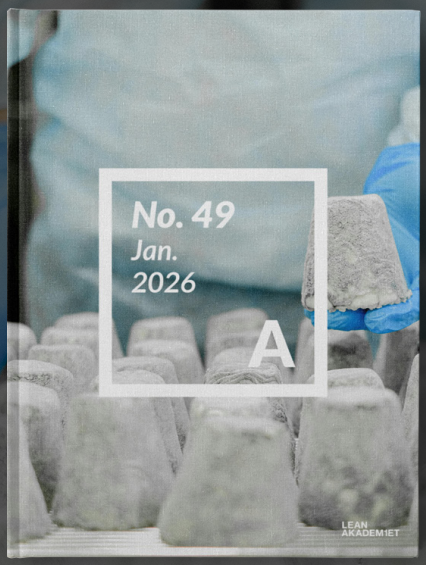
© 2026
Lean Akademiet



#KVALITETSSIKRING OG DATA-ANALYSE
#DATAANALYSE OG DASHBOARDING #SIX SIGMA



23 JANUAR 2026

MÅLESYSTEMANALYSE MED KATEGORISKE DATA - NÅR VI IKKE MÅLER, MEN VURDERER

I mange organisationer opstår usikkerheden ikke i analysen, men langt tidligere i processen, nemlig i selve vurderingen. Når udfaldet bygger på menneskelige fortolkninger frem for tal, stilles der særlige krav til målesystemet.



KION SCHMELTZER

Kion udvikler uddannelser med fokus på synergien mellem Lean og Six Sigma samt varig forbedring.

23 JANUAR 2026

I mange organisationer opstår usikkerheden ikke i analysen, men langt tidligere i processen, nemlig i selve vurderingen. Når udfaldet bygger på menneskelige fortolkninger frem for tal, stilles der særlige krav til målesystemet.

Denne artikel sætter fokus på netop den type målinger, hvor kvalitet og stabilitet ofte tages for givet, men sjældent dokumenteres systematisk.

Når data bygger på vurderinger frem for målinger

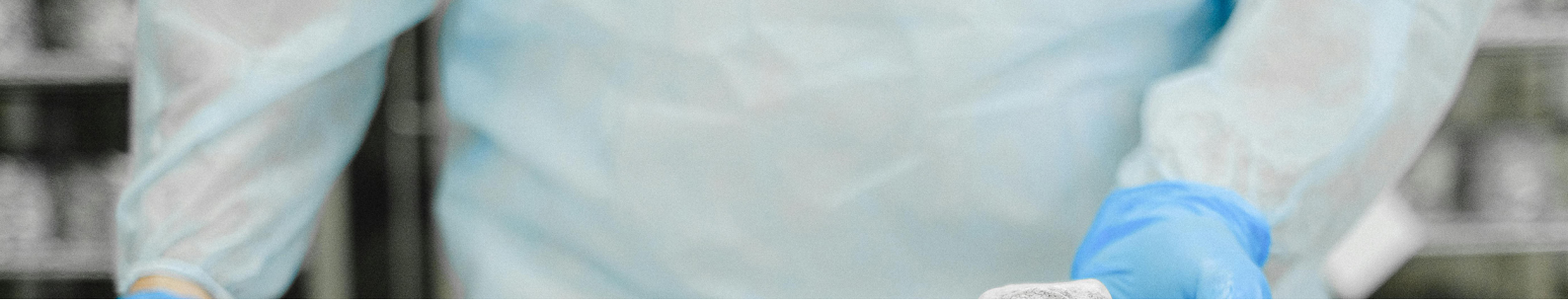
Ikke alle målinger foretages med skydelære, termometer eller mikrometer. I mange processer bygger vores data på menneskelige vurderinger, hvor udfaldet ikke er et tal, men en kategori. Det kan være OK eller defekt, godkendt eller afvist, lav, middel eller høj.

Denne type vurderinger er udbredt i både produktion og service. Visuel kontrol ved indgangs- og udgangsinspektion, vurdering af overflader, fejlklassificeringer og kundeklager er blot nogle af de områder, hvor kategoriske data spiller en central rolle.

Her opstår ofte en velkendt udfordring: To personer kan se på det samme emne og nå frem til forskellige konklusioner. Den ene vurderer emnet som acceptabelt, den anden som defekt. Intuitivt kan løsningen synes enkel. Vi skal bare træne medarbejderne bedre.

Problemet er, at træning alene sjældent er nok. Uden et stabilt og reproducerbart målesystem risikerer organisationen, at variationen i data primært afspejler forskelle i vurderinger og ikke reelle forskelle i kvalitet eller procesperformance.

Præcis som ved numeriske målinger er det derfor nødvendigt at analysere selve målesystemet. Også når målesystemet består af mennesker, instruktioner og vurderinger frem for måleinstrumenter.



I denne artikel ser vi nærmere på målesystemanalyse for kategoriske data. Fokus er på Kappa-analyse, som gør det muligt objektivt at vurdere, om vores vurderinger er stabile, ensartede og bedre end tilfældige gæt. Artiklen bygger videre på den generelle introduktion til MSA og udgør anden del i serien om målesystemanalyse.

Kappa analyse

Der findes forskellige typer af kappa-analyser, der er designet til at vurdere graden af enighed mellem bedømmere eller klassifikatorer i forskellige scenarier. Valget af kappa-type afhænger af datatyperne og vurderingssituationen. Her er de vigtigste kappa-typer:

Kappa-type	Bedømmere	Datatype	Typisk anvendelse
Cohen's Kappa	2	Nominale/kategoriske	Simpel enighed mellem to bedømmere
Weighted Kappa	2	Ordinale	Når fejl tæt på hinanden vægtes mindre
Fleiss' Kappa	3 eller flere	Nominale/kategoriske	Enighed blandt flere bedømmere
Scott's Pi	2	Nominale/kategoriske	Simpel enighed, alternativ til Cohens's Kappa
Krippendorff's Alpha	2 eller flere	Alle typer	Fleksibelt mål til flere bedømmere og datatyper
Quadratic Weighted Kappa	2	Ordinale	Når fejl langt fra hinanden vægtes højere
Conger's Kappa	Flere	Nominale/kategoriske	Fleksibel generalisering af Cohen's Kappa
Interobserver Agreement	Flere	Nominale/kategoriske	Til visuelle fund eller medicinske vurderinger
Time-Weighted Kappa	Flere over tid	Alle typer	Måling af enighed over tid med vægtning

Tabel 1 - Kappa typer
Oversigt over Kappa-typer og deres anvendelse.

Opsætning af MSA med kategoriske data

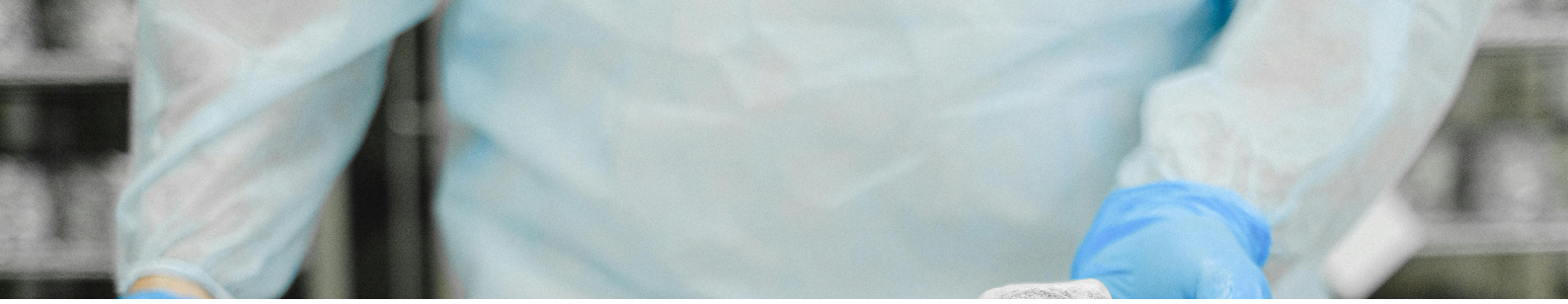
Før du kan udføre analysen, skal følgende være på plads:

Måleobjekter

Et repræsentativt sæt prøver eller emner – typisk 20-30 stk.
Det er vigtigt, at datasættet afspejler virkeligheden, og at der er ca. 50 % gode og 50 % dårlige prøver, så beregningerne bliver statistisk meningsfulde.

Operatører

Et antal personer (typisk 3) som udfører vurderingerne.



Vælg ikke specialisterne, men de faktiske brugere af målesystemet. De skal have adgang til normale instruktioner, men ellers arbejde præcis som i hverdagen – samme lys, hjælpemidler osv.

Måleplan

En struktureret plan, fx:

- 30 emner × 3 operatører × 3 gentagelser = 270 registreringer
- Rækkefølgen skal være tilfældig, så der ikke opstår systematik i vurderingerne

Analyse: Kappa-værdien

Når dataene er indsamlet, analyseres graden af enighed mellem operatørerne – både individuelt og på tværs.
Her bruger vi Kappa-analyse, som korrigerer for tilfældig enighed.

Cohen's Kappa (eksempel)

Kappa-værdien (κ) måler, hvor meget bedre enigheden er, end hvad man kunne forvente ved rent gætteri:

Kappa-værdien (κ) angiver, i hvilken grad enigheden overstiger det niveau, man ville forvente ved tilfældigt gæt.
P0 = observeret enighed (fx 85 %)
Pe = forventet enighed ved tilfældige vurderinger.

En Kappa på 1,0 betyder perfekt enighed. 0,0 betyder ingen bedre enighed end tilfældigt, og negative værdier betyder faktisk værre end tilfældigt – du ville måle bedre med bind for øjnene.

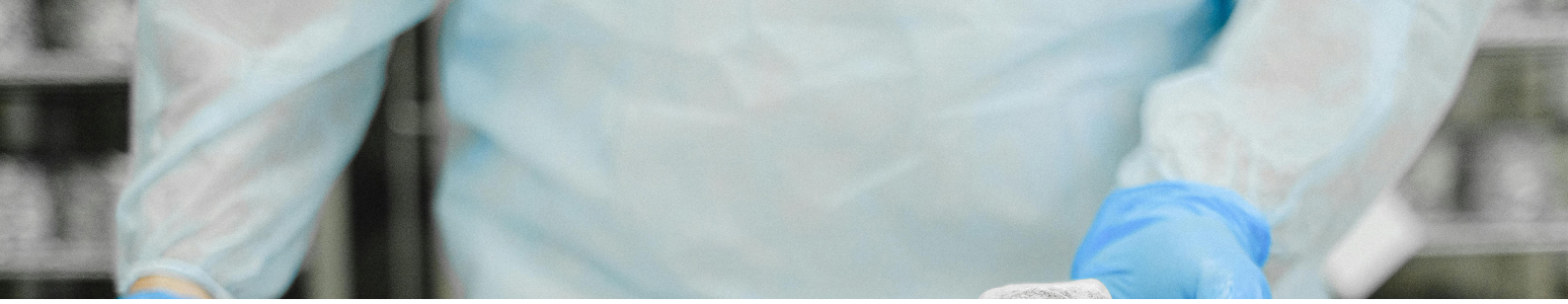
$$k = \frac{P_0 - P_e}{1 - P_e}$$

Figur 1 - Eksempel på Cohen's Kappa

Når vi har flere kategorier

Hvis vores data ikke bare er 'OK/Defekt', men fx 'Ingen fejl', 'Mindre fejl', 'Alvorlig fejl', bør vi bruge Weighted Kappa.

Her vægtes uenigheder forskelligt – fx er det værre at forveksle 'OK' med 'Alvorlig fejl' end



med 'Mindre fejl'.

Kappa bruges til at vurdere reproducérbarhed, altså om bedømmere konsekvent kan nå til samme vurdering, både individuelt (over tid) og sammenlignet med hinanden. Den kan dog også bruges til at se om målerne kan 'ramme' den sande værdi (altså nøjagtighed).

Kappa-værdien ligger mellem -1 og 1:

- 1.0: Perfekt enighed mellem bedømmere
- 0.0: Ingen bedre enighed end man ville forvente tilfældigt
- < 0: Enighed er dårligere end tilfældig (meget usædvanligt)

Kappa-værdi (*k*)	Fortolkning
< 0.00	Ingen enighed (værre end tilfældigt)
0.00 - 0.20	Svag enighed
0.21 - 0.40	Moderat enighed
0.41 - 0.60	Middel enighed
0.61 - 0.80	God enighed
0.81 - 1.00	Meget god/fremragende enighed

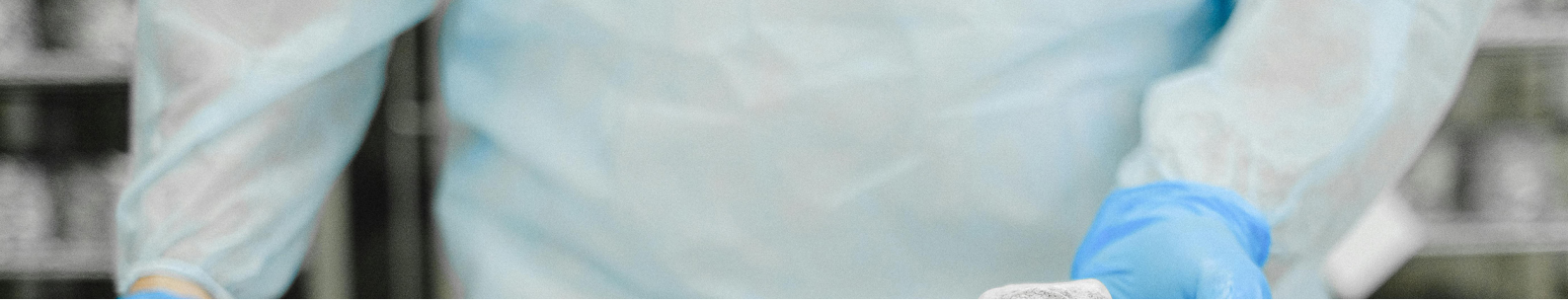
Tabel 2 - Klassiske regler for Kappa
Udvidet fortolking: Landis og Koch, ofte brugt i forskningsverden.

Kappa-værdi (*k*)	Fortolkning
> 0.90	Fremragende enighed
0.70 - 0.90	Acceptabel enighed
< 0.70	Uacceptabel enighed

Tabel 3 - Klassiske regler for Kappa
Tommelfingerregel for fortolkning af Kappa.

Hvornår bruges hvad?

Tommelfingerreglen (>0.90,>0.70) bruges ofte i industriel praksis eller kvalitetsstyring, hvor høje krav stilles til nøjagtighed og konsistens (fx produktion, medicinsk udstyr og lignende).



Den udvidede Landis & Koch-skala bruges mere i forskning og sociale studier, hvor en lidt lavere enighed ofte accepteres.

Hvorfor Kappa er bedre end procentvis enighed

Procentvis enighed (P0) fortæller, hvor ofte to vurderinger er ens, men siger intet om, hvor meget af det der skyldes tilfældigheder.

Hvis fx 95 % af alle dele normalt er 'OK', kan to personer opnå 95 % enighed bare ved altid at sige OK – uden egentlig at være gode vurderere.

Kappa korrigerer for denne chance-enighed - og giver dermed et mere retvisende billede af målesystemets reelle kvalitet.

Hvad kan vi bruge resultaterne til?

Resultaterne fra en målesystemanalyse med kategoriske data giver vigtig indsigt i, hvordan vores vurderinger faktisk fungerer i praksis. Først og fremmest kan analysen vise, om operatørerne vurderer ens fra gang til gang. Hvis den samme person træffer forskellige beslutninger ved gentagne vurderinger af det samme emne, er målesystemet ikke reproducérbart, og data kan ikke anvendes som et stabilt beslutningsgrundlag.

Analysen giver samtidig et billede af, om de prøver, der indgår i undersøgelsen, reelt afspejler den variation, som forekommer i den daglige drift. Hvis datasættet ikke er repræsentativt for virkeligheden, risikerer man at drage konklusioner, som ikke holder, når målesystemet anvendes i praksis.

