Metodevalidering

Mona Høysæter Fenstad, overlege AIT

ST. OLAVS HOSPITAL
TRONDHEIM UNIVERSITY HOSPITAL

Validering

- Objektive bevis for at spesifiserte krav for tilsiktet bruk er innfridd
- At analysen fungerer til det den er tenkt brukt til

- Innføre nytt måleområde
- Endring av metode i forhold til produsentens anbefalinger

Verifisering

- Dokumentere at metoden fungerer som forventet i eget laboratorium
- Spesifiserte krav (kvalifisering)

- Tilstrekkelig dokumentasjon gjennom vitenskapelige publikasjoner, fra produsent eller referanselaboratorium

Evaluering

- Det settes ikke krav på forhånd
- Laveste nivå av validering

Momenter til vurdering

- Klinisk brukbarhet
- Praktisk egnethet
  - Robusthet
  - IT
- Kostnadsberegning
- Feilkilder
  - Carry-over
  - Interferens
- Analytisk kvalitet
  - Deskriptiv statistikk
  - Slutningsstatistikk

Klinisk brukbarhet

- Kan analysen gi legene hjelp til riktig behandling av pasienten?
- Er den kliniske populasjonen testen er prøvd ut i godt beskrevet?
- Er populasjonen relevant for vår bruk?
- Klarer testen å skille mellom ulike sykdomstilstander?
- Medfører testen risiko for pasienten?

Praktisk egnethet

- Vil analysen/metoden fungere teknisk i laboratoriet?
  - arbeidsflyt, plassbehov
- Brukervennlighet
  - Ergonomi, Støy, HMS
- Grad av automasjon
  - IT-oppkobling
- Kapasitet, analysetid
- Prøvemateriale
  - Holdbarhet
  - Lot-til-lot variasjon
- Kontroll, kalibrering
- Service, vedlikehold
- Lagerplass til reagenser og forbruksvarer
- Personell og opplæringsbehov

Bobjgalindo https://commons.wikimedia.org/

Kostnadsberegning

- **Pris per test/Pris per pasient**
- **Takstkode (inntekt)**

- Sammenligning med alternativ metode
  - Erstatter etablert metode?
  - Metode tilgjengelig på eksisterende teknologisk platform?

- Samfunnsøkonomisk nytteverdi?
  - 40/60

Valideringsplan

- Innledning; vurdering av klinisk brukbarhet, praktisk egnethet og kostnadsberegning
- Omfang av validering
- Praktisk gjennomføring
  - Analytisk kvalitet
  - Referansegrenser
- Kriterier

Valideringsrapport

- Rådata av analyseresultat
- Statistiske beregninger
- Diskusjon, vurderinger og konklusjoner som er gjort med bakgrunn i funn og litteratur
- Referanser, pakningsvedlegg

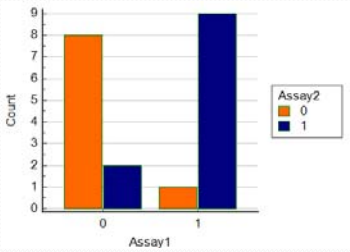
Analytisk kvalitet

Type test	Nøyaktighet/ Riktighet[a]	Måleområde	Presisjon[b]	Sannsynlighet[c]	Statistikk
Kvantitativ Resultat på en skala inkludert desimaler mellom to gitte målverdier	X	X	X	X	Korrelasjon/ Regresjon Linearitet Standardavvik (SD) Variasjonskoeffisient (CV)
Semi-kvantitativ Kvantitativt resultat plasseres i ulike kategorier	X	X			Venn-diagram/Bayesiansk statistikk Sensitivitet/Spesifisitet/Nøyaktighet(FP/FN) Kappa – statistikk
Kvalitativ Testen gir to mulige utfall, pos/neg	X				

a : Sammenligning med referansemetode
b : Repeterbarhet og reproduserbarhet
c : Mulig å beregne en statistisk sannsynlighet for verdien?

Deskriptive variabler (kategoriske data)

- Bar-chart



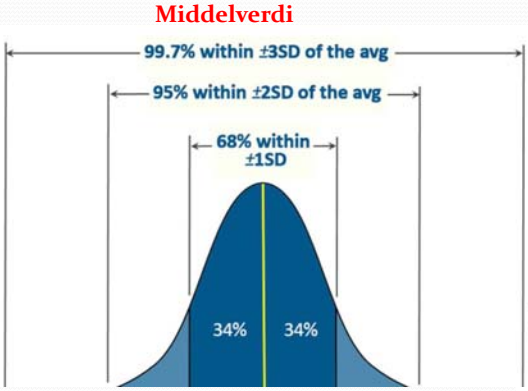
- Krysstabell/Venn-diagram
  - Tall
  - Prosent
  - Odds eller Risk
  - Odds ratio eller Risk ratio
  - Sensitivitet / Spesifisitet
  - Prediktiv verdi

Outcome	Group 0	Group 1
0	8	2
1	1	9
	9	11

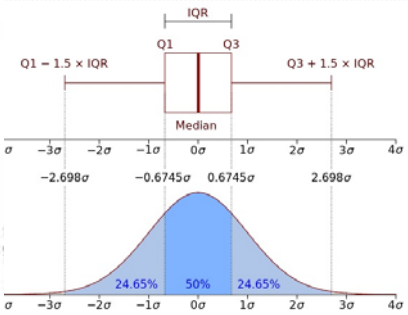
Deskriptive variabler (kontinuerlige data)

- Utfallsrom, maksimum, minimum, frekvens
- Sentral tendens

Middelverdi



Median



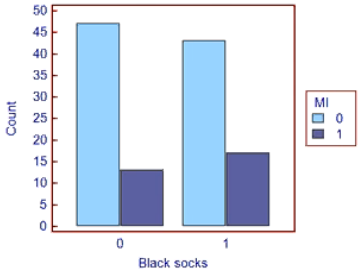
Hjerteinfarkt og svarte sokker

	MI = 1 (Disease present)	MI = 0 (Healthy)	
Black socks = 1	17 (28%)	43 (72%)	60
Black socks = 0 (coloured socks)	13 (22%)	47 (78%)	60
	30	90	120

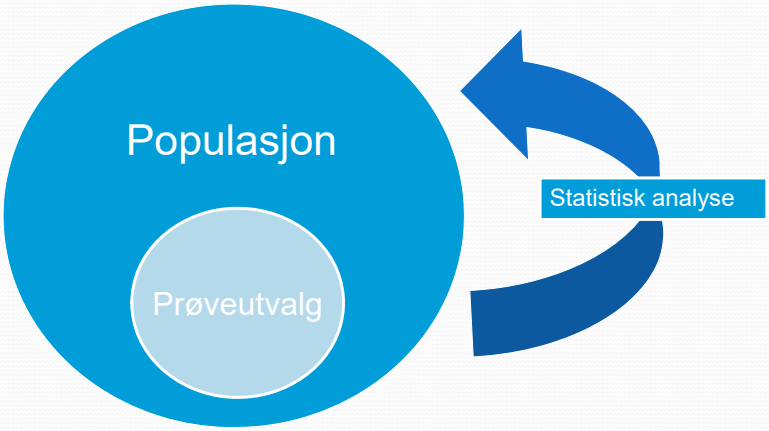
$$Odds\ MI\ in\ black\ socks = \frac{17}{43} = 0.395$$

$$Odds\ MIcoloured\ socks = \frac{13}{47} = 0.276$$

$$Odds\ ratio = \frac{0.395}{0.276} = 1.43$$

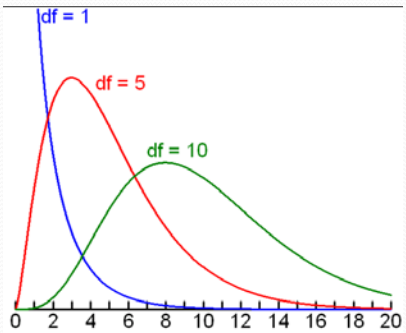


Slutningsstatistikk



Chi Sqare statistikk

- Samplingvariasjon gir sokke-resultatet. Finnes det slik variasjon, da fins det også en samplingdistribusjon!



$$d.f. = (number\ rows - 1) \times (number\ columns - 1)$$

For a 2x2 table d.f. = (2-1)x(2-1) = 1

For a 3x3 table d.f. = (3-1)x(3-1) = 4

For a 6x10 table d.f. = (6-1)x(10-1) = 45

Chi Sqare-statistikk

- Lag en krysstabell for null-hypotesen med forventede verdier ut fra insidens av hjerteinfarkt i gruppen totalt
- Lag krysstabell med differansen mellom forventet og observert
- Kvadratet av forskjellen / forventet verdi

Expected:

	MI = 1	MI = 0	
Black socks = 1	15 (25%)	45 (75%)	60
Black socks = 0	15 (25%)	45 (75%)	60
	30 (25%)	90 (75%)	120

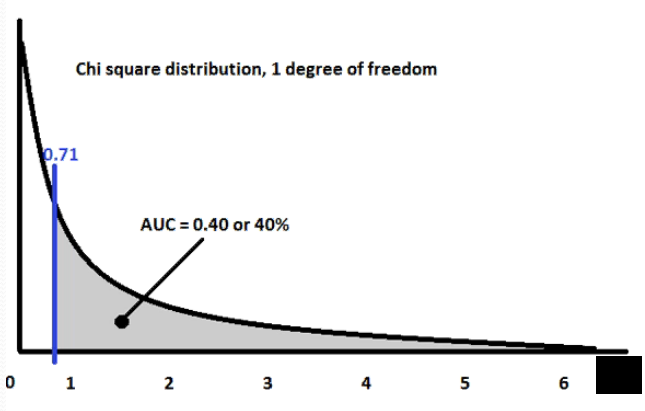
Observed:

	MI = 1	MI = 0	
Black socks = 1	17 (28%)	43 (72%)	60
Black socks = 0	13 (22%)	47 (78%)	60
	30	90	120

The size of difference between observed & expected, squared, and divided by expected

	MI = 1	MI = 0
Black socks = 1	4 / 15	4 / 45
Black socks = 0	4 / 15	4 /45

Total = (4/15)+(4/15)+(4/45)+(4/45)
= 0.71



Hjerteinfarkt og Diabetes

Expected (if null hypothesis is true):

	MI = 1	MI = 0	
Diabetes = 1	15 (25%)	45 (75%)	60
Diabetes = 0	15 (25%)	45 (75%)	60
	30 (25%)	90 (75%)	120

Observed:

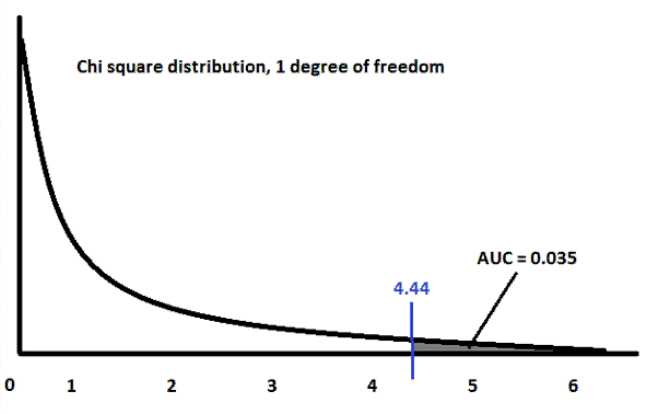
	MI = 1	MI = 0	
Diabetes = 1	20 (33%)	40 (66%)	60
Diabetes = 0	10 (17%)	50 (83%)	60
	30	90	120

The size of difference between observed & expected, squared, and divided by expected

	MI = 1	MI = 0
Black socks = 1	25 / 15	25 / 45
Black socks = 0	25 / 15	25 /45

$$X^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$

= (25/15)+(25/15)+(25/45)+(25/45)
= 4.44



www.medicalstats.org

Av og til snubler menneskene over en sannhet. De fleste skynder seg videre som om ingenting hadde hendt.
Winston Churchill

Det gjelder å finne en sannhet som er sannhet for meg.
Søren Kierkegaard

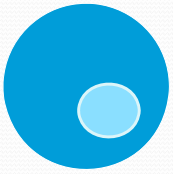
Selv om den ikke er kjent, slutter ikke sannheten å være sann.
Richard Bach

Visdom finnes kun i sannheten.
Johann Wolfgang von Goethe

«Sann verdi»

Klinisk brukbarhet

- Er den kliniske populasjonen testen er prøvd ut i godt beskrevet?
  - Populasjonen relevant?
  - Ulike sykdomstilstander?



Metodesammenligning

- Definitiv metode: gir helt riktig nivå
- Referansemetode: ubetydelig avvik fra riktig nivå
- Sertifiserte referansematerialer
- Konsensusverdi (gjennomsnitt av måling fra mange ulike laboratorier)

Metodesammenligning (kvalitative)

	Positiv	Negativ	
Positiv	18	2	20
Negativ	6	74	80
			100

Sensitivitet: positive av sanne positive 18/20
Spesifisitet: negative av sanne negative 74/80
Falske negative : 2/100 (2%)
Falske positive: 6/100 (6%)

Cohen`s Kappa (κ)

	Positiv	Negativ	
Positiv	25	25	50
Negativ	25	25	50
			100

- Samsvar: 92/100 (0,92)
- Tilfeldighet: 50/100 (0,50)
- Kappa: 0,92-0,5 / 1-0,5 = 0,84

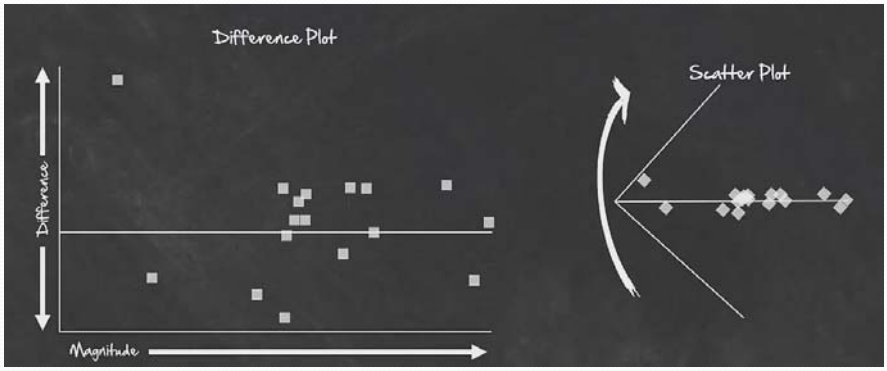
- Vektet kappa; flere kategorier, størrelsen på uenigheten

Metodesammenligning (kvantitative)

- Bland Altmann
- Mountain Plot
- Deming regresjon
- Passing Bablok regresjon

Bland Altmann

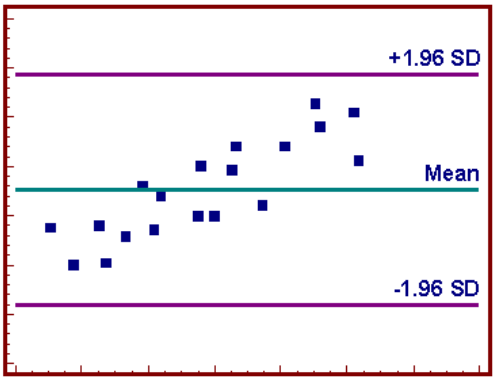
- Difference plot
  - Når skalaen på forskjell mot mengde/størrelse er veldig forskjellig
  - Mindre tettpakket enn rundt diagonalen, mer visuell informasjon
  - Informasjon om endring med størrelse på målingene



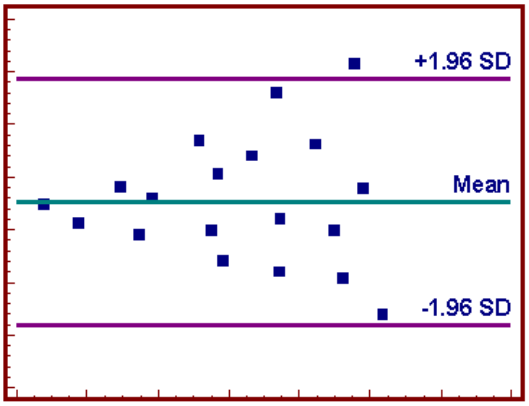
Bland Altmann

- Plot against (x-axis)
  - Vanlig: forskjell mot gjennomsnittet av de to metodene (anbefalt av forfatterne)
  - Evt. forskjell mot gullstandardmetode
- Plot differences
  - Som % eller ratios: nyttig når variabiliteten øker med størrelsen på målingene

