEOPHYSICAL INSTITUTE

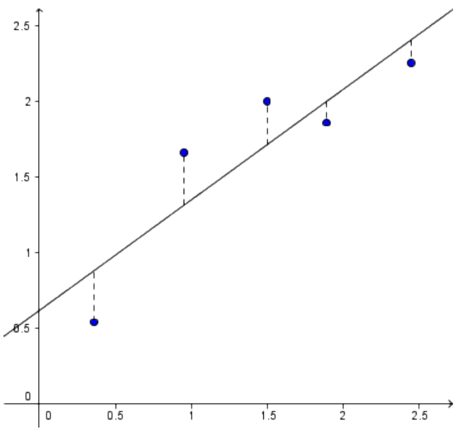
Funksjonstilpasning

Vedlegg til Ekte data oppgave

January 13, 2016

Lineær regresjon

Studier av sammenhengen mellom to variabler ved målinger er ofte motivert ut fra behov for å kunne forutsi den ene variabelen fra den andre. Utviklingen av global temperatur og klimaendringer er et eksempel på dette. I slike sammenhenger er det vanlig å la x betegne den uavhengige variabelen, også kalt inn-variabelen, og la y betegne responsen, eller ut-variabelen. Formålet er å finne en mulig sammenheng mellom verdiene. Første skritt er å plotte spredningsdiagrammet slik vi gjorde i forrige oppgave. Hvis en lineær sammenheng bekreftes av korrelasjonskoeffiesienten kan vi prøve å tilpasse en funksjon.
en linje $y = a + bx$ er bestemt ved to konstanter: Høyden over origo, a, og stigningstallet, b, dvs mengnde y øker med når x øker med en enhet. Se figur.



Vi ønsker å bestemme de verdier a^* og b^* for henholdsvis a og b som gjør at regresjonslinjen $y = a + bx$ er best mulig tilpasset våre data. Et mye benyttet prinsipp er minste kvadraters metode. Prinsippet går ut på å minimere kvadratsummen.
Vi har et tallmateriale $(x_1, y_1), ..., (x_n, y_n)$ og skal tilpasse en rett linje $y^* = a^* + b^* x$. Minste

Morven Muilwijk
E-mail: morven.muilwijk@student.uib.no

kvadraters metode gir da følgende formel til å bestemme a^* og b^*:

$$b^* = \frac{S_{xy}}{S_x^2}, a^* = \bar{y} - b^* \bar{x} \tag{0.1}$$

der $\bar{x} og \bar{y}$ er middelverdiene, og S_x^2 og S_{xy} er definert som i forrige oppgave.

Lineære modeller

Mattematikk brukes mye til å finne sammenhenger mellom forskjellige størrelser i naturfag, teknologiske fag og samfunnsfag. Slike sammenhenger kaller vi for matematiske modeller. Noen ganger kan modellene være svært nøyaktige mens i noen tilfeller gir modeller bare en grov beskrivelse av situasjonen. I visse tilfeller kan modellen gi et helt feilaktig bilde av virkeligheten.

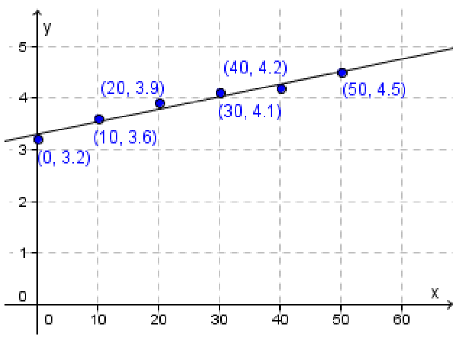
Når vi sammenhengen mellom to størrelser x og y er gitt ved en rett linje i et koordinatsystem, har vi $lineær vekst$. Det finnes da to tall a og b slik at

$$y = ax + b \tag{0.2}$$

Disse konstantene kan vi finne både med og uten digitale hjelpemiddel. Vi sier da at vi har brukt en $lineær matematisk$ modell.

Lineær regresjon

Vi kan bruke digitale hjelpemiddler til å finne den rette linjen som passer best til et datasett. Da bruker vi en metode som heter $regresjon$.



Oppgave

1. Last inn/åpne temperaturdata fra bøyen Gabriel for en hel dag fra et Excel ark. Lag en tabell som viser overflatetemperaturen i vannet (eller lufttemperaturen) fra midnatt til klokken 12 midt på dagen. La dine x verdier være ca antall timer etter midnatt fra 1 til 12 og y-verdier temperaturen.