Derudover kan resultaterne afdække forskelle mellem operatørerne. Systematiske afvigelser peger ofte på et behov for målrettet træning eller fælles kalibrering af vurderingerne, så alle arbejder ud fra samme forståelse af, hvad der er acceptabelt, og hvad der ikke er.

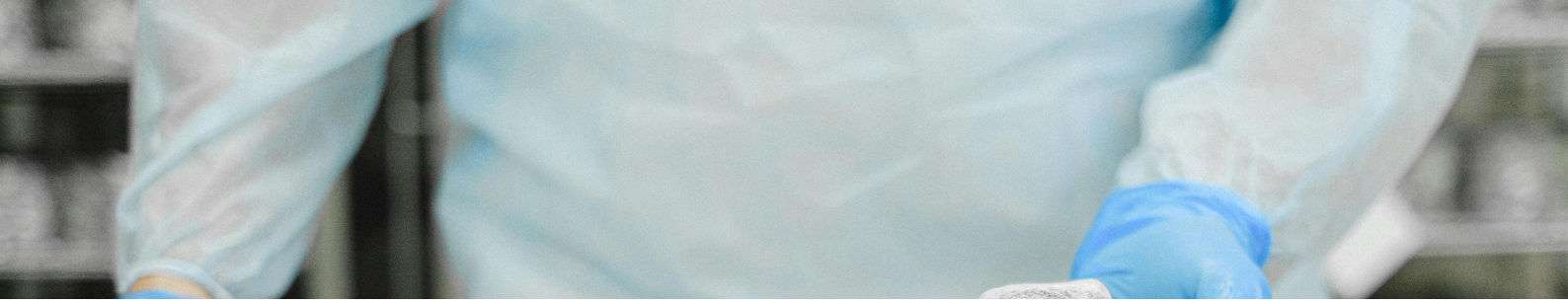
Endelig peger en Kappa-analyse ofte direkte på konkrete forbedringstiltag. Det kan være behov for tydeligere defektgrænser, bedre visuelle referenceeksempler eller mere præcise instruktioner. På den måde bliver målesystemanalysen ikke blot en vurdering af datakvaliteten, men et aktivt værktøj til at styrke både kvalitetssikring og procesforbedring.

Konklusion

Målesystemanalyse er en forudsætning for valide data og dermed for datadrevne beslutninger. Uanset om der arbejdes med numeriske målinger eller menneskelige vurderinger, er det afgørende at forstå, hvor stor en del af variationen der stammer fra selve målesystemet.

Når målesystemet ikke er stabilt og reproducérbart, risikerer organisationer at optimere på et forkert grundlag. Ved at arbejde systematisk med MSA skabes der transparens i data, fælles forståelse på tværs af organisationen og et langt stærkere fundament for både kvalitetsstyring og procesforbedring.

Et godt målesystem reducerer usikkerhed, minimerer interne diskussioner og gør det muligt at fokusere forbedringsindsatsen dér, hvor den reelt skaber værdi.

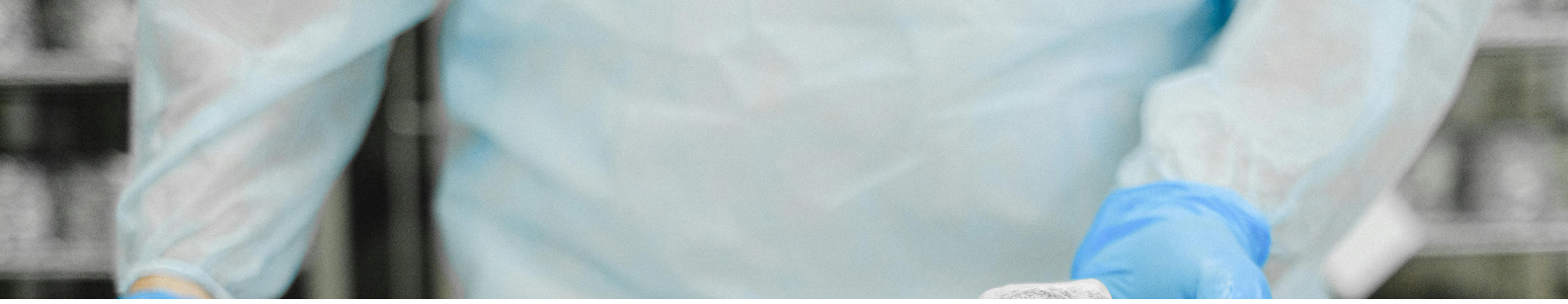


Fra vurdering til måling

I denne artikel har vi haft fokus på målesystemanalyse for kategoriske data, hvor udfaldet bygger på vurderinger frem for tal. I næste artikel i serien flytter vi fokus til kontinuerte data og ser nærmere på klassiske MSA-metoder som Gage R&R, bias, linearitet og stabilitet.

Her går vi i dybden med, hvordan numeriske målesystemer analyseres, og hvordan du sikrer, at variationen i dine data afspejler processen og ikke selve målingen.

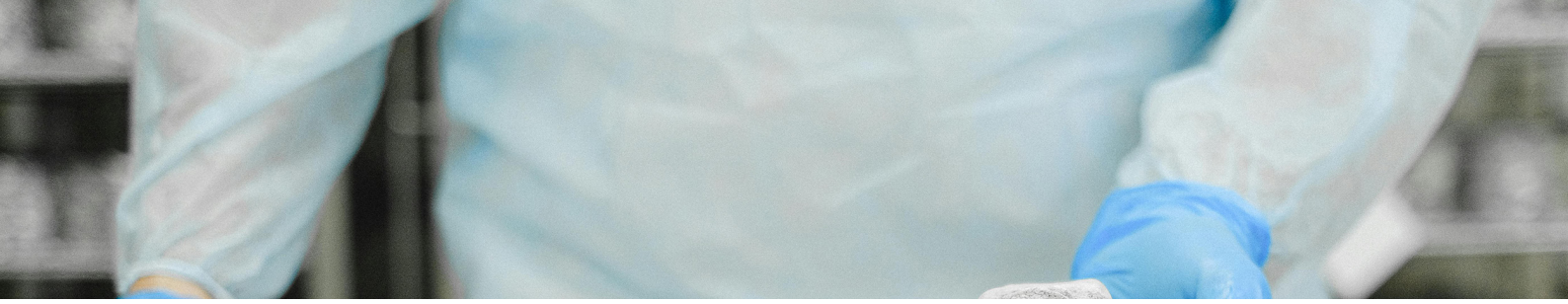
Få faglig viden og lad dig inspirere i vores **LA Library**



Tabel 1 - Oversigtstabel over MSA-typer

Oversigt over MSA-typer opdelt på datatyper og formål.

Kappa-type	Bedømmere	Datatype	Typisk anvendelse
Cohen's Kappa	2	Nominale/kategoriske	Simpel enighed mellem to bedømmere
Weighted Kappa	2	Ordinale	Når fejl tæt på hinanden vægtes mindre
Fleiss' Kappa	3 eller flere	Nominale/kategoriske	Enighed blandt flere bedømmere
Scott's Pi	2	Nominale/kategoriske	Simpel enighed, alternativ til Cohens's Kappa
Krippendorff's Alpha	2 eller flere	Alle typer	Fleksibelt mål til flere bedømmere og datatyper
Quadratic Weighted Kappa	2	Ordinale	Når fejl langt fra hinanden vægtes højere
Conger's Kappa	Flere	Nominale/kategoriske	Fleksibel generalisering af Cohen's Kappa
Interobserver Agreement	Flere	Nominale/kategoriske	Til visuelle fund eller medicinske vurderinger
Time-Weighted Kappa	Flere over tid	Alle typer	Måling af enighed over tid med vægtning