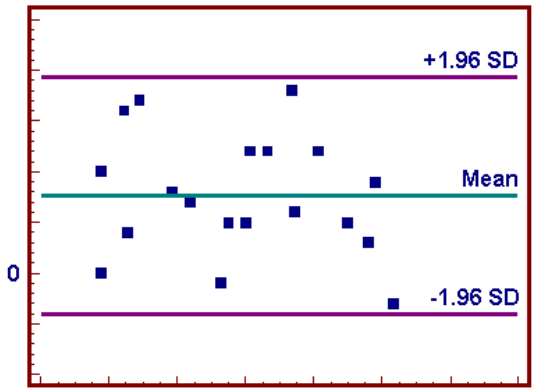
Proposjonal forskjell



Variabilitet øker med størrelse på målingene

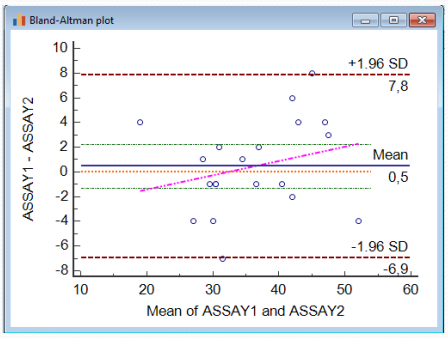


Systematisk forskjell



Bland Altmann

- Tilvalg
  - Line of equality : systematiske forskjeller
  - 95% CI for middelverdi av forskjellene: hvis «line of equality» ikke er innenfor dette intervallet er det signifikante systematiske forskjeller
  - Regression line of differences: proporsjonale forskjeller



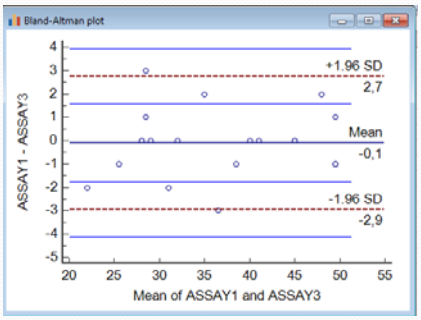
Limits of agreement

- If the differences are normally distributed, we would expect 95% of the differences to lie between mean difference (d) -1,96SD and mean d+1,96SD. We can then say that nearly all pairs of measurements by the two methods will be closer together than these extreme values, which we call 95% limits of agreement
- These values define the range within which most differences between measurements by the two methods will lie
- **Provided differences within the observed limits of agreement would not be clinically important** we could use the two measurement methods interchangeably

Bland, Altman 1999

Limits of agreement

95% CI for limits of agreement : hvis disse er for vide kan metoden være uakseptabel, selv om beregnede limits vurderes som tilfredsstillende



Samsvar mellom metoder

- ”Agreement” eller samsvar : hva er godt nok?
  - Dette kan ikke avgjøres med statistikk
  - Klinisk skjønn
  - Situasjonsbestemt
  - Fortrinnsvis forhåndsdefinert

  - Ex. Blodtrykksmåling: 10mmHg ok? 30mmHg ikke?
  - Hb? Klinisk liten forskjell på 12.2 og 12.0
    - Sporbarhet
    - ”Følgefeil”

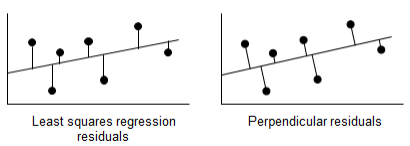
Repeterbarhet

- Hvis en av metodene har dårlig repeterbarhet, vil nødvendigvis samsvar mellom metodene bli dårlig

(rart hvis den har bedre samsvar med en annen metode enn med seg selv ☺)

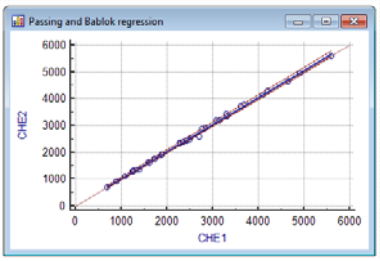
Passing-Bablok regresjon

- Lineær regresjonsmodell som er uavhengig av prøvedistribusjon og tilfeldige målefeil for begge metoder
  - Krever ikke input med gjentatte målinger eller CV for metodene
- Bruk «calculate perpendicular residuals» (vs least sqare method) som anbefalt i artikkelen til Passing og Bablok



Passing og Bablok regresjon

- Intercept (A) ≠ 0 : systematisk forskjell (testes ved 95% konfidensintervallet)
- Slope (B) ≠ 1 : Proposjonal forskjell
- Tilfeldige forskjeller (SD av residualene)