2. Bruk linear regresjon i Geogebra til å tilpasse en lineær kurve til datasettet. Hva er likningen til linjen?

3. Bruk lineær regresjon i Excel til å tilpasse en lineær kurve til datasettet. Hva er likningen til kurven?

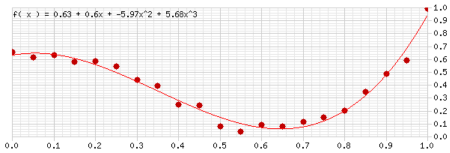
Morven Muilwijk
E-mail: morven.muilwijk@student.uib.no

UNIVERSITY OF BERGEN
GEOPHYSICAL INSTITUTE

4. Dersom vi antar at modellen stemmer, hva vil temperaturen være klokken 3 på ettermiddagen? Bruk digitale hjelpemidler til å finne svaret. (Hint: bruk "skjæring mellom to objekter" i Geogebra)

Polynomfunksjoner

Utrykkene $2x+3$ og x^2+3x-5 kaller vi polynomer. Den høyeste eksponenten til variabelen i et polynom kaller vi $graden$ til polynomet. Når vi skriver polynomer, ordner vi alltid leddene etter graden. På samme måte som vi tilpasset en lineær kurve (polynom av første grad) til et datasett kan vi tilpasse polynomer av høyere grad til et datasett. Fra polynomene kan vi regne ut ekstremalpunkt og eventuelle nullpunkt. Polynomtilpassning gjør vi ved hjelp av digitale hjelpemidler som Geogebra eller Excel. Polynomtilpasning fungerer ofte veldig bra for et visst antall verdier, men gir også ofte dårlig resultat utenfor definisjonsområdet av $x-verdier$. Det vil si at det er ikke alltid en like god modell. Ofte kan et polynom beskrive en del av verdier som svinger mye opp og ned, men skal man lage en modell må man da bruke en svingende funksjon som for eksempel sinus eller cosinus.



Eksponentialfunksjoner og Logaritmer

Vi vet at en eksponentialfunksjon har et funksjonsytrykk på formen

$$f(x) = a \cdot k^x \tag{0.3}$$

der a og k er to konstanter. For at utrykket skal være definert for alle x, må k være et positivt tall.

Logaritmiske funksjoner er funksjoner på formen

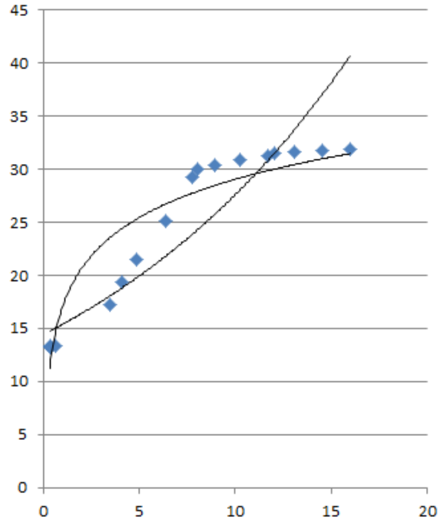
$$g(x) = a \cdot ln(x) \tag{0.4}$$

der ln er den naturlige logaritmen.

På lignende måte som lineære funksjoner og polynom kan også eksponentialfunksjoner og logaritmiske funksjoner beskrive ulike prosesser og tilpasses et datasett.

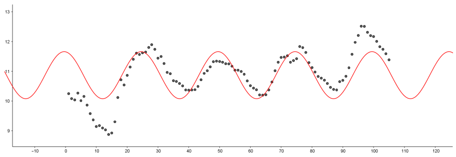
Morven Muilwijk
E-mail: morven.muilwijk@student.uib.no

UNIVERSITY OF BERGEN
GEOPHYSICAL INSTITUTE



Sinusregresjon

Svært ofte observer vi tilnærmet gjentagende mønstre i naturen. Sykluser og gjentagende mønstre må beskrives med sykliske funksjoner skal modellene gi mening. En sinusfunksjon er et eksempel på en syklisk funksjon, dvs at den gjentar seg selv hele tiden. Noe som svinger opp og ned hele tiden kan derfor godt beskrives ved hjelp av en sinuskurve. Temperatur er et eksempel på en variable som til tider endrer seg syklisk.



Figuren viser et eksempel på sinusregresjon på temperaturmålinger tatt av overflatevannet av bøyen Gabriel.