«Uncertainty of the Uncertainty»
Del 4 av 6

v/Rune Øverland, Trainor Elsikkerhet AS

Innledning

Dette er del fire i artikkelserien om «Uncertainty of the Uncertainty».

I dag skal jeg vise deg utledningen av formelen:

$$\sigma_m \approx \frac{1}{\sqrt{n-1}} \cdot s \qquad [1]$$

, og hvorledes man bestemmer standard usikkerhet for en måleserie, og rapporterer riktig.

Utledning av formel og talleksempel

Vi har et måleoppsett, som er vel definert. Noen parametere holdes konstante, mens andre, som noen omgivelsesparametere (lufttrykk, omgivelsestemperatur og kvaliteten på spenningsforsyningen), er ukontrollerte og varierer naturlig. Dette oppfattes som «målestøy».

Vi har en måleserie med to observasjoner:

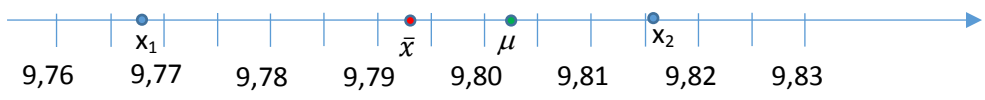
Observasjon nr.	Avlest måleverdi (Volt)
1 [x_1]	9,76788
2 [x_2]	9,81591
Måleseriens middelverdi [$\bar{x}$]	9,791895

Det at måleverdiene varierer skyldes «støy» i måleoppsettet som vi tillater.

Statistikeren plasser de observerte måleverdiene [x_1] og [x_2] på tallstigen. Det samme gjør han (eller henne) med måleseriens kalkulerte gjennomsnittsverdi [$\bar{x}$]. Alle disse tre verdiene er kjente. Statistikeren antar videre at variasjonene i målingene skyldes «støy», og at verdiene x_1 og x_2 er normalfordelt (Gauss-kurven) om $\bar{x}$.

En hemmelighet som nesten blir avslørt

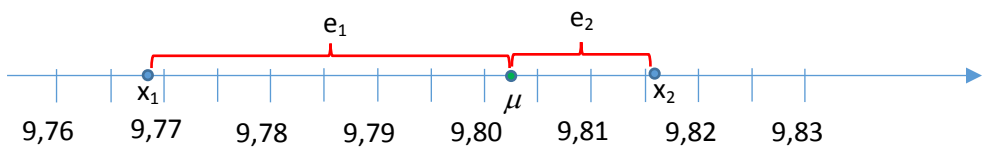
Du og jeg har en hemmelighet! Vi vet at gjennomsnittsverdien for hele populasjonen [μ], ved å ta «Godzilla» av enkeltobservasjoner, kan bevises å være lik μ=9,80125, men det vet ikke statistikeren! Hold på hemmeligheten.



Statistikeren settes på prøve

Vi setter statistikeren på en kunnskapsprøve og ber han fortelle oss verdien på populasjonens standardavvik [σ].

Imidlertid, statistikeren har ingen kjennskap om hvor den sanne verdien [μ] ligger, som representerer måleseriens sanne gjennomsnittsverdi (som du og jeg vet er 9,80125).



For å rapportere måleoppsettets standardavvik [σ], trenger statistikeren i utgangspunktet å vite verden på den sanne gjennomsnittsverdien [μ] for måleoppsettet. Da kunne han ha beregnet måleavvikene:

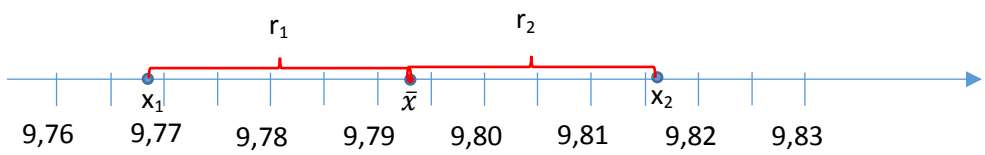
$e_1 = [x_1 - \mu]$ [2]

$e_2 = [x_2 - \mu]$ [3]

Men, da den sanne gjennomsnittverdien [μ] er ukjent for ham, kan statistikeren heller ikke vite noe sikkert om verdiene på [e_1] og [e_2].

Residualene er løsningen

En vei utenom dette er å jobbe med residual. Disse kan han tallsette.