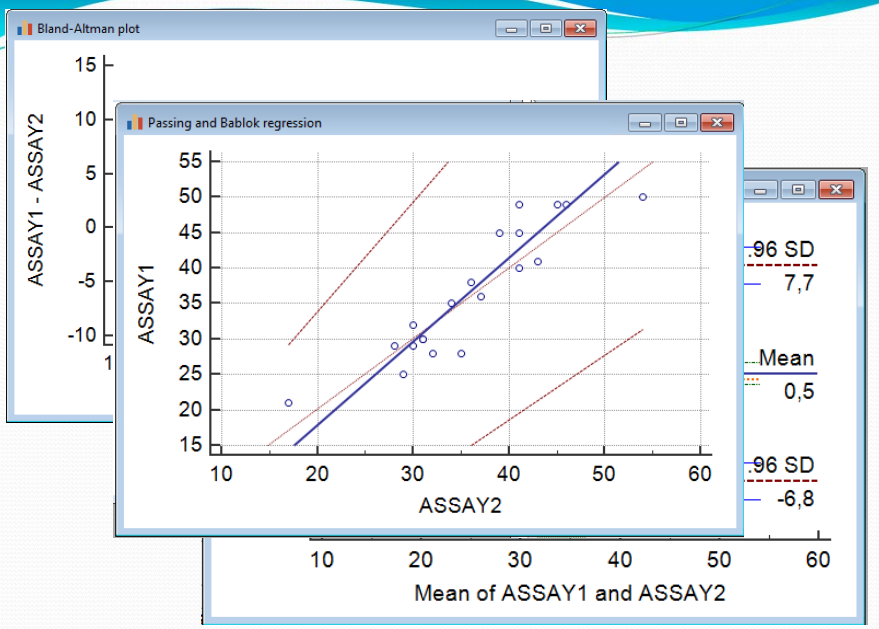
- Linear model validitiy
  - tester bare om metoden er brukbar
  - sier ikke om metodene er sammenlignbare/like gode!
  - P lavere enn 0.05 betyr at P&B metoden ikke kan brukes

Regression Equation

y = 22.904132 + 1.008326 x

Systematic differences	
Intercept A	22.9041
95% CI	-3.9496 to 54.0386
Proportional differences	
Slope B	1.0083
95% CI	0.9943 to 1.0234
Random differences	
Residual Standard Deviation (RSD)	35.0689
± 1.96 RSD Interval	-68.7351 to 68.7351
Linear model validity	
Cusum test for linearity	No significant deviation from linearity (P=0.74)

Save predicted values - Save residuals



Passing and Bablok regression

Variable X	ASSAY2
Variable Y	ASSAY1
Sample size	20

	Variable X	Variable Y
Lowest value	17,0000	21,0000
Highest value	54,0000	50,0000
Arithmetic mean	36,0000	36,4500
Median	35,5000	35,5000
Standard deviation	8,1305	9,0814
Standard error of the mean	1,8180	2,0307

Regression Equation

y = -5,807487 + 1,179144 x

Systematic differences	
Intercept A	-5,8075
95% CI	-18,1364 to 2,8750
Proportional differences	
Slope B	1,1791
95% CI	0,9167 to 1,5455
Random differences	
Residual Standard Deviation (RSD)	2,6514
± 1.96 RSD Interval	-5,1967 to 5,1967
Linear model validity	
Cusum test for linearity	No significant deviation from linearity (P=0,36)

Help (F1) Print Save predicted values - Save residuals

Passing and Bablok regression

Variable X	ASSAY2
Variable Y	ASSAY1
Sample size	100

	Variable X	Variable Y
Lowest value	17,0000	21,0000
Highest value	54,0000	50,0000
Arithmetic mean	36,0000	36,4500
Median	35,5000	35,5000
Standard deviation	7,9646	8,8960
Standard error of the mean	0,7965	0,8896

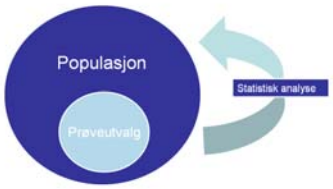
Regression Equation

y = -5,807487 + 1,179144 x

Systematic differences	
Intercept A	-5,8075
95% CI	-12,6250 to 0,0000
Proportional differences	
Slope B	1,1791
95% CI	1,0000 to 1,3750
Random differences	
Residual Standard Deviation (RSD)	2,5409
± 1.96 RSD Interval	-4,9801 to 4,9801
Linear model validity	
Cusum test for linearity	Significant deviation from linearity (P<0,01)

Sammendrag

- Validering – mer enn analytisk kvalitet
- Deskriptiv statistikk vs slutningsstatistikk
  - Godt nok samsvar?
  - Forkaste eller beholde null-hypotesen (metodene er like)?



Visdom finnes kun i sannheten. *Johann Wolfgang von Goethe*

Det gjelder å finne en sannhet som er sannhet for meg. *Søren Kierkegaard*