



Notat

11. december 2017
Proj.nr. 2004280
EVO/DBN/MT

Dokumentation af måleudstyr

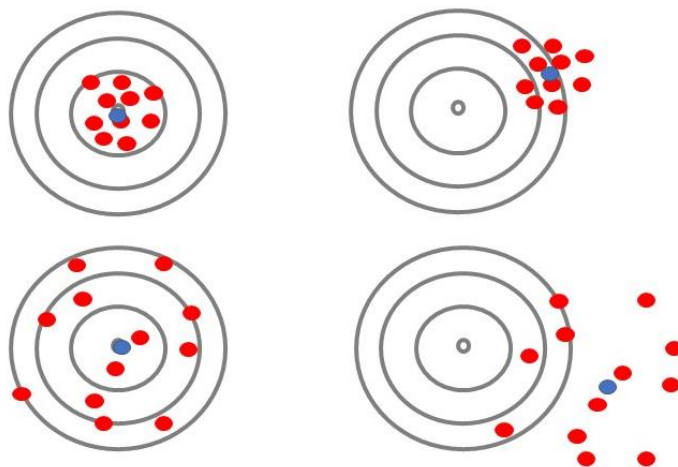
Version 2

Problemstilling En brugers naturlige forventning til et måleudstyr er, at måleresultatet er den sande værdi for det pågældende emne. I praksis vil man ikke kunne opnå fuldstændig ens resultater ved gentagne målinger. Hvor god overensstemmelsen er, vil afhænge af udstyrets *nøjagtighed (accuracy)*, som er det fælles udtryk for tilfældige variationer og eventuel systematik.

Ved udvikling af måleudstyr er det naturligvis vigtigt at kende de faktorer, der kan påvirke nøjagtigheden, så man kan justere designet, så påvirkningen bliver mindst muligt.

Formål Formålet med denne vejledning er at anvise metoder til at beskrive nøjagtighed og metoder til afdække, hvilke faktorer der kan påvirke nøjagtigheden.

Nøjagtighed er synonym med måleusikkerhed (*uncertainty*), som er den hyppigst anvendte betegnelse.



Indhold

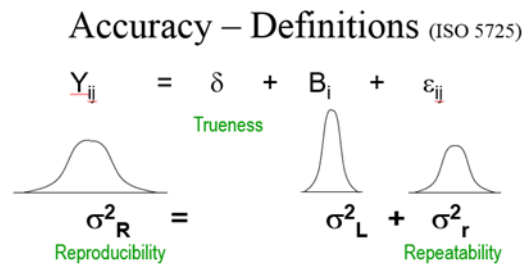
0	Introduktion	3
1	Metoder til beskrivelse af nøjagtighed/måleusikkerhed	3
1.1	Hvordan bruges viden om nøjagtigheden?	4
2	Guide to the Expression of Uncertainty in Measurements (GUM)	5
3	Faktorer med mulig indflydelse på måleresultatet (fiskebensdiagram)	6
4	Produktrisiko (FMEA)	8
5	Robusthed og fejlbudget	8
6	Reference og sporbarhed	9
7	Måleresultater, som er enten "OK" eller "ikke OK"	9
8	Produktrisikovurdering for OK/ej OK – målesystemer	11
9	Stikprøvestørrelse	11
10	Proceskontrol, SPC – kontrolkort	12
11	Referencer og links	13
	BILAG – Eksempler	14
	Forsøgsplan til fastlæggelse af præcision	14
	Bestemmelse af korrekthed	14
	Bestemmelse af reliabilitet – "signal-støj-forhold"	15
	Eksempel – bestemmelse af repeterbarhed, reproducerbarhed og reliabilitet	15
	Forsøgsplan til analyse af robusthed	15
	Youden-design	16
	Eksempel – Faktorer med indflydelse på måling af fedtindhold i trimmings	16
	Eksempel: Beregning af varians på kvotient (uafhængige input-variable)	18
	Eksempel: Fejlbudget for CT-skanning	19

0 Introduktion

<i>Indledning</i>	Vejledningen er baseret på standarden <i>Laboratoriernes analyse- og prøvningsmetoders præcision</i> , som er defineret og beskrevet i den internationale standard, ISO 5725 [1]. Endvidere anvendes Guide to the Expression of Uncertainty in Measurements (GUM) [2].
<i>Gentagelser</i>	I begge standarder kræves data fra gentagelser, som er opnået ved målinger på identisk materiale eller ved gentagne, uafhængige målinger på det samme emne under kontrollerede målebetingelser. "Identisk materiale" kan fx være prøver fra en farsblanding, som er opnået ved udtagning af stikprøver, der repræsenterer hele batchen og derefter homogeniseret, se [9]. Gentagne målinger på det samme emne, fx en slagtekrop, forudsætter, at emnet ikke ændres ved deformation, at temperaturen ikke ændres eller lignende. Det er nødvendigt i hvert enkelt tilfælde at definere, hvad der menes med gentagne målinger.
	1 Metoder til beskrivelse af nøjagtighed/måleusikkerhed En grundlæggende antagelse er, at måleresultater kan antages at følge en normalfordeling – eventuelt efter log-transformation eller lign. transformation. Både standardafvigelse og spredning anvendes som betegnelse for den ene parameter, σ , der beskriver normalfordelingen og er kvadratroden af variansen, σ^2 . Det er almindeligt at beskrive den estimerede standardafvigelse (opnået ved forsøg) med s . Den anden parameter betegnes typisk med μ og den estimerede værdi med for eksempel \bar{x} . Nedenstående definitioner svarer til standarden ISO 5725.
<i>Nøjagtighed (måleusikkerhed)</i>	En målemetodes <i>nøjagtighed (Accuracy)</i> består af metodens <i>korrekthed (Trueness)</i> og <i>præcision (Precision)</i> .
<i>Korrekthed</i>	<i>Korrekthed</i> refererer til overensstemmelse mellem måleresultater og en accepteret referenceværdi og udtrykkes normalt i form af <i>bias</i> (systematisk afvigelse).
<i>Præcisionen</i>	<i>Præcision</i> referer til graden af overensstemmelse mellem måleresultater. Denne opdeles i repeterbarhed og reproducerbarhed og udtrykker henholdsvis den mindste og den største variation i resultaterