Statistisk behandling av kalibreringsresultatene – Del 2.

v/ Rune Øverland, Trainor Elsikkerhet AS

Denne artikkelserien handler om statistisk behandling av kalibreringsresultatene.

I den første artikkelen så vi på hvordan estimere en populasjon basert på en eller flere stikkprøver (utvalg).

I denne artikkelen tar vi for oss den grafiske presentasjonen av innsamlede data ved kalibrering. Du får informasjon om hvordan vi beregner sentralverdien for en måleserie, og hvorledes vi kan uttrykke spredningen (variasjonen) av måleserien. Basert på sentralverdiene tegner vi opp en graf.

Den neste artikkelen har kalibreringskurve (lineær regresjon) som hovedtema. Dette er en teknikk for å estimere verdien av koeffisienter for en rett linje ($\hat{Y} = \hat{a}_1 \cdot x + \hat{a}_0$).

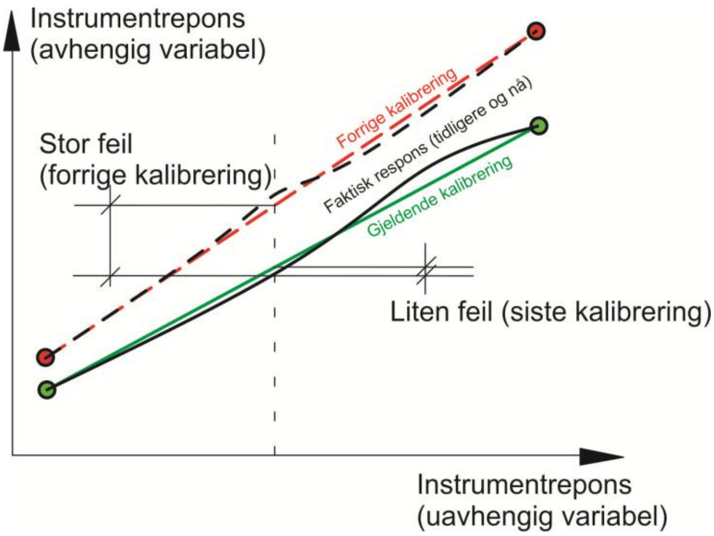
Den fjerde artikkelen tar for seg tilpasning, det vil si hvor godt vår kalibreringskurve er tilpasset verdiene vi fant ved kalibreringen. Vi bestemmer en korrelasjonskoeffisient (R^2).

Den femte, og sannsynligvis den siste artikkelen for denne gang, tar for seg et usikkerhetsintervall (Confidence interval) som vi legger rundt linjen. Den neste måleserien (kalibreringen) vil med 95 prosent sannsynlighet ligge innenfor dette usikkerhetsintervallet. Vi kommer også innom «Prediction Interval».

Kalibrering

Kalibrering handler om tillitt, og oftest mellom selger og kjøper. Instrumenter som har gal eller uakseptabel instrumentrespons gir feilmålinger ved salg, og produkter med feil kvalitet.

Kalibrering er sammenligning av en instrumentrespons mot responsen fra en annet referanseinstrument med kjent måleusikkerhet. Kalibreringen bestemmer avviket mellom de to responsene.



Figur 1: Et instrument er rekalibrert.

Vi har et scenario hvor instrumentresponsen, her representert med den svarte stiplede linjen er blitt endret. Instrumentresponsen er nå tilsvarende den svarte heltrukne linjen.

Før vi foretok ny kalibrering, hadde vi relativ stor avlesningsfeil (forskjell mellom rød stiplet linje og svart heltrukken linje).

Etter kalibrering fikk vi etablert en ny estimert lineær linje, her representert med grønn heltrukken linje. Avlesningsfeilen i dette er tilfelle er innenfor toleransegrensene for linearitet og måleusikkerhet.

Type A: Evaluering av måleusikkerhet

Type 'A' er en statistisk evaluering av instrumentresponser basert på en serie med observasjoner. En serie består av flere, uavhengige observasjoner av instrumentresponsen med de samme målebetingelsene, slik som vist i Figur 4.

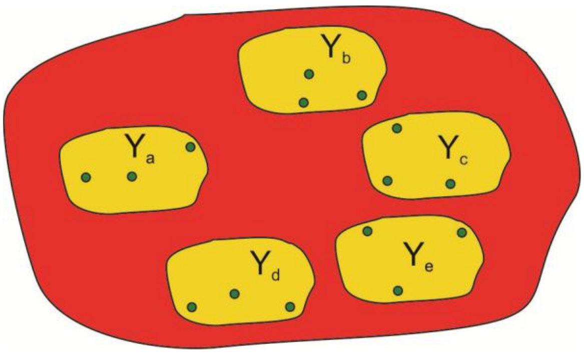
Noen begreper: elementer, utvalg og populasjon

Det som særmerker statistikken er at den anvendes på fenomener hvor det inngår usikkerhet, og hvor man ønsker å ta hensyn til denne ved behandling av dataene. Usikkerheten er ofte et resultat av en har et begrenset antall tilfeller av et større antall som en vet lite om på forhånd.

Grunnlaget for å få større viten er å foreta observasjoner (målinger) av de egenskapene ved instrumentet vi ønsker å undersøke. Vi utfører en rekke eksperimenter eller forsøk.

Generelt kan vi si; Elementene, som typisk er avlesninger av instrumentresponsene, er «byggesteinene» i statistikken. En serie med observasjoner utgjør et utvalg. Antall elementer i utvalget kalles utvalgsstørrelsen, og betegnes med symbolet n.

Gjør man flere serier (har flere måleserier), utgjør utvalgene en populasjon. En populasjon kan bestå av et bestemt antall utvalg, eller generelt sett et ubestemt antall utvalg. Populasjonsstørrelsen betegnes med symbolet N.



Figur 2: Vi tenker oss et en populasjon (rødt område som vi for eksempel kan navngi Y_1) består av N = 15 elementer. I dette eksemplet har vi fordelt elementene på 5 utvalg (oransje områder hvor vi tilsvarende kunne navngitt dem Y_{1a}, Y_{1b}, Y_{1c}, Y_{1d}, og Y_{1e}). Vi håper at elementene (grønne prikker) i et utvalg har en egenskap som representerer egenskapen i populasjonen. Under en slik forutsetning er

det tilstrekkelig å teste egenskapene til et utvalg, for så å relatere dette til egenskapene for populasjonen. Da sparer vi både tid og krefter. Utvalgsstørrelsen er her n = 3, i stedet for 15.

Målet med våre undersøkelser er å komme frem til et utsagn (ligning) eller lovmessigheter som vi med en viss statistisk sannsynlighet kan ha håp om gjelder for hele populasjonen, selv om de bare bygger på undersøkelser av et utvalg.

Innhenting av elementer (datapar)

Vi skal nå innhente elementer, og danne utvalg og populasjoner.

Vi tenker oss en mengdemåler som tidligere har gjennomført en omfattende kalibrering.

I Microsoft Excel har vi satt kalibreringsdataene inn i en tabellform. Vi har kjørt til sammen fem serier, hver med N = 15 enkeltkalibreringer. Første serie hadde en påtrykt strømningsrate på X_1 = 100,00 m^3/h. Den første avlesningen ga en instrumentrespons på $y_{1.1}$ = 99,90 m^3/h. Den andre avlesningen ga en instrumentrespons på $y_{1.2}$ = 100,05 m^3/h. Den femtende avlesningen ga oss verdien $y_{1.15}$ = 99,86 m^3/h.

Denne prosedyren gjentok vi for strømningsratene X_2 = 300, X_3 = 500, X_4 = 700 og X_5 = 900 m3/h.

De avleste responsverdiene fra instrumentet er satt inn i tabell 1. Den siste avlesningen av responsverdien fra instrumentet var $y_{5.15}$ = 899,64 m^3/h.

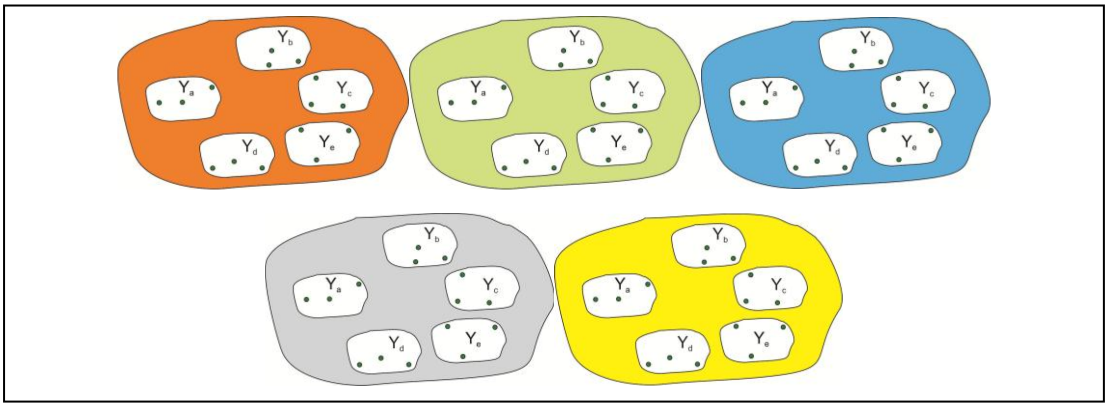
	X_1 = 100,00	X_2 = 300,00	X_3 = 500,00	X_4 = 700,00	X_5 = 900,00
1	99,90	300,45	500,83	700,25	899,50
2	100,05	300,07	500,22	700,33	899,33
3	100,17	300,00	500,11	698,80	900,45
4	99,99	299,66	499,45	700,55	901,30
5	99,86	299,80	499,30	700,22	899,56
6	99,94	299,90	499,70	700,11	899,32
7	100,01	299,65	499,90	699,85	900,32
8	100,12	300,25	499,99	699,81	900,98
9	100,05	300,28	500,50	700,05	899,78
10	100,01	300,01	500,11	700,09	899,23
11	99,88	299,95	501,00	699,50	900,23
12	100,12	299,73	499,26	699,25	900,00
13	100,00	300,22	499,83	700,68	900,25
14	100,04	300,13	499,72	700,45	900,11
15	99,86	299,90	500,08	700,06	899,64

Tabell 1: Kolonnevis har vi de fem måleseriene. For hver måleserie ble det gjort N = 15 avlesninger av strømningsraten.

Sub-populasjoner i stedet for populasjon

I vårt tilfelle har vi nå tilsammen 75 datapar; 15 stykk for strømningsraten X_1 = 100 m3/h, 15 stykk for strømningsraten X_2 = 300 m3/h og så videre.

For den videre statistiske behandlinger tar vi denne globale populasjonen og deler denne inn i fem sub-populasjoner.



Figur 3: Fem sub-populasjoner, én for de frem kalibreringspunktene. Og hver sub-populasjon består av 15 datapar, det vil si 15 elementer. I en senere artikkelen skal vi se på fordelen av å dele en sub-populasjon inn i flere utvalg. Her vises at hver sub-populasjon er delt inn i utvalg med 3 elementer.

Bestemme sentralverdien av en måleserie

Når verdien av en målestørrelse bestemmes ved å ta en serie enkeltmålinger, er beste estimat av målestørrelsen tatt for å være det aritmetiske gjennomsnittet av enkeltmålingene. Vi kaller dette for sentralverdien (μ_i) for populasjonen.

Dermed, hvis observasjonene (avlesning av instrumentresponsen) av målestørrelsen, Y_i, er $y_{i,1}$, $y_{i,2}$, $y_{i,3}$, ..., $y_{i,j}$, ... $y_{i,N}$, er det beste estimatet av Y_i:

$$\mu_{Yi} = \frac{y_{i,1} + y_{i,2} + \ ... + y_{i,N}}{N} = \frac{1}{N}\sum_{j=1}^{N} y_{i,j} \qquad [1]$$

For hver av kalibreringsratene bruker vi Excel-funksjonen = Gjennomsnitt (tall1;tall2) for enkelt å beregne sentralverdien (aritmetisk gjennomsnitt) av populasjonen.

B17 | =GJENNOMSNITT(B2:B16)

	A	B	C	D	E	F
1		X_1 =100,00	X_2 =300,00	X_3 =500,00	X_4 =700,00	X_5 =900,00
2	1	99,90	300,45	500,83	700,25	899,50
3	2	100,05	300,07	500,22	700,33	899,33
4	3	100,17	300,00	500,11	698,80	900,45
5	4	99,99	299,66	499,45	700,55	901,30
6	5	99,86	299,80	499,30	700,22	899,56
7	6	99,94	299,90	499,70	700,11	899,32
8	7	100,01	299,65	499,90	699,85	900,32
9	8	100,12	300,25	499,99	699,81	900,98
10	9	100,05	300,28	500,50	700,05	899,78
11	10	100,01	300,01	500,11	700,09	899,23
12	11	99,88	299,95	501,00	699,50	900,23
13	12	100,12	299,73	499,26	699,25	900,00
14	13	100,00	300,22	499,83	700,68	900,25
15	14	100,04	300,13	499,72	700,45	900,11
16	15	99,86	299,90	500,08	700,06	899,64
17	Gjennomsnitt	100	300	500	700	900,00

Figur 3: Excel-funksjonen =GJENNOMSNITT(B2:B16) beregner sentralverdien for verdiene i celleområdet B2 til og med B16. Funksjonen er skrevet inn i celle B17. Videre; i celle C17 beregnes sentralverdien μ_2.

I vår eksempel oppsummerer vi sentralverdiene slik:

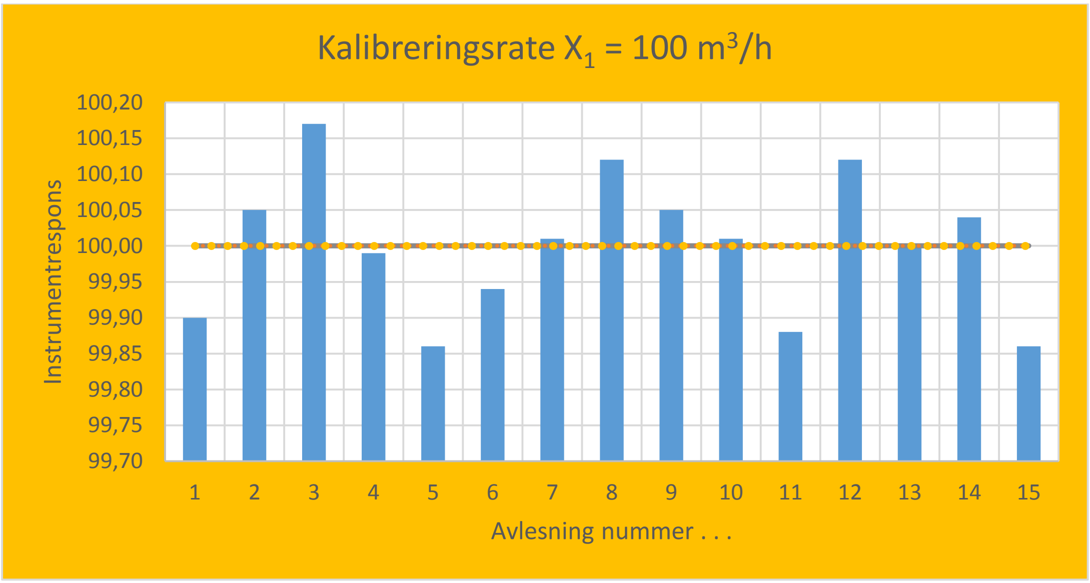
Sentralverdier (μ_i) for de ulike populasjonene (fremkommer i rad 17 i regnearket)				
μ_1 = 100 m^3/h	μ_2= 300 m^3/h	μ_3 = 500 m^3/h	μ_4= 700 m^3/h	μ_5 = 900 m^3/h

Tabell 2: Vi har kalkulert sentralverdiene for de fem populasjonene (måleseriene).

Grafisk presentasjon av populasjonen (Y_i) og sentralverdien (μ_i) for sub-populasjoner

Excel kan grafisk presentere dataene fra vår tabell.

Vi velger celleområdet [B2:B16] og klikker på menyen SETT INN, og velger KOMBINASJONSDIAGRAM. Diagrammet har jeg konfigurert; lagt på litt tekster på aksene, samt lagt inn en linje som representerer sentralverdien for μ_1-populasjonen.

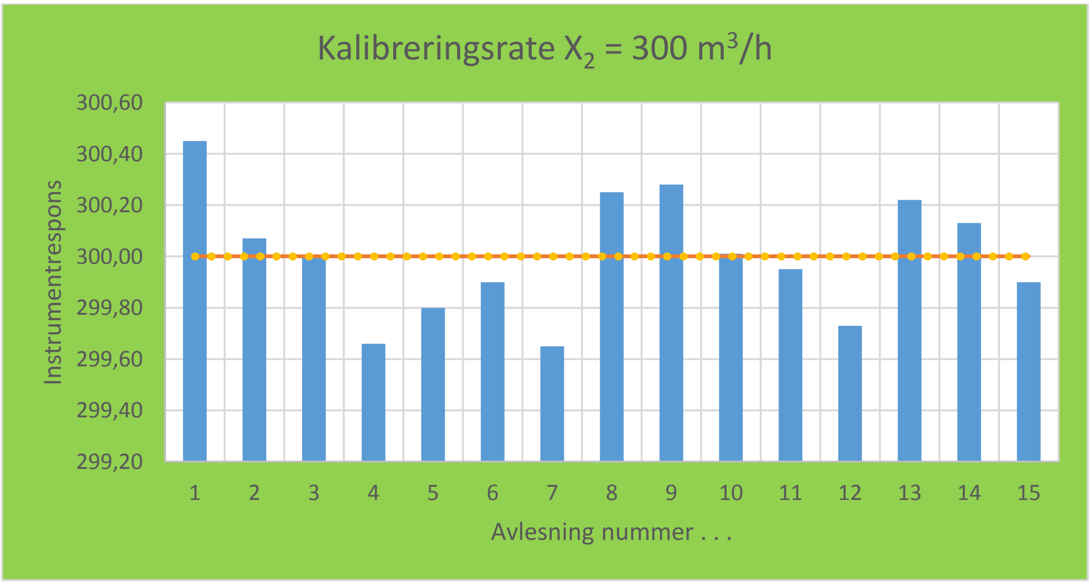


Figur 4: Grafisk presentasjon av responsdataene Y_1 fra mengdemåleren basert på en strømningsrate lik X_1= 100 m^3/h (sub-populasjon 1). Instrumentresponsene på y-aksen har små, tilfeldige variasjoner omkring sentralverdien μ_1.

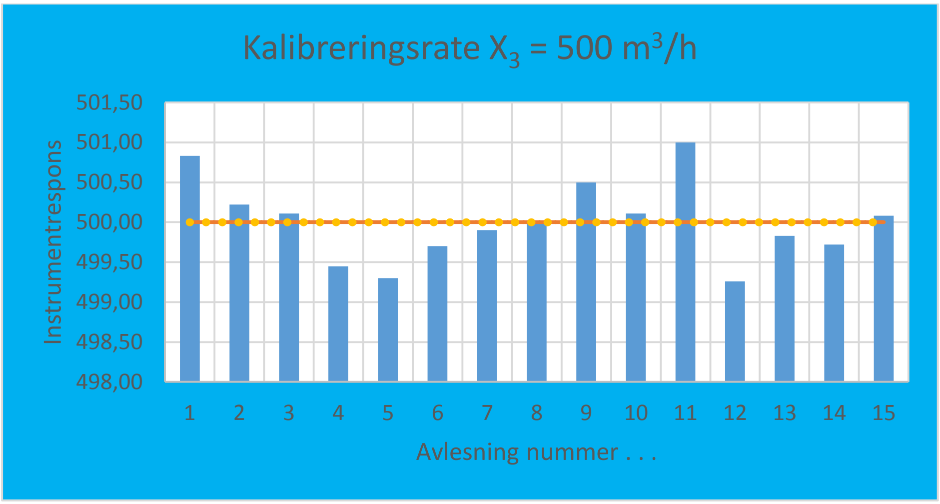
Vi har to typer av klassifisering av usikkerhetsbidrag; Type A og Type B.

Ved Type A gjøres en serie med observasjoner hvor måleoppsettet ikke endres. Variasjonen av enkeltmålingene av en serie observasjoner, gjort under identiske betingelser, bestemmes ved bruk av en statistisk analytisk metode. Vi får altså tak i de eksterne usikkerhetene. Vi trenger med andre ord ikke kjenne til de interne kildene til usikkerhetene.

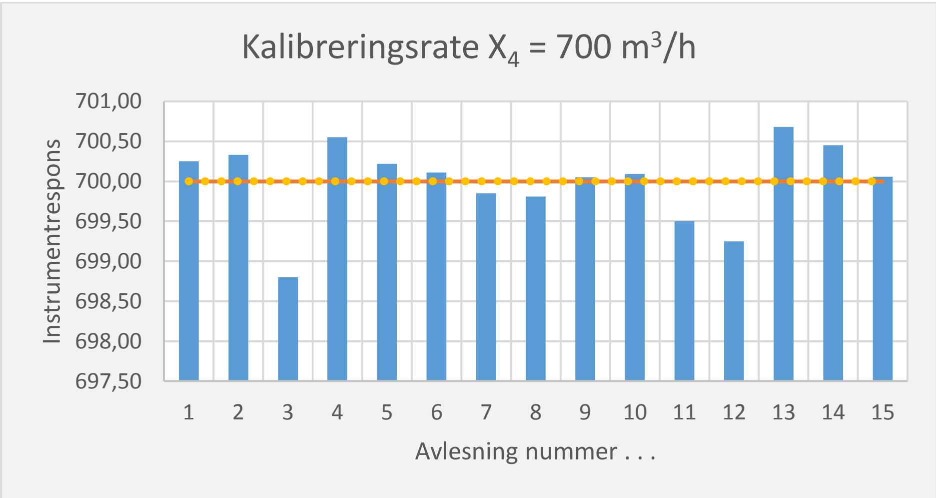
Figurene 4, 5, 6 og 7 viser de grafiske presentasjonene av måleseriene for strømningsratene $X_2 = 300$, $X_3 = 500$, $X_4 = 700$ og $X_5 = 900$ m^3/h.



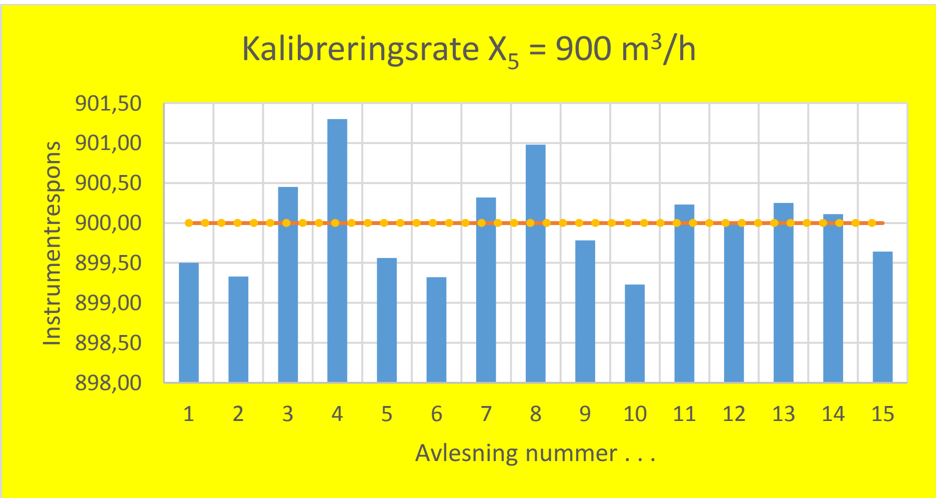
Figur 5: *Grafisk presentasjon av responsdataene Y_2 fra mengdemåleren basert på en strømningsrate lik X_2 = 300 m^3/h (sub-populasjon 2).*



Figur 6: *Grafisk presentasjon av responsdataene Y_3 fra mengdemåleren basert på en strømningsrate lik $X_3 = 500$ m^3/h (sub-populasjon 3).*



Figur 7: *Grafisk presentasjon av responsdataene Y_4 fra mengdemåleren basert på en strømningsrate lik $X_4 = 700$ m^3/h (sub-populasjon 4).*



Figur 8: *Grafisk presentasjon av responsdataene Y_5 fra mengdemåleren basert på en strømningsrate lik $X_5 = 900$ m^3/h (sub-populasjon 5).*

Elementvariasjon i en sub-populasjon

Det neste vi skal se på et i hvilken grad det er variasjon i våre fem sub-populasjoner. Vi la Excel beregne dette for oss med funksjonen STAAV.P.

B18 =STDAV.P(B2:B16)

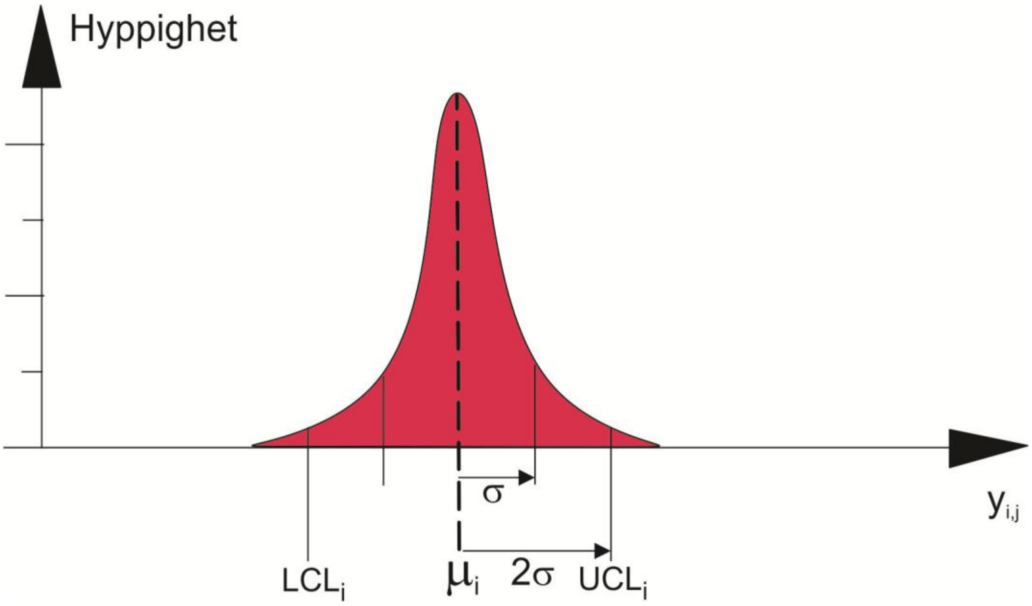
	A	B	C	D	E	F
1		X_1 = 100,00	X_2 = 300,00	X_3 = 500,00	X_4 = 700,00	X_5 = 900,00
2	1	99,90	300,45	500,83	700,25	899,50
3	2	100,05	300,07	500,22	700,33	899,33
4	3	100,17	300,00	500,11	698,80	900,45
5	4	99,99	299,66	499,45	700,55	901,30
6	5	99,86	299,80	499,30	700,22	899,56
7	6	99,94	299,90	499,70	700,11	899,32
8	7	100,01	299,65	499,90	699,85	900,32
9	8	100,12	300,25	499,99	699,81	900,98
10	9	100,05	300,28	500,50	700,05	899,78
11	10	100,01	300,01	500,11	700,09	899,23
12	11	99,88	299,95	501,00	699,50	900,23
13	12	100,12	299,73	499,26	699,25	900,00
14	13	100,00	300,22	499,83	700,68	900,25
15	14	100,04	300,13	499,72	700,45	900,11
16	15	99,86	299,90	500,08	700,06	899,64
17	Gjennomsnitt	100	300	500	700	900,00
18	Standard avvik	0,09	0,23	0,49	0,48	0,59

Figur 9: Excel-funksjonen =STDAV.P beregner standard avvik for en populasjon. Funksjonen er skrevet inn i rad 18 i vårt eksempel. Det beregnes verdien på ett standard avvik slik beskrevet av en normalfordeling (Gauss-kurve).

Når variasjonen følger normalfordelingskurven, også omtalt som Gausskurven, vil cirka 68,3 prosent av populasjonen ligge innenfor et intervall begrenset av ett standardavvik.

Standard avvik for populasjonen er kvadratroten av variansen.

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(y_{i,j} - \mu_{Yi})^2} \qquad [2]$$



Figur 10: Basert på normalfordelingskurven vil ca 95 prosent av elementene i populasjonen ligge innenfor ± 2 standardavvik (± 2σ). LCL (Lower Confidence Limit) definerer nedre grenseverdi, mens UCL (Upper Confidence Limit) definerer øvre grenseverdi i usikkerhetsintervallet.

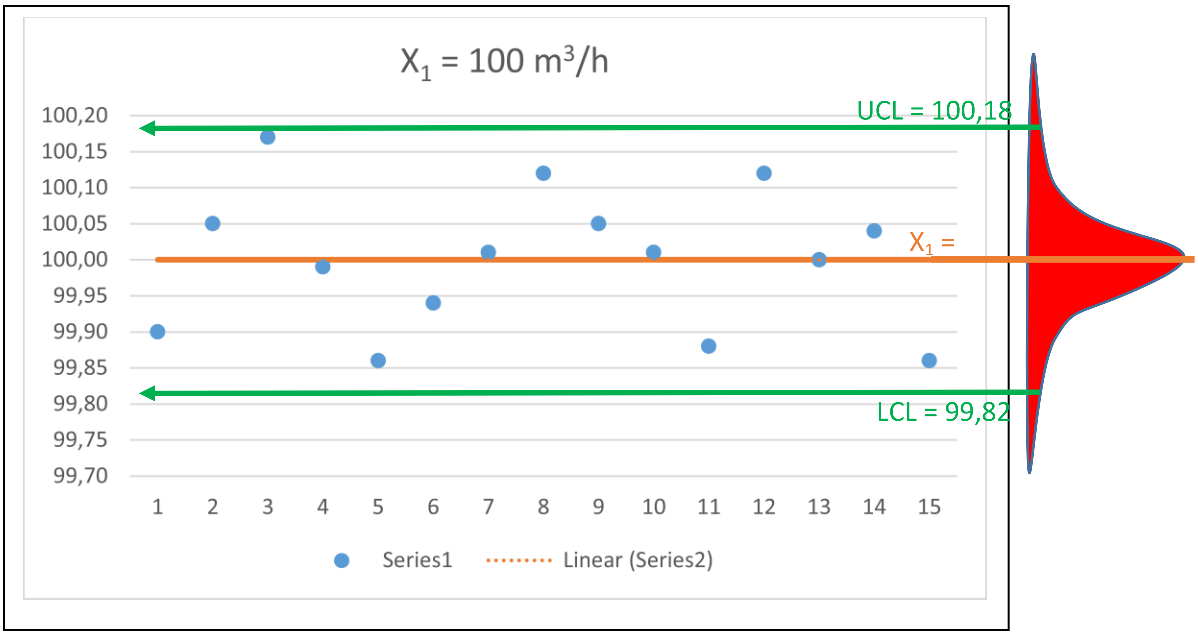
Vi kan oppsummere dette i en tabell slik:

Utvidet usikkerhetsområde (CI = 95 %) for de ulike populasjonene (dekningsfaktor z = 1,96 ≈ 2) U = Utvidet usikkerhetsområde = 2 · Standard avvik									
μ_1= 100 m³/h σ_1 = 0,09		μ_2= 300 m³/h σ_2 = 0,23		μ_3 = 500 m³/h σ_3 = 0,49		μ_4= 700 m³/h σ_4 = 0,48		μ_5= 900 m³/h σ_5 = 0,59	
LCL_{y1}	UCL_{y1}	LCL_{y2}	UCL_{y2}	LCL_{y3}	UCL_{y3}	LCL_{y4}	UCL_{y4}	LCL_{y5}	UCL_{y5}
$\mu_1 - 2\sigma_1$ = 100 – 0,18 = 99,82	$\mu_1 + 2\sigma_1$ = 100 + 0,18 = 100,18	$\mu_2 - 2\sigma_2$ = 300 – 0,46 = 299,54	$\mu_2 + 2\sigma_2$ = 300 + 0,46 = 300,46	$\mu_3 - 2\sigma_3$ = 500 – 0,98 = 499,02	$\mu_3 + 2\sigma_3$ = 500 + 0,98 = 500,98	$\mu_4 - 2\sigma_4$ = 700 – 0,96 = 699,04	$\mu_4 + 2\sigma_4$ = 700 + 0,96 = 700,96	$\mu_5 - 2\sigma_5$ = 900 – 1,08 = 898,92	$\mu_5 + 2\sigma_5$ = 900 + 1,08 = 901,08

Tabell 2: Kalkulasjon av øvre grenseverdi for usikkerhetsintervall (UCL: Upper Confidence Limit) og nedre grenseverdi for usikkerhetsintervall (LCL: Lower Confidence Limit) for hver av de fem sub-populasjonene. I dette tilfellet er omfatter usikkerhetsintervallet 95 prosent av måleserien.

Grafisk fremstilling av utvidet usikkerhet for en sub-populasjon

Figuren illustrerer plasseringen av den øvre og nedre usikkerhetslinjen for måleserien kalibrering med X_i = 100 m^3/h.



Figur 11: Grafisk presentasjon av data for sub-populasjon 1. Vi har tegnet inn en normalfordelings-kurve til høyre. De grønne linjene er 2 standard avvik fra populasjonens sentralverdi (gjennomsnittsverdi).

Grafisk ser vi at alle våre (15) målepunkter ligger mellom UCL og LCL. Vår distribusjon av observerte data følger ikke en perfekt normalfordelingen. Teorien for en normal-fordeling er at enkeltverdier hyppigst forekommer ved eller nærheten av sentralverdien. Hyppigheten av enkeltverdier faller desto lengre vekk fra sentralverdien vi beveger oss.

Overfor UCL og nedenfor LCL representerer i dette tilfellet 5 prosent av populasjonen, i og med at 95 prosent skal ligge innenfor.

Dette kan vi uttrykke slik:

$CL_{1-\alpha}$ [3]

I vårt eksempel er α = 0,05.

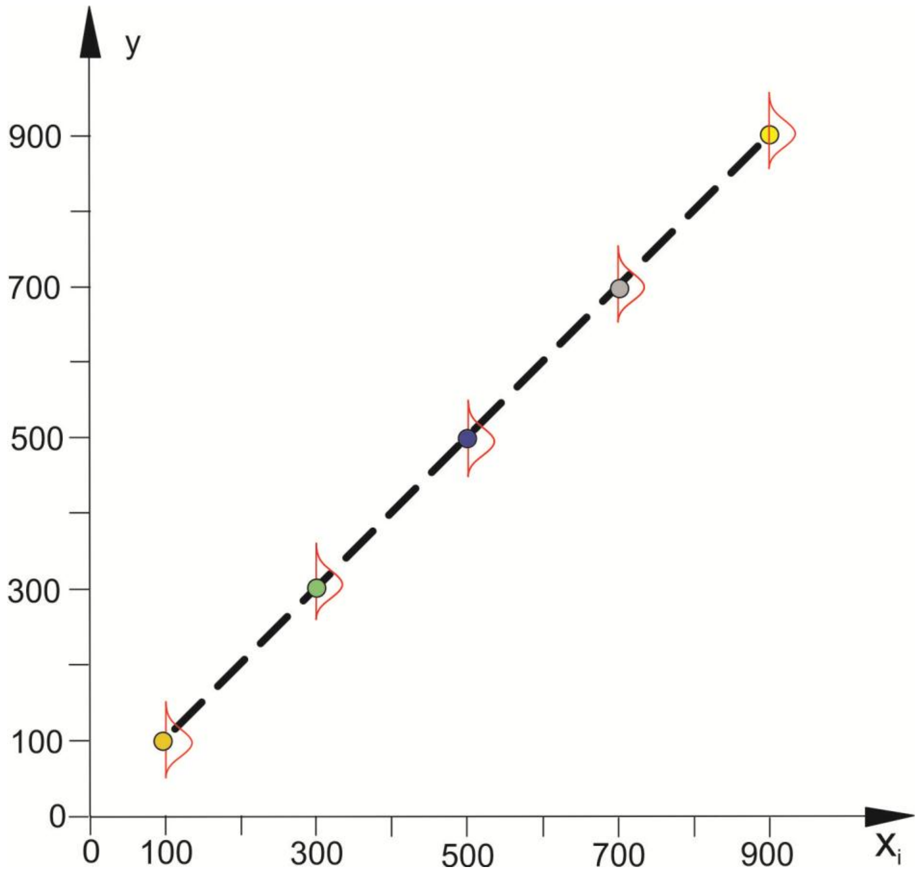
0,05 tilsvarer som kjent 5 prosent.

Vi kan derfor uttrykke det slik:

$CL_{1-0,05} = Cl_{0,95}$ [4]

Trekke en linje gjennom sentralverdiene til de ulike populasjonene

Vi trekker en linje gjennom tallparene (X_1, μ_1), (X_2, μ_2), (X_3, μ_3), (X_4, μ_4) og (X_5, μ_5) for de 5 sub-populasjonene. I vårt tilfelle vil dette være en tenkt rett linje da vi forutsetter linearitet i området X_1 til og med X_5.



Figur 12: Vi kan tegne inn en tenkt rett linje som går gjennom sentralverdiene for hver sub-populasjon. For hver av datasettene har vi tegnet inn Gausskurven (røde kurver), som vi har oppskalert. Fra tabell 2: For eksempel så hadde vi for µ1 verdiene LCL_{y1} = *99,82* m^3/h *og* UCL_{y1} = *100,18. I denne generelle figuren ser det ut som* UCL_{y1} *ligger rundt 150, mens* LCL_{y1} *ligger rundt 50. Legg for øvrig merke til at usikkerheten foreløpig kun er knyttet til y-verdiene.*

Linjen kan beskrives med følgende likning:

$\mu_{y|x} = a_1 \cdot x + a_0$ [5]

Den sanne gjennomsnittlige Y for en gitt verdi av x er lik koeffisienten a_0 pluss koeffisienten a_1 multiplisert med x. Koeffisienten a_0 har verdien lik det punktet linjen skjærer Y-linjen, og koeffisienten a_1 er stigningstallet (vinkelen) for linjen.

Imidlertid, i praksis kjenner vi ikke populasjonen til $X_1 = 100$, $X_2 = 300$, $X_3 = 500$, $X_4 = 700$ og $X_5 = 900$ m3/h! Da måtte vi i så fall gjøre uendelig antall kalibreringer (∞) for hvert av X_1 og så videre; ikke nøye oss med N = 15. Med referanse til Figur 1: Så langt i resonnementet har jeg betraktet et utvalg som hele populasjonen, og jeg skulle ikke nøyd meg med N = 15! Og, vi kan derfor heller ikke tegne opp den tenkte rette linjen heller! Egenskapene til datasettet $x_{1,1}$ til $x_{1,15}$ tilsvarer ikke helt sikkert egenskapene til datasettet $x_{1,1}$ til $x_{1,\infty}$.

Linjen vil i praksis alltid være omgitt av Type A usikkerheter; det vil si små, tilfeldige variasjoner (målestøy).

Vi kan derfor *estimere* en linje, og denne uttrykker vi slik:

$$\hat{Y} = \mu_{y|x} + \varepsilon = (a_1 \cdot x + a_0) + \varepsilon \qquad [6]$$

Den estimerte verdien $\hat{Y}$ er lik den sanne verdien pluss en feilstørrelse lik ε.

Den nye likningen omskrives slik:

$$\hat{Y} = \hat{a}_1 \cdot x + \hat{a}_0 \qquad [7]$$

Terminologi

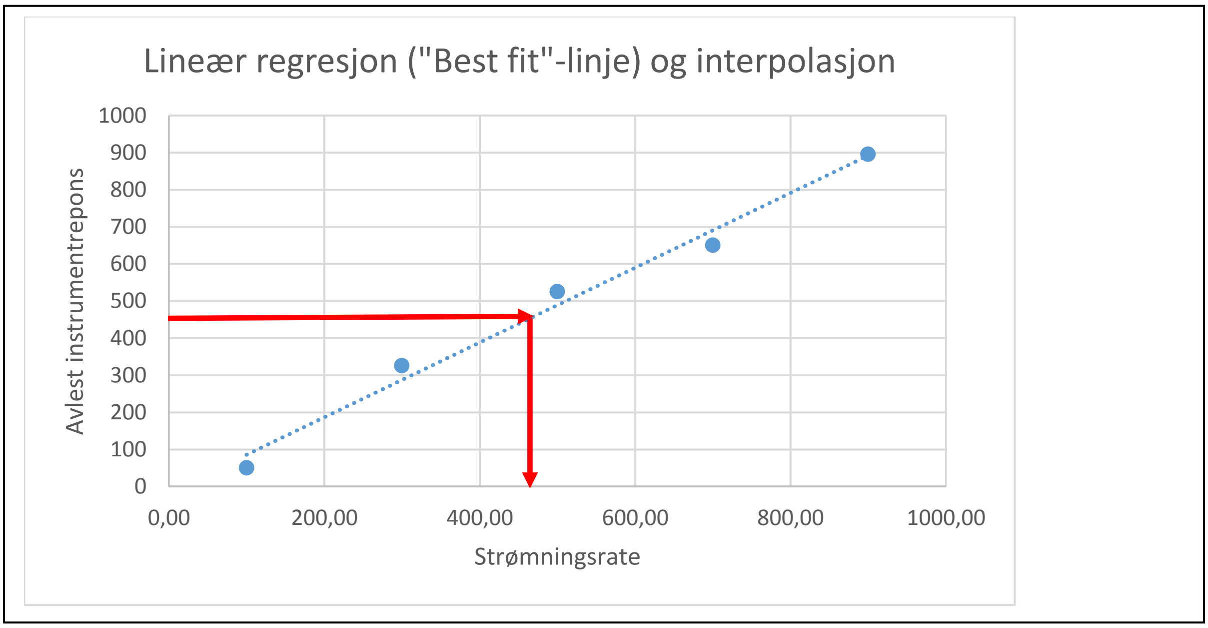
∞	Uendelig
a_0	Koeffisient. Har verdien y når x = 0.
$\hat{a}_0$	Estimert koeffisient. Har verdien y når x = 0.
a_1	Koeffisient. Har verdien på stigningstallet til en rett linje
$\hat{a}_1$	Estimert koeffisient. Har verdien på stigningstallet til en rett linje
CL	Confidence Interval
j	Teller, vanligvis i forbindelse med et summasjonstegn
LCL	Lower Confidence Limit
n	Antall elementer i et utvalg
N	Antall elementer i en populasjon
U	Utvidet usikkerhet (= 2 · Standard avvik)
UCL	Upper Confidence Limit
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
Y_{1a}	Navn på et utvalg i populasjonen Y_1
Y_{1b}	Navn på et utvalg i populasjonen Y_1
Y_{1c}	Navn på et utvalg i populasjonen Y_1
Y_{1d}	Navn på et utvalg i populasjonen Y_1
Y_{1e}	Navn på et utvalg i populasjonen Y_1
$\hat{Y}$	Variabel. Innehar forventningsverdien.

z	Dekningsfaktor (= 1,96 ≈ 2 når konfidensnivået er 95 prosent).
α	Koeffisient. Innehar verdien på andelen av elementet som er over UCL og under LCL.
ε	Feil
μ_1	Sentralverdi (aritmetisk gjennomsnittsverdi) av populasjon Y_1
μ_2	Sentralverdi (aritmetisk gjennomsnittsverdi) av populasjon Y_2
μ_3	Sentralverdi (aritmetisk gjennomsnittsverdi) av populasjon Y_3
μ_4	Sentralverdi (aritmetisk gjennomsnittsverdi) av populasjon Y_4
μ_5	Sentralverdi (aritmetisk gjennomsnittsverdi) av populasjon Y_5
μ_{Yi}	Sentralverdi populasjon Y_i
$\mu_{y\|x}$	Den sanne gjennomsnittlige Y for en gitt verdi av x
σ	Standard avvik for en populasjon
σ^2	Varians for en populasjon
σ_1	Standard avvik for populasjon Y_1
σ_2	Standard avvik for populasjon Y_2
σ_3	Standard avvik for populasjon Y_3
σ_4	Standard avvik for populasjon Y_4
σ_5	Standard avvik for populasjon Y_5

Neste artikkel . . . .

I den neste artikkelen skal vi se på lineær regresjon. Dette er en metode for å estimere verdiene for koeffisientene $\hat{a}_0$ og $\hat{a}_1$ slik at den rette linjen har en optimal tilpasning til vårt datasett.

Vi kommer også inn på interpolasjon, det vil si hvordan vi kan estimere verdien på inngangsstørrelsen x basert på kjennskap til den kalkulerte instrumentresponsen $\hat{Y}$.



Figur 13: Lineær regresjon $\hat{Y} = \hat{a}_1 \cdot x + \hat{a}_0$. Vi estimerer verdiene på koeffisientene $\hat{a}_0$ og $\hat{a}_1$ slik at en rett linje oppnår «Best fit» til våre datasett. Deretter kan vi interpolere slik at vi kan bestemme x-verdien ved kjennskap til en avlest instrumentrespons.