

Statistisk behandling av kalibreringsresultatene – Del 3.

v/ Rune Øverland, Trainor Elsikkerhet AS

Denne artikkelserien handler om statistisk behandling av kalibreringsresultatene.

Denne artikkelen har kalibreringskurve (lineær regresjon) som hovedtema. Dette er en teknikk for å estimere koeffisienter for en rett linje.

$$\hat{Y} = \hat{a}_1 \cdot x + \hat{a}_0 \quad [1]$$

I den forrige artikkelen tok vi for oss den grafiske presentasjonen av innsamlede data ved kalibrering. Vi beregnet sentralverdien for en måleserie, og viste hvorledes vi kunne uttrykke spredningen (variasjonen) av måleserien. Basert på sentralverdiene tegnet vi opp en graf.

Datapar

Vi tenker oss et scenario hvor vi har gjort en kalibrering i fem punkter for vår mengdemåler. I den første måleserien påtrykte vi en strømningsrate på $X_1 = 100,00 \text{ m}^3/\text{h}$, og avleste instrumentresponen $99,90 \text{ m}^3/\text{h}$. Deretter kalibrerte vi punkt 2, og så videre. Verdiene er satt inn i tabell 1.

Sub-populasjon 1 ($X_1, Y_{1.1}$)	Sub-populasjon 2 ($X_2, Y_{2.1}$)	Sub-populasjon 3 ($X_3, Y_{3.1}$)	Sub-populasjon 4 ($X_4, Y_{4.1}$)	Sub-populasjon 5 ($X_5, Y_{5.1}$)
(100 , 99,90)	(300 , 300,45)	(500 , 500,83)	(700 , 700,25)	(900 , 899,50)

Tabell 1: Ét datapar fra hver av de fem sub-populasjonene).

For disse dataparene kalles gjerne X for den uavhengige variable, mens y kalles den avhengige variable. Y-verdien avhenger på en eller annen måte av verdien av X.

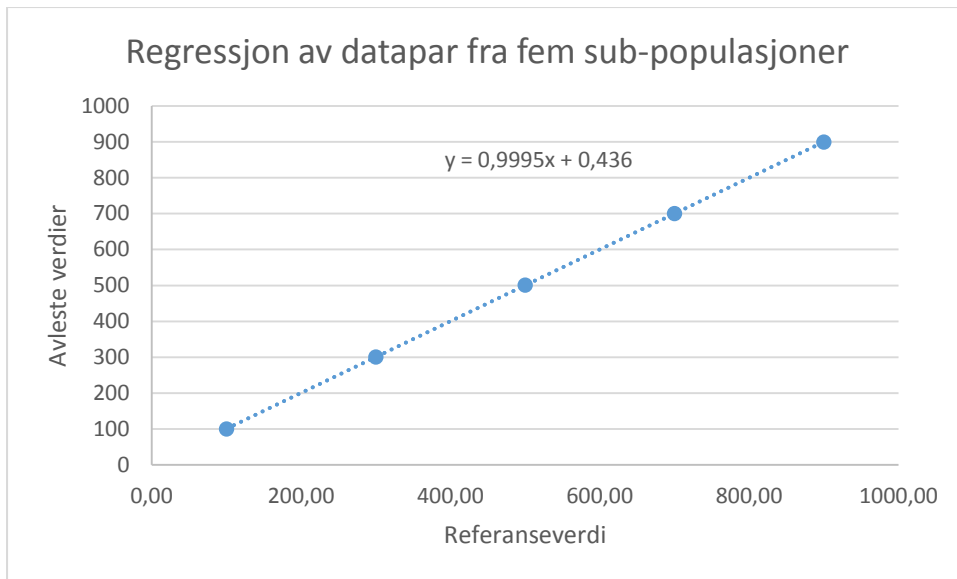
Regresjon

Vi bruker Microsoft Excel[®] til å visualisere en tenkt rett linje som har «Best fit» til tallparene.

Vi arrangerer dataparene slik; med x-verdiene i venstre kolonne og y-verdiene i høyre.

X	y
100,00	99,90
300,00	300,45
500,00	500,83
700,00	700,25
900,00	899,50

Området merkes, og vi setter inn et punktdiagram. Vi konfigurerer diagrammet ved å legge til en lineær linje. Videre ber vi Excel om å vise oss koeffisientene, og derav likningen for «Best fit»-linjen.



Figur 1: Microsoft Excel[®] tegner opp dataparene. Vi konfigurerer grafikken ved å be om presentasjon av «Vis formel i diagrammet». Kalibreringskurven presenteres her som stiplet blå linje.

Excel foreslår basert på de fem tallparene følgende estimatlikning for instrumentresponsen:

$$\hat{Y} = \hat{a}_1 \cdot x + \hat{a}_0 \quad [2]$$

$$\hat{Y} = 0,9995 \cdot x + 0,436 \quad [3]$$

Utvidet usikkerhet for en kalibreringskurve

Basert på «Best fit»-linjen kan vi estimere \hat{Y} -verdier ved ulike x-verdier.

X	$\hat{Y} = 0,9995 \cdot x + 0,436$	\hat{Y}	$y_{i,1}$	Residual ($y_{i,1} - \hat{Y}$)
100	$\hat{Y} = 0,9995 \cdot 100 + 0,436$	100,39	99,90	-0,49
300	$\hat{Y} = 0,9995 \cdot 300 + 0,436$	300,29	300,45	+0,16
500	$\hat{Y} = 0,9995 \cdot 500 + 0,436$	500,19	500,83	+0,64
700	$\hat{Y} = 0,9995 \cdot 700 + 0,436$	700,09	700,25	+0,16
900	$\hat{Y} = 0,9995 \cdot 900 + 0,436$	899,97	899,50	-0,49

Tabell 2: Beregning av Residual. Denne verdien representerer forskjellen mellom faktisk avlest verdi og beregnet avlest verdi basert på «Best Fit»-linjen.

Fra tabellen ser vi at estimert instrumentrespons \hat{Y} avviker noe fra den avleste instrumentresponsen ved kalibrering $y_{i,1}$.

Avviket mellom disse responsverdiene kalles residual. Dette er en «feil» som introduseres ved at regresjonslinjen ikke perfekt kan så gjennom alle tallparene samtidig.

«Best Fit» og koeffisientene \hat{a}_0 og \hat{a}_1

Hva er definisjon på en «Best Fit»-linje?

Jo, det fins et sett med koeffisienter på \hat{a}_0 og \hat{a}_1 som gjør at summen av alle positive og negative residualer er lik null.

Residual ($y_{i.1} - \hat{Y}$)
-0,49
+0,16
+0,64
+0,16
-0,49
Sum av residual = 0 når $\hat{a}_0 = 0,436$ og $\hat{a}_1 = 0,995$

Måleforskriften krav til linearitet

Dersom instrumentet skal brukes på en målestasjon, skal kalibreringen og målingen oppfylle krav i Måleforskriften til Oljedirektoratet.

I paragraf 21 fremkommer det at målesløyfene skal kalibreres med et antall punkter slik at linearitetsavvik avdekkes innenfor arbeidsområdet.

La oss ta noen eksempler:

$y_{i.1}$	X	Avvik [m^3/h] ($x - y_{i.1}$)	Avvik [% av målt verdi] $\left(\frac{x - y_{i.1}}{x}\right)$
99,90	100	-0,10	-0,10
300,45	300	0,45	0,15
500,83	500	0,83	0,17
700,25	700	0,25	0,04
899,50	900	-0,50	-0,06

Tabell: Beregning av linearitet i prosentverdi

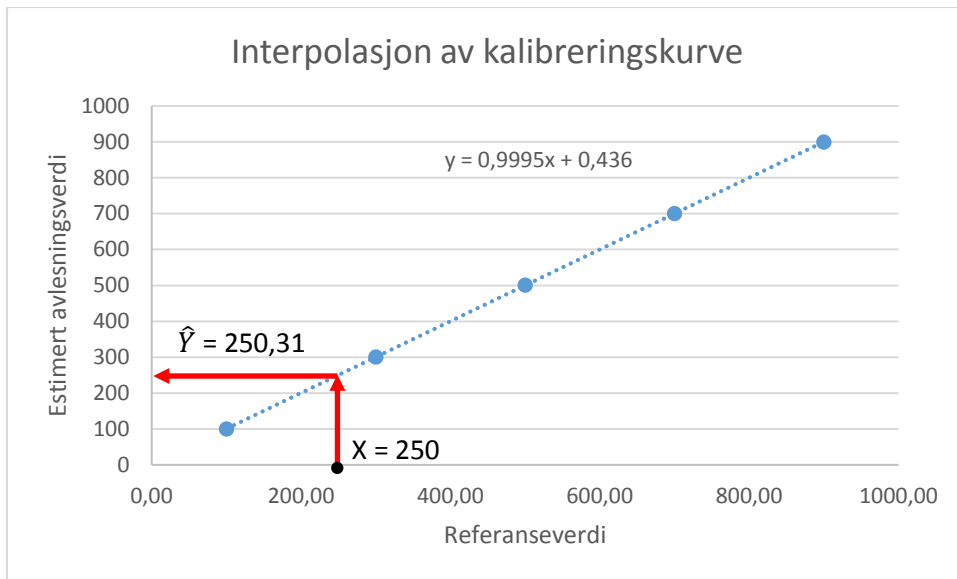
Kalibreringskurve: Interpolasjon: $x \rightarrow \hat{Y}$

Likningen for den estimerte \hat{Y} kan på generelt grunnlag brukes for å kalkulere estimert responsverdier ved en spesifikk x-verdi.

For eksempel; når den påtrykte strømningsraten x er $250 m^3/h$, hvilken instrumentrespons forventer vi?

$$\hat{Y} = 0,9995 \cdot 250 + 0,436 = 250,31 [m^3/h]$$

[4]



Figur 3: Her illustrerer vi interpolasjon av et tallpar. X-verdien ligger mellom to av våre tidligere kalibreringsverdier (100 og 300 m³/h). Basert på «Best Fit»-linjen kalkulerer vi estimert avlesningsverdi.

Kalibreringskurve: Revers interpolasjon: $\hat{Y} \rightarrow x$

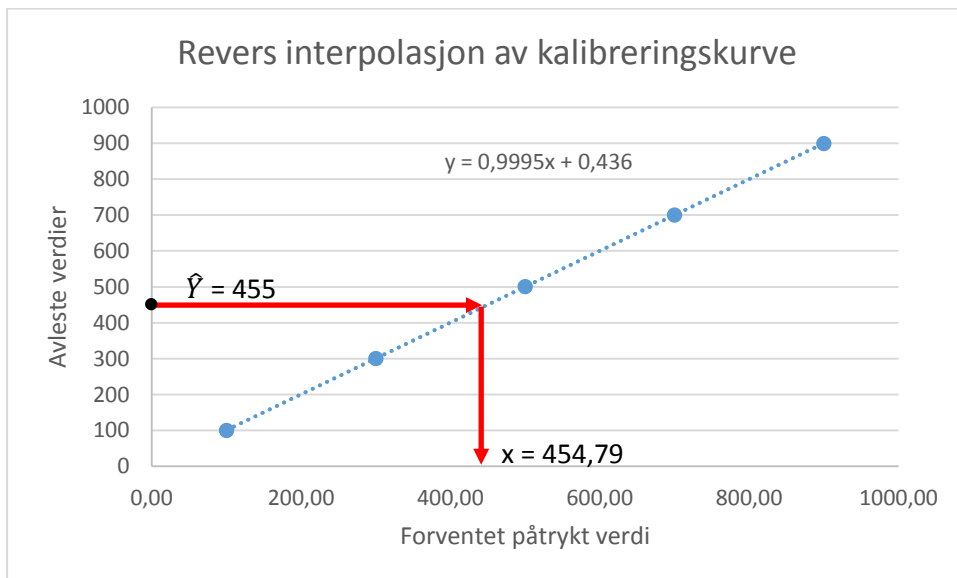
Vi kan selvfølgelig interpolere motsatt retning; slik:

For eksempel; når den avleste strømningsraten \hat{Y} er 455 m³/h, hvilken estimert x-verdi forventer vi?

Vi snur litt på likningen; slik

$$x = \frac{\hat{Y} - 0,436}{0,9995} = \frac{455 - 0,436}{0,9995} = 454,79 \text{ [m}^3/\text{h]}$$

[5]



Figur 4: Når vi avleser instrumentverdien 455 m³/h, har vi den forventede påtrykte strømningsraten lik 454,79 m³/h basert på «Best Fit»-linjen. Den avleste instrumentverdien 455 m³/h ligger mellom to

tidligere kalibreringsverdier (300,45 og 500,83). Vi benytter interpolasjon for å beregne den forventede påtrykte strømningsraten.

