ISTx1002 Usikkerhet og støy i målinger

Feilforplantning i målesystemer; Kalibrering

Stefanie Muff, Institutt for matematiske fag, NTNU Trondheim

Oktober 24 og 30, 2023

Plan for i dag (15:15-16:00) og mandag 15:15-16:00

- Tema 1: Feilforplantning i målesystemer
- Tema 2: Kalibrering

Pensum og læringsressurser

Husk lenken til den eksterne modulsiden:

https://wiki.math.ntnu.no/istx1002/2023h/start

Pensum del 2:

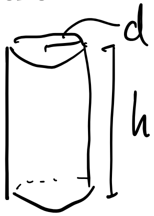
- **Korte videoer**: (by Charles H. A. Curry)
  - Målefunksjoner og kombinert standardavvik (7:08)
  - Kalibrering (5:06)
- Denne forelesningen
- **Disse slides** med alle notater og beregninger som er vist fram

Målesystemer

Et **målesystem** beskriver en sammenheng mellom én eller flere inngangsstørrelser og den størrelsen vi egentlig vi måle.

Eksempel:

- Vi skal måle volumet til en gjenstand som er en sylinder med høyde h og diameter d. Volumet er gitt ved



$$V(d,h) = \pi d^2 h \ .$$

- **Målefunksjonen** defineres som $\underbrace{V(d,h)}_{f(d,h)}$ og ses som funksjon av variablene d og h.

Feilforplantning i målesystemer

$u(X) = \sigma(X)$

Vi har sett på standard usikkerhet i en eneste variabel. Nå skal vi gjøre ting litt mer kompliserte:

$X_1 + X_2$

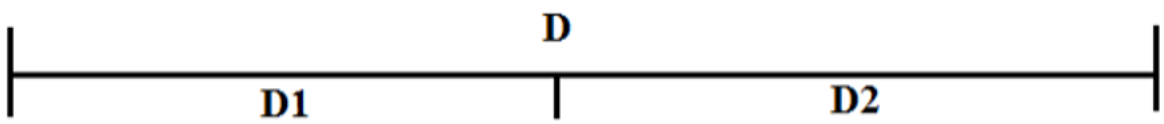
1) Usikkerhet i en lineær *målefunksjon* av flere variabler
2) Usikkerhet i en ikke-lineær *målefunksjon* av en variabel $f(X) = X^2$
3) Usikkerhet i en ikke-lineær *målefunksjon* av flere variabler $f(X, Y)$ for ukorrelerte X, Y
4) Usikkerhet i en ikke-lineær *målefunksjon* av flere variabler $f(X, Y)$ for korrelerte X, Y

$V(d, h)$

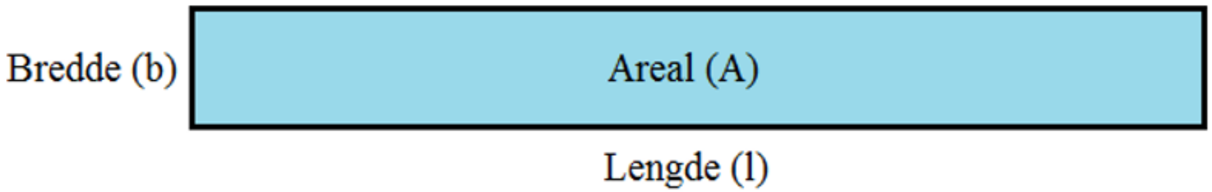
Kombinert standard usikkerhet (eller “feilforplantning”)

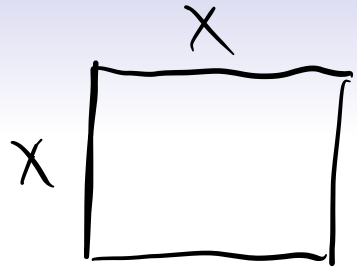
Eksempler:

- Mål summen av flere deler (X_1, X_2,..), slik at du tar $X_1 + X_2 + ... + X_n$ (oppgave 3a og b i prosjektet).



- Mål lengde (l) og bredde (b) til en rektangel og beregn arealet $A = f(l, b) = l \cdot b$ (oppgave 3c-e i prosjektet).





$f(X) = X^2$

- Mål en størrelse X, men så er du interessert i en transformert versjon, for eksempel X^2 (tenk at du vil måle arealeat til en firkant).

- Mål masse m og hastighet v av et objekt og beregn bevegelsesenergi $\underbrace{f(m,v)} = \frac{1}{2}mv^2$.

Og alle størrelser vi måler har forskjelle grader av usikkerhet som "forplanter" seg videre til den endelige størrelsen vi egentlig vil måle.

1) Lineære kombinasjoner av variabler

P1 P2

Problemstilling: Vi har to variabler, X_1 og X_2, som måler steglengden til to forskjellige personer. Vi vet at de typiske steglengdene er $\mu_1 = 75$cm og $\mu_2 = 82$cm med standard usikkerhet $u(X_1) = 6$ og $u(X_2) = 9$cm. Vi antar steglengdene er uavhengige ($\mathrm{Cov}(X_1, X_2) = 0$).

a) Hva er forventningsverdien og standard usikkerheten i summen av 5 uavhengige steg gått av person 1?

$$f(X_1) = 5 \cdot X_1 \qquad u(f(X_1)) = ?$$

b) Hva er standard usikkerheten i gjennomsnittig steglengde til de to personene?

$$f(X_1, X_2) = \frac{X_1 + X_2}{2} \qquad u(f(X_1, X_2)) = ?$$

Generell regel:

For en kombinert størrelse

$$Y = f(X_1, \ldots, X_n) = \overbrace{a_1 X_1 + a_2 X_2 + \ldots + a_n X_n}^{\text{lineær kombinasjon}}$$

konstant tall

kan vi beregne variansen som

$$\mathrm{Var}(Y) = a_1^2 \mathrm{Var}(X_1) + a_2^2 \mathrm{Var}(X_2) + \ldots + a_n^2 \mathrm{Var}(X_n)$$
$$\boxed{+ 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathrm{Cov}(X_i, X_j)} \ .$$

hvis $\mathrm{Cov}(X_i, X_j) = 0$

$\Rightarrow$ så kan vi glemme den delen.

Husk: $\mathrm{Var}(X) = u^2(X)$, fordi standard usikkerheten er bare standardavviket av størrelsen.

Nå kan vi beregne løsningen for problemstillingen ovenfor:

$\mu_1 = 75\text{cm}, \; u(X_1) = 6\text{cm} \Rightarrow var(X_1) = 36$

$\mu_2 = 82\text{cm}, \; u(X_2) = 9\text{cm} \Rightarrow var(X_2) = 81$

a) $f(X_1) = 5 \cdot X_1 \Rightarrow E(f(X_1)) = 5 \cdot \underbrace{E(X_1)}_{75\text{cm}} = 5 \cdot 75 = \underline{\underline{375\text{cm}}}$

$var(f(X_1)) = 25 \cdot var(X_1) = 25 \cdot 36$

$\Rightarrow u(f(X_1)) = \sqrt{25 \cdot 36} = 5 \cdot 6 = \underline{\underline{30\text{ cm}}}$

b) $f(X_1, X_2) = \frac{1}{2} \cdot X_1 + \frac{1}{2} \cdot X_2 \; ; \; E(f(X_1, X_2)) = \frac{1}{2} \cdot 75 + \frac{1}{2} 82 = \underline{\underline{78.5\text{cm}}}$

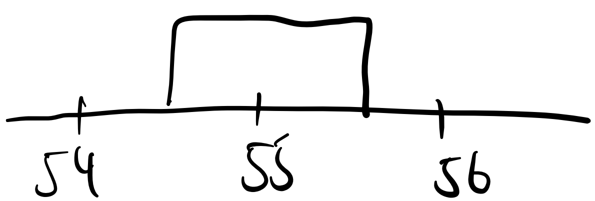
$var(f(X_1, X_2)) = \frac{1}{2^2} var(X_1) + \frac{1}{2^2} var(X_2) = \frac{1}{4}(36 + 81)$

$\Rightarrow u(f(X_1, X_2)) = \sqrt{\frac{1}{4}(36+81)} = \frac{1}{2}\sqrt{36+81} = \underline{\underline{5.41\text{ cm}}}$

2) Transformert versjon av en variabel

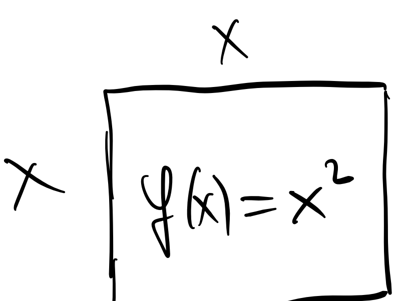
Problemstilling: Du måler lengden X til en firkant med 55cm (nøyaktigheten er bare 1cm), og så vil du beregne arealet til firkanten. Hva er standard usikkerheten i arealmålingen?

- 55cm med en standard usikkerhet (standardavvik) som er firkantfordelt, og derfor...



$$\Rightarrow u(X) = \frac{0.5}{\sqrt{3}} = 0.289$$

- Men hvordan bruker vi det når vi må gange X med seg selv...?



$$u(X^2) = ?$$

Usikkerheten i en transformert variabel $f(X)$

Generell regel:
Når vi har en variabel X som vi kan måle og vet usikkerheten $u(X)$, men vi er interessert i en transformert version $f(X)$ (for eksempel $f(X) = X^2$, $\log(X)$,...), kan vi bruke det følgende:

$f(X) = e^X$

Trick: Taylor approximasjon
For en gitt måleverdi μ har vi omtrent

$$\begin{aligned} f(X) &\approx f(\mu) + f'(\mu)(X - \mu) \\ &= \underbrace{[f(\mu) - f'(\mu)\mu]}_{a} + \underbrace{f'(\mu)}_{b} \cdot X \end{aligned}$$

$a + b \cdot X$

fra

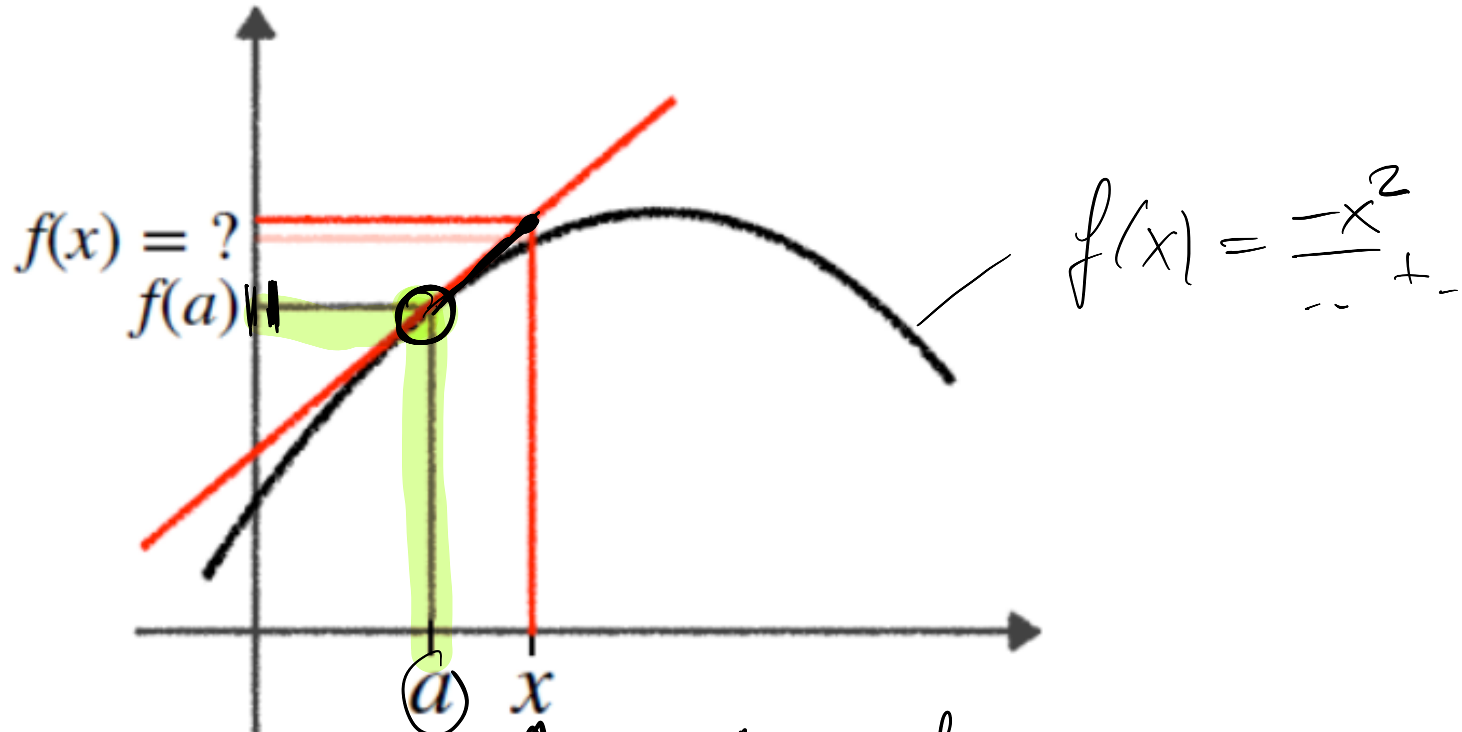
Derfor har vi omtrent var til $f(X)$

$$\mathrm{Var}(f(X)) = \underbrace{f'(\mu)^2}_{b^2} \cdot \underbrace{\mathrm{Var}(X)}_{u^2(X)} .$$

Eller: $\quad u(f(X)) = \sqrt{f'(\mu)^2 \cdot u^2(X)}$.

$\left[\mathrm{var}(a + b \cdot X) = b^2 \cdot \mathrm{var}(X) \right]$

Visualisering: Taylor approximasjon i 1 dimension:

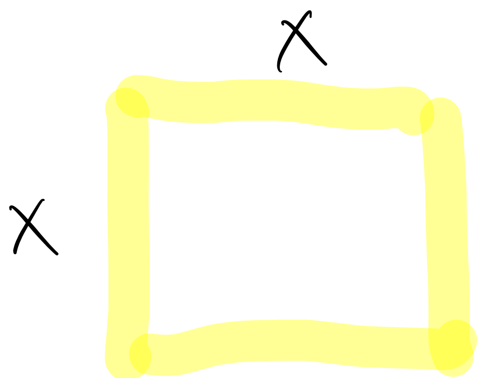


$\Delta y = \Delta f$

$$f(x) \approx f(a) + \boxed{f'(a)} \cdot (x-a)$$

$f'(a)$ heter følsomhet eller følsomhetsfaktor.

Nå kan vi beregne arealet og approximert usikkerhet til arealet for firkanten:



$$x_0 = 55\,\text{cm} \qquad u(X) = 0.289\,\text{cm}$$

$$A = x_0 \cdot x_0 = 3025\,\text{cm}^2$$

$$f(x) = x^2$$

$$f'(x) = 2 \cdot x$$

$\Rightarrow$ Derfor

$$u(f(X))\Big|_{x_0} = \sqrt{\left(f'(x_0)\right)^2 \cdot u^2(X)}$$

$$= \sqrt{(2x_0)^2 \cdot 0.289^2}$$

$$= \sqrt{(2 \cdot 55)^2 \cdot 0.288^2} = 31.75\,\text{cm}^2$$

Relativ usikkerhet

$$\left.\frac{u(f(x))}{f(x)}\right|_{x=55} = \frac{31.75}{55^2} = 0.0105$$

$$\approx 1\%$$

3) Ikke-lineær *kombinasjon* av flere variabler $f(X_1, X_2)$, med ukorrelerte X_1, X_2

Problemstilling: Vi vil beregne bevegelsesenergi til et objekt og måler

- masse $m_0 = 0.45, \overbrace{u(m)}^{\sigma(m)} = 0.01kg$
- hastighet $v_0 = 10.8, u(v) = 0.05m/s$

$\sigma(v)$

Hva er usikkerheten i den beregnete begevelsesenergien?

$$f(m, v) = \frac{1}{2}mv^2$$

Hvordan går vi frem i flere dimensioner?

Generell regel:

Trick: Taylor approximasjon (igjen) for flere dimensioner

For gitte måleverdier μ_1 og μ_2 av en funksjon $f(X_1, X_2)$ har vi omtrent

Var?

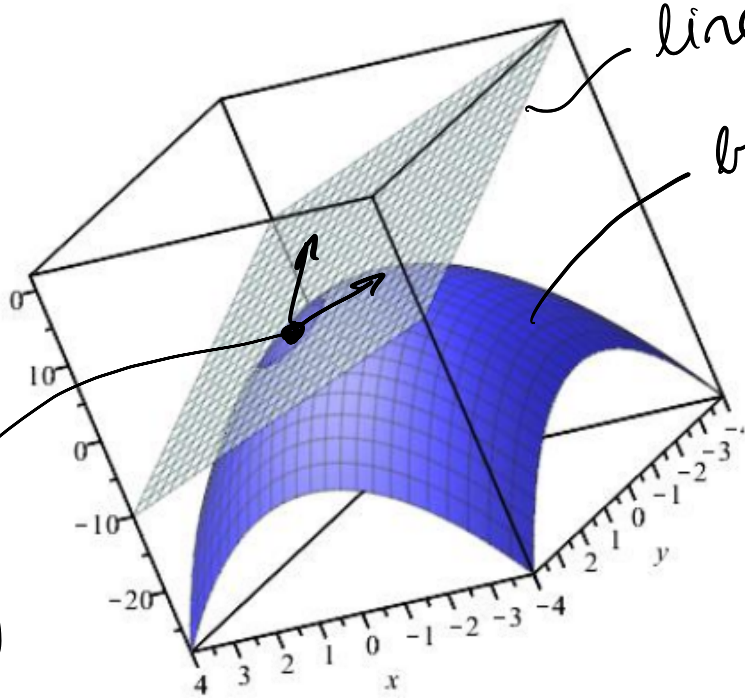
$$\underline{f(X_1, X_2)} \approx \boxed{f(\mu_1, \mu_2)} + \underbrace{\frac{\partial f}{\partial X_1}(\mu_1, \mu_2)}_{a}(X_1 - \mu_1) + \underbrace{\frac{\partial f}{\partial X_2}(\mu_1, \mu_2)}_{b}(X_2 - \mu_2)$$