Nå kan statistikeren tallfeste residualene:

$r_1 = [x_1 - \bar{x}]$ [4]

$r_2 = [x_2 - \bar{x}]$ [5]

Generelt har han:

$r_i = x_i - \bar{x}$ [6]

Variansen er lik summen av kvadratet av residualene dividert med antall observasjoner [n] i måleserien:

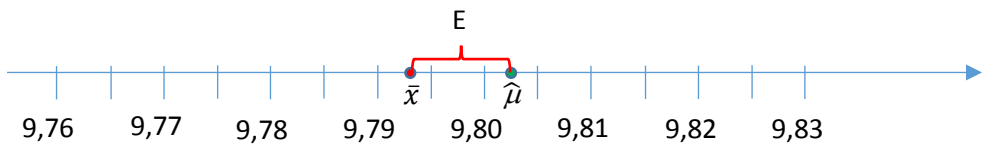
Måleseriens varians = [s^2] $= \frac{Summen\ av\ (kvadratet\ av\ residualene)}{Antall\ observasjoner\ i\ måleoppsettet}$ [7]

$= \frac{\sum[x_i - \bar{x}]^2}{n}$ [8]

'E' er systematisk feil («measurement bias»)

Statistikeren definerer størrelsen [E]. Dette er forskjellen mellom måleseriens kalkulerte gjennomsnittsverdi [$\bar{x}$] og den for ham ukjente, sanne gjennomsnittsverdien i måleserien [$\hat{\mu}$].

$E = \bar{x} - \hat{\mu}$ [9]



For et sett med måleobservasjoner har statistikeren:

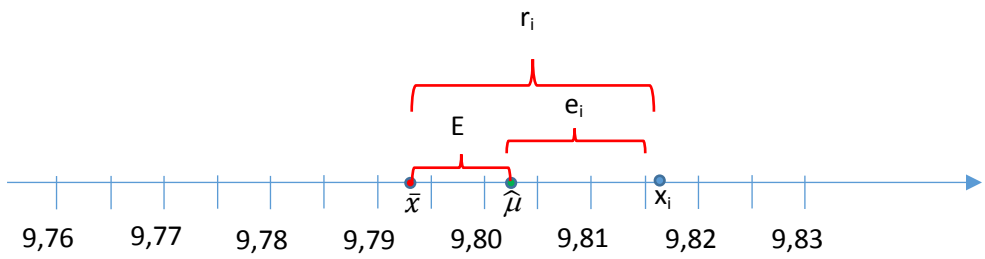
$E = (\bar{x}) - \hat{\mu}$ [10]

$= \left(\frac{\sum[x_i]}{n}\right) - \hat{\mu}$

$= \left(\frac{\sum[x_i - \hat{\mu}]}{n}\right) = \frac{\sum[e_i]}{n}$

[E] er lik gjennomsnittsverdien av avvikene ($\frac{\sum[e_i]}{n}$).

Han kan nå skrive følgende viktige setning for en generell måleobservasjon:



$r_i = (x_i - \bar{x})$ [11]

$= (e_i - E)$ [12]

I og med at statistikeren tilfeldigvis har plassert [$\hat{\mu}$] til høyre for [$\bar{x}$], blir [E] i hans eksempel negativ. Derfor blir det likevel riktig å si at

$r_i = e_i - E$ [13]

Derfor:

Måleseriens varians = [s^2]	Kommentarer [14]
= $\frac{1}{n}\sum(x_i - \bar{x})^2$	Erstatter s^2 med $\frac{1}{n}\sum(x_i - \bar{x})^2$ [15]
= $\frac{1}{n}\sum(e_i - E)^2$	Erstatter $(x_i - \bar{x})$ med $(e_i - E)$ [16]
= $\frac{1}{n}\sum(e_i - E)(e_i - E)$	Løser opp parentesen $()^2$ [17]
= $\frac{1}{n}\sum(e_i)^2 - 2E[\frac{1}{n}\sum(e_i)] + E^2$	Kalkulerer parentesen ()() [18]
= $\frac{1}{n}\sum(e_i)^2 - 2E[E] + E^2$	Erstatter $[\frac{1}{n}\sum(e_i)]]$ med [E] [19]
= $\frac{1}{n}\sum(e_i)^2 - 2E^2 + E^2$	Trekker sammen 2E[E] til $2E^2$ [20]
= $\frac{1}{n}\sum(e_i)^2 - E^2$	Subtraherer $- 2E^2 + E^2$ [21]

Måleseriens varians = [s^2] = $\frac{1}{n}\sum(e_i)^2 - E^2$ gjelder for ett målesett á [n] observasjoner.