Kalibreringskurve: Ekstrapolarisasjon

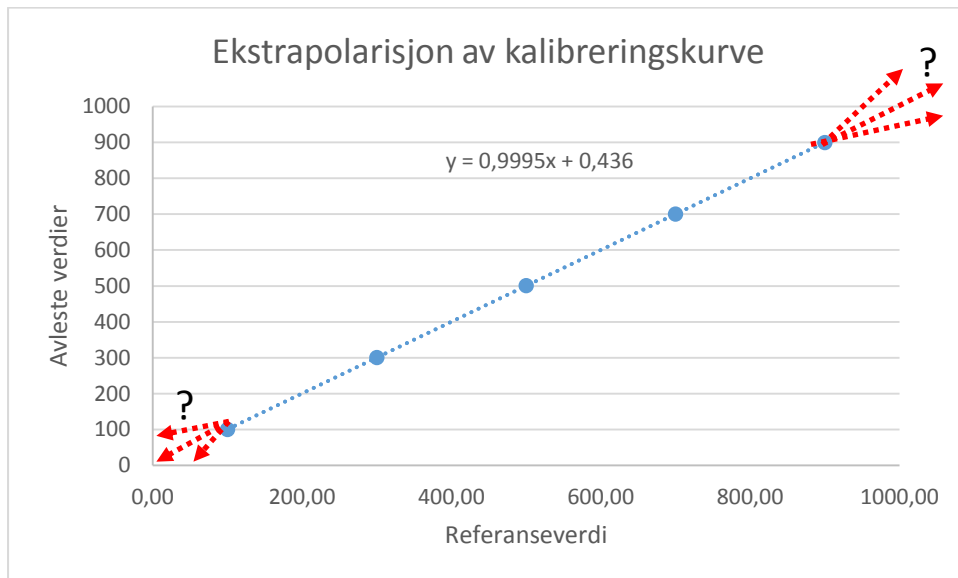
Den laveste x-verdien vi kalibrerte ved var $X_1 = 100 \text{ m}^3/\text{h}$, og den høyeste $X_5 = 900 \text{ m}^3/\text{h}$.

Vi har forutsatt at det er linearitet mellom disse ytterpunktene. Men, hva med utenfor?

Svaret er: Det kommer an på.

Skal instrument brukes på en fiskal målestasjon er svaret nei fordi Måleforskriften definerer i paragraf 13 at måleutstyret skal kunne måle alle planlagte hydrokarbonstrømmer uten at noen del av måleutstyret opererer utenfor sitt arbeidsområde.

For andre bruksområder enn Måleforskriften kan det være at vi kan forutsette at lineariteten fortsetter utenfor det kalibrerte området, men noen god garanti har vi ikke.



Figur 5: I enkelte tilfeller er det ikke lov til å ekstrapolere resultater. Blant annet gjennom Måleforskriften til Oljedirektoratet.

Terminologi

Terminologi benyttet i denne artikkel:

\hat{a}_0	Estimert koeffisient. Har verdien y når $x = 0$.
\hat{a}_1	Estimert koeffisient. Har verdien på stigningstallet til en rett linje
x	Variabel (inngangsstørrelse)
X_1	Verdi på inngangsstørrelsen x
X_2	Verdi på inngangsstørrelsen x
X_3	Verdi på inngangsstørrelsen x
X_4	Verdi på inngangsstørrelsen x
X_5	Verdi på inngangsstørrelsen x

Y_1	Navn på populasjon 1
Y_2	Navn på populasjon 2
Y_3	Navn på populasjon 3
Y_4	Navn på populasjon 4
Y_5	Navn på populasjon 5
$Y_{1,1} - Y_{1,15}$	15 avlesninger på instrumentresponser når instrumentet påføres X_1
$Y_{2,1} - Y_{2,15}$	15 avlesninger på instrumentresponser når instrumentet påføres X_2
$Y_{3,1} - Y_{3,15}$	15 avlesninger på instrumentresponser når instrumentet påføres X_3
$Y_{4,1} - Y_{4,15}$	15 avlesninger på instrumentresponser når instrumentet påføres X_4
$Y_{5,1} - Y_{5,15}$	15 avlesninger på instrumentresponser når instrumentet påføres X_5

Tabell 3: Terminologi benyttet i denne artikkelen

Neste artikkel

I den neste artikkelen skal vi se Residual. Dette er en teknikk for å bestemme hvor godt en «Best Fit»-linje kan beskrive dataparene.