$u(X_1), u(X_2) \longrightarrow u(f(X_1, X_2))$

Visualisering: Taylor approximasjon 2 dimensioner:

lineær funksjon

buet funksjon



f(a,b)

$$f(x,y) \approx f(a,b) + \frac{\partial f}{\partial x}(a,b)\cdot(x-a) + \frac{\partial f}{\partial y}(a,b)\cdot(y-b)$$

$\frac{\partial f}{\partial x}(a,b)$ og $\frac{\partial f}{\partial y}(a,b)$ heter følsomhet eller følsomhetsfaktorer.

Og derfor er

$$\mathrm{Var}(f(X_1, X_2)) \approx \underbrace{\left(\frac{\partial f}{\partial X_1}(\mu_1, \mu_2)\right)^2}_{a^2} \mathrm{Var}(X_1) + \underbrace{\left(\frac{\partial f}{\partial X_2}(\mu_1, \mu_2)\right)^2}_{b^2} \mathrm{Var}(X_2)$$

$$\Rightarrow u(f(X_1, X_2)) \approx \sqrt{a^2 \mathrm{Var}(X_1) + b^2 \mathrm{Var}(X_2)}$$

Nå kan vi beregne bevegelsesenergien og approximert usikkerhet:

$m_0 = 0.45$, $u(m) = 0.01\,\text{kg}$

$v_0 = 10.8\,\frac{m}{s}$, $u(v) = 0.05\,\frac{m}{s}$

$f(m,v) = \frac{1}{2} \cdot m \cdot v^2$

$\underline{\underline{f(m_0, v_0)}} = \frac{1}{2} m_0 v_0^2 = \frac{1}{2} 0.45\, 10.8 = \underline{\underline{26.24}}\ \frac{km^2}{s^2}$

Vi trenger: $\frac{\partial f}{\partial m} = \frac{1}{2} \cdot v^2 \qquad \frac{\partial f}{\partial v} = m \cdot v$

$\Rightarrow \left.\frac{\partial f}{\partial m}\right|_{m_0, v_0} = \frac{1}{2} v_0^2 = 58.32, \quad \left.\frac{\partial f}{\partial v}\right|_{m_0 v_0} = m_0 v_0 = \underline{\underline{4.86}}$

$\Rightarrow$ Kombinert standardusikkerhet:

$$\underline{\underline{u(f(m_0,v_0))}} = \sqrt{\underbrace{\left(\frac{\partial f}{\partial m}\bigg|_{m_0,v_0}\right)}^{2} \cdot u^2(m) + \left(\frac{\partial f}{\partial v}\bigg|_{m_0,v_0}\right)^2 \cdot u^2(v)}$$

$$= \sqrt{58.32^2 \cdot 0.01^2 + 4.86^2 \cdot 0.5^2} = \underline{\underline{2.50}} \frac{kg\,m^2}{s^2}$$

$k=2$ $\qquad f(m_0,v_0) = 26.24 \frac{kg\,m^2}{s^2}$

med dekningsintervall $(21.24, 31.24)$

$-2u(\ldots)$ $\qquad +2u(\ldots)$

4) Ikke-lineær *kombinasjon* av flere variabler $f(X, Y)$, med korrelerte X, Y

X_1 X_2

Generell regel:

$$\begin{aligned}\mathrm{Var}(f(X_1, X_2)) \approx & \left(\frac{\partial f}{\partial X_1}(\mu_1, \mu_2)\right)^2 \mathrm{Var}(X_1) \\ & + \left(\frac{\partial f}{\partial X_2}(\mu_1, \mu_2)\right)^2 \mathrm{Var}(X_2) \\ & \boxed{+ 2 \cdot \frac{\partial f}{\partial X_1} \frac{\partial f}{\partial X_2} u(X_1, X_2)} \, ,\end{aligned}$$

kan bli ganske stor!

hvor $u(X_1, X_2) = u(X_1)u(X_2)\mathrm{Cor}(X_1, X_2)$.

Det må dere ikke lære utenat.

Men husk: sterkere korrelasjoner mellom X_1 og X_2 betyr større avvik fra tilfelle 3) hvor vi hadde uavhengige variabler.

Kalibrering – grunnlegende idé

- Prosessen hvor vi presiserer forhold mellom målte og ekte verdier (eller nominell og ekte verdi).
- Retting av **systematisk feil**, for eksempel i et måleinstrument.
- For å kalibrere, sammenligner vi resultatene av målinger fra et instrument med verdier fra en annen kilde som vi vet er mer nøyaktig (3-5 ganger nøyaktigere).
- Eksempler: Mål en kjent kilde på 100V, en kjent lengde på 10m eller en kjent vekt på 1kg.

Kalibrering eksempel: Voltmeter

Vi måler en kjent kilde på 100V med to voltmetere, 10 ganger hver:

	A	B
1	100,5	90,0
2	95,5	92,0
3	101,5	91,5
4	104,0	90,5
5	100,5	89,5
6	103,0	90,0
7	99,5	91,0
8	101,0	89,5
9	98,5	88,5
10	103,0	92,0
gjennomsnitt	100,7	90,45
standardavvik	2,49	1,17

−0.7

+ 9.55

- Hva vet vi nå om de to voltmetere? Hva ville du gjort hvis du må bruke en av disse to voltmetere?

  → Ville bruke B, fordi det har lavt standardavvik!

- Hva er *usikkerheten* i de to estimatene av systematisk feil?

- Hvilket voltmeter vil du heller bruke, og hvorfor?

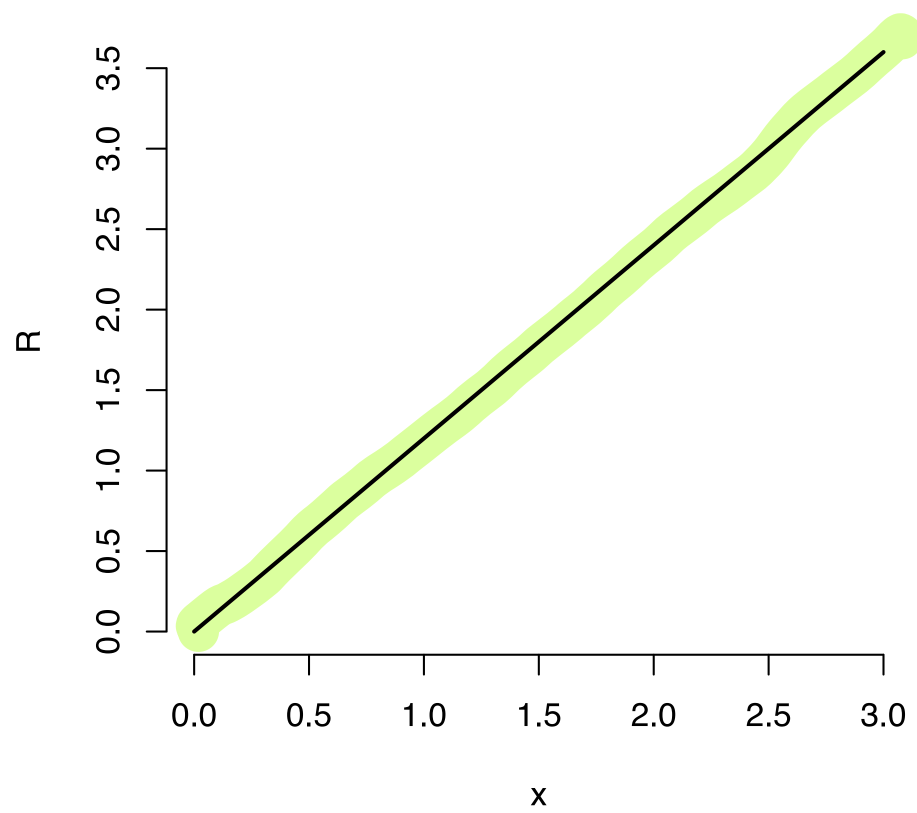
- Hvordan går du frem hvis du egentlig vil måle en spenning på omtrent 200V eller 25V med en av de to voltmetere?

  → Vanskelig å si

  ⇒ Vi trenger en ny kalibrering på 200V
  25V.

Kalibreringskurver

- Det er mulig at samme voltmeter som underestimerer når den skulle måle 100V, er mye nærmere den ekte verdien når den skulle måle 200V.
- Mer generelt måler vi en størrelse R som har en lineær sammenheng med en størrelse x vi er interessert i.



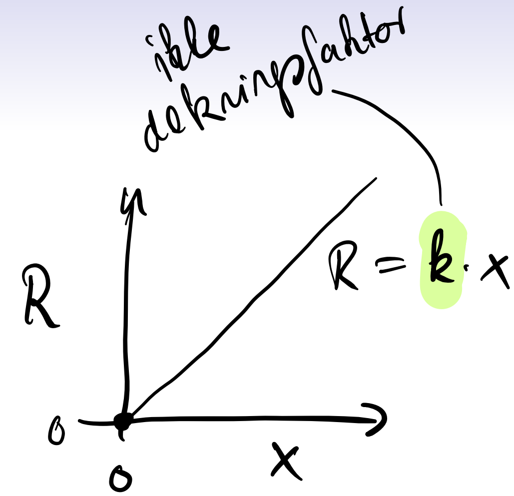
- Kalibreringskurven kan beskrives som

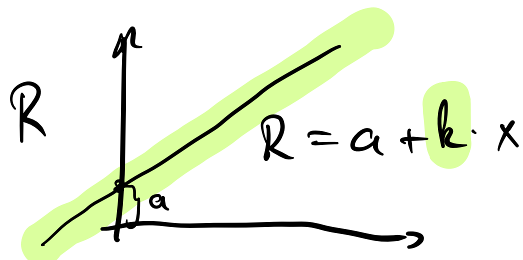
$$R = k \cdot x \ ,$$

hvor k skal estimeres, eller

$$R = a + k \cdot x$$

hvis response ikke kan antas a være lik når $x = 0$.





- Når vi kan anta at R er normalfordelt, kan vi bruke minste kvadratsumme (lineær regresjon) for å finne k.

$R \sim N(\ldots, -)$

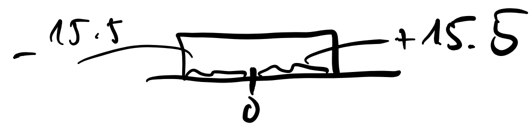
Eksempel 1: Kalibrering av et lodd

Her regner vi eksempelet for kalibrering av et lodd:

Et lodd med ukjent masse M_x på ca 10g skal kalibreres mot et 10g referanselodd.

vi vet masse

1. Referanseloddet hadde ved siste kalibreringen massen $M_r = 10.005g$ og er antatt normalfordelt med standardavvik av $22.5mg$.
2. Mulig drift i referanseloddets masse siden kalibreringen: $\delta M_r = 0$ med maksimal $\pm 15.5mg$.



3. Målt differanse: Ukjent lodd - referanselodd: Ved 20 gjennomtatte målinger er gjennomsnitt $M_{diff} = 20mg$ med standardavvik $S = 64.6mg$, og derfor er standardavviket for middelverdien $S/\sqrt{20} = 14.4mg$.

$u(M_{diff})$

Oppgave: Angi et godt estimat for verdien av det ukjente loddet og standard usikkerheten til den verdien.

Løsning: <u>Målefunksjonen</u> er

$$M_x = M_r + \underbrace{\delta M_r}_{0} + M_{diff} = 10.005\,g + 0 + 0.020\,g$$

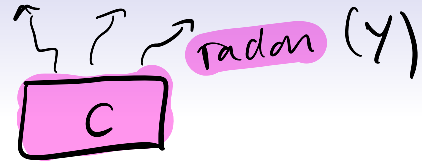
$$= 10.025\,g$$

$u(M_r) = 22.5\,mg$

$u(\delta M_r) = \frac{15.5}{\sqrt{3}} = 8.95\,mg$

$u(M_{diff}) = 14.4\,mg$

$$u(M_x) = \sqrt{u^2(M_r) + u^2(\delta M_r) + u^2(M_{diff})} = \underline{\underline{28.17\,mg}}$$

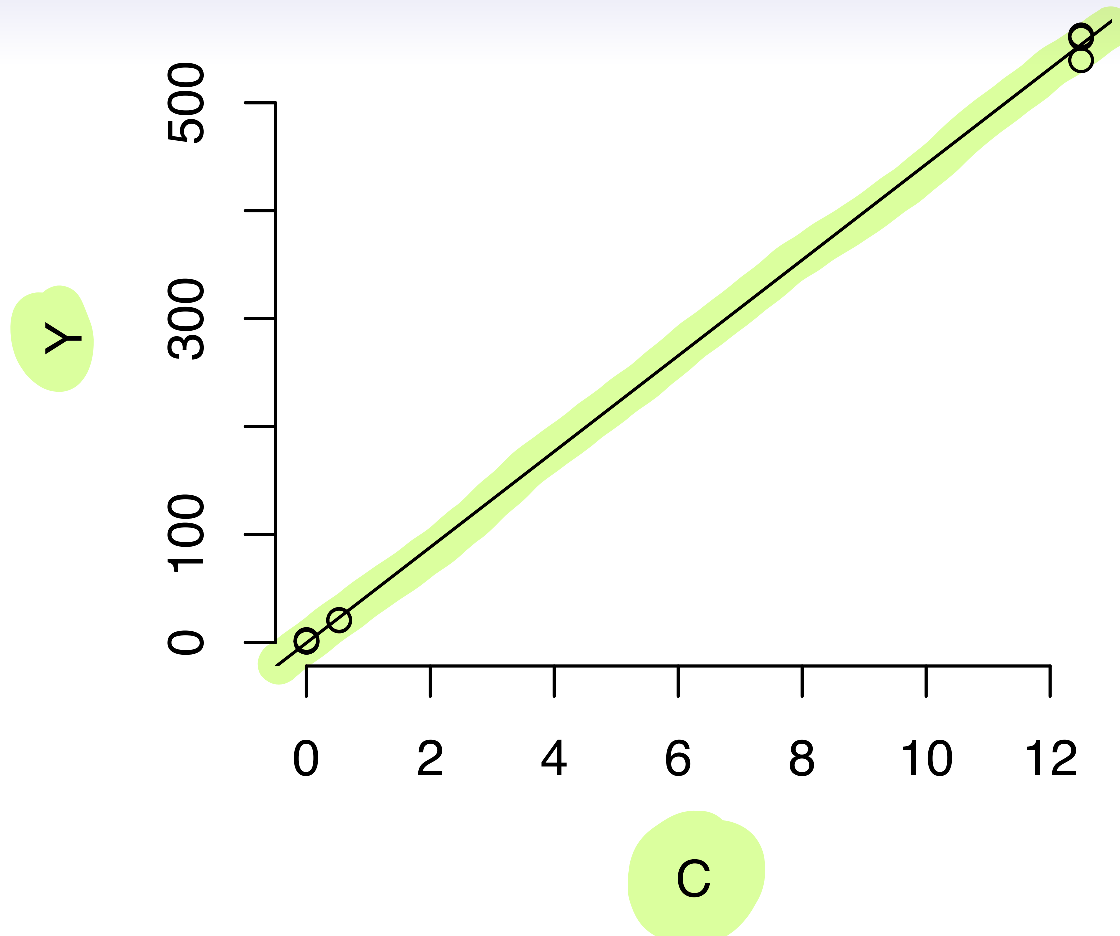


Eksempel 2: Kalibreringskurve

Vi er interessert i konsentrasjonen C av en radioaktiv stoff i en bergart. Siden det er vanskelig å måle den direkte, kan vi måle radon (en radioaktiv gass) som dannes av stoffen i bergarten. For å finne ut på relasjonen mellom konsentrasjon i steinen (C) og radon i luften (Y), måler vi åtte ganger strålingsintensiteten i tre kjente konsentrasjoner av uran:

vanskelig

prøve	Y_i	C_i
1	0.64	0.0036
2	0.67	0.0036
3	2.19	0.0036
4	20.35	0.53
5	20.80	0.53
6	539.4	12.5
7	560.2	12.5
8	562.4	12.5



Vi kan tilpasse en lineær regresjonslinje:

$$\boxed{Y = k \cdot C}$$

med estimert $\hat{k} = 44.36$ og $sd(\hat{k}) = u(\hat{k}) = 0.45$.

Anta vi har målt en spesifisk stråleintensitet $y = 120$ i luften. Så vil vi vite to ting:

1) Hva er den estimerte konsentrasjonen i bergarten?

2) Hva er standard usikkerheten i estimatet?

Løsning:

1)

$$\hat{c} = \frac{y}{\hat{k}} = \frac{120}{44.36} = 2.71$$

2) Usikkerheten er litt mer vanskelig å bestemme, og det finnes mer teoretiske og mer empiriske måter. Her gjør vi det empirisk og beregner forskjellen mellom de estimerte C_i verdiene med de observerte (det er et slags "residual", men for x-variablen istendenfor y)

$$W_i = \frac{Y_i}{\hat{k}} - C_i \ ,$$

og så tar vi standardavviket av alle W_i, $s = \sqrt{\frac{(W_i - \overline{W})^2}{n-1}} = 0.16$