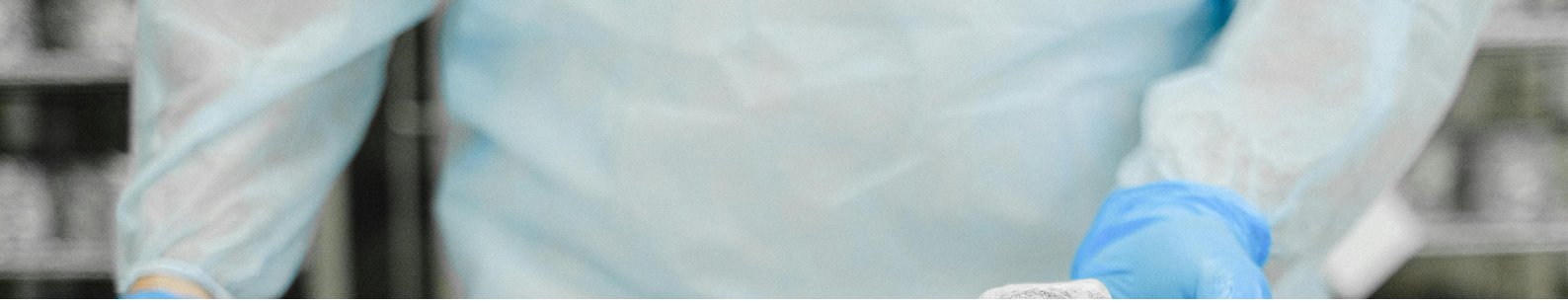
Figur 1 - Eksempel på Cohen's Kappa

Kappa-værdien (κ) angiver, i hvilken grad enigheden overstiger det niveau, man ville forvente ved tilfældigt gæt.

P0 = observeret enighed (fx 85 %)
Pe = forventet enighed ved tilfældige vurderinger.

En Kappa på 1,0 betyder perfekt enighed. 0,0 betyder ingen bedre enighed end tilfældigt, og negative værdier betyder faktisk værre end tilfældigt – du ville måle bedre med bind for øjnene.

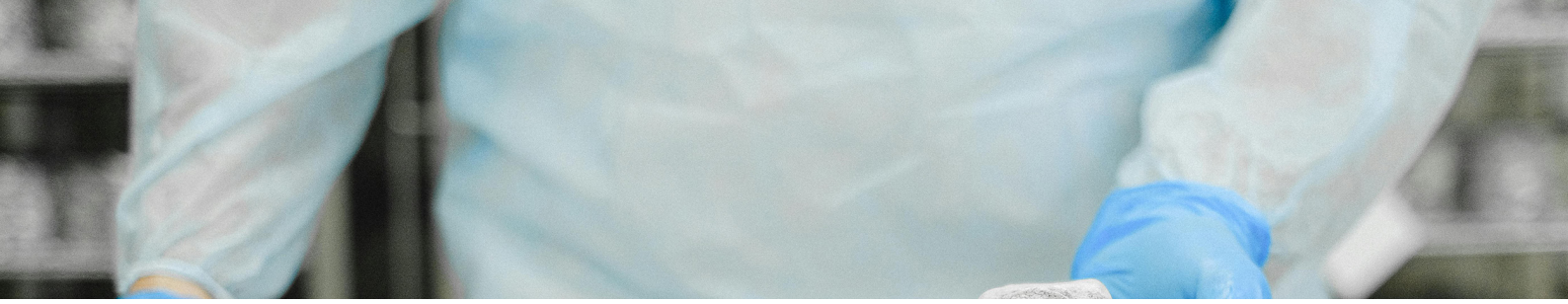
$$k = \frac{P_0 - P_e}{1 - P_e}$$



Tabel 2 - Klassiske regler for Kappa

Udvidet fortolking: Landis og Koch, ofte brugt i forskningsverden.

Kappa-værdi (*k*)	Fortolkning
< 0.00	Ingen enighed (værre end tilfældigt)
0.00 - 0.20	Svag enighed
0.21 - 0.40	Moderat enighed
0.41 - 0.60	Middel enighed
0.61 - 0.80	God enighed
0.81 - 1.00	Meget god/fremragende enighed



Tabel 3 - Klassiske regler for Kappa

Tommelfingerregel for fortolkning af Kappa.

Kappa-værdi (*k*)	Fortolkning
> 0.90	Fremragende enighed
0.70 - 0.90	Acceptabel enighed
< 0.70	Uacceptabel enighed