Som før, statistikeren kan foreta nye, mange måleoppsett med [n] observasjoner. Han kan da kalkulere gjennomsnittlig varians for målesettene. Statistikeren får da følgende resultat:

$\overline{s^2} = \sigma^2 - (\sigma_m)^2$ [22]

, hvor

$\overline{s^2}$ er gjennomsnittlig varians til målesettene

σ^2 er variansen til populasjonen

$(\sigma_m)^2$ er variansen til måleoppsettets gjennomsnittsverdi

Sentralgrenseteoremet

Fra en tidligere artikkelserie tok jeg for meg sentralgrenseteoremet. Der konkluderte statistikeren med at:

	Kommentar
$\sigma^2 = n \cdot (\sigma_m)^2$	Populasjonens varians = n multiplisert med utvalgsvariansen [23]
$(\sigma_m)^2 = \frac{\sigma^2}{n}$	Dividerer med n på hver side av ligningen [24]
$\sqrt{(\sigma_m)^2} = \sqrt{\frac{\sigma^2}{n}}$	Trekker kvadratroten på hver side av likningen [25]

$\sigma_m = \frac{\sigma}{\sqrt{n}}$	Rydder. [26]
$\sigma = \sqrt{n} \cdot \sigma_m$	Alternativ måte å skrive ligningen [23] på [27]

Uttrykket ($\sigma_m = \frac{\sigma}{\sqrt{n}}$) er sentralgrenseteoremet.

Standard usikkerhet for måleoppsettenes gjennomsnittsverdi [σ_m] er lik populasjonens standardavvik [σ] dividert med antall observasjoner [n] i hver måleserie.

Standard usikkerhet for måleoppsett

Statistikeren kan skrive:

$\overline{s^2} = [\sigma^2] - (\sigma_m)^2$	Formel hentet fra [22]. Gjennomsnittlig varians ($\overline{s^2}$) er lik differansen mellom populasjons-variansen og utvalgsvariansen. [28]
$\overline{s^2} = [n \cdot (\sigma_m)^2] - (\sigma_m)^2$	Erstatter [σ^2] med uttrykket fra sentralgrenseteoremet $\sqrt{n} \cdot \sigma_m$ [29]
$\overline{s^2} = (\sigma_m)^2 \cdot (n - 1)$	Setter $(\sigma_m)^2$ utenfor fellesparentesen [30]
$(\sigma_m)^2 = \frac{\overline{s^2}}{(n - 1)}$	Dividerer høyre og venstre side med (n – 1), og rydder [31]
$\sqrt{(\sigma_m)^2} = \sqrt{\frac{\overline{s^2}}{(n - 1)}}$	Trekker kvadratroten på venstre og høyde side [32]
$\sigma_m \approx \frac{1}{\sqrt{n-1}} \cdot s$	Da er formelen [1] utledet som lovet! [33]

, hvor

s er kalkulert standard usikkerhet for én spesifikk måleserie

n er antall observasjoner i denne måleserien

σ_m er standard usikkerhet for en gjennomsnittlig måleserie

Oppsummering

For å kalkulere måleoppsettets standard usikkerhet, bruk:

$$\sigma_m \approx \frac{1}{\sqrt{n-1}} \cdot s \qquad [34]$$

, hvor

n er antall observasjoner i måleserien

s er kalkulert standard usikkerhet for én spesifikk måleserie

Altså, når vi vi skal rapportere den mest representative verdien fra måleoppsett (gjennomsnittverdien), bruker vi beregnet standard usikkerhetsverdi for å karakterisere hvorledes de individuelle måleobservasjoner i måleoppsettet varierer (grunnet «støy»).

$\bar{x}\ (\sigma_m)$

For eksempel slik:

9,82(0,05) volt.

Neste del om «Uncertainty of the Uncertainty»

I neste del i denne artikkelserien skal jeg vise utledningen av formelen for uttrykket «Uncertainty of the Uncertainty»:

$$\frac{1}{\sqrt{2v}} \qquad [35]$$

På gjensyn!

Med vennlig hilsen
Trainor Elsikkerhet AS
Rune Øverland
Senioringeniør
Tønsberg juni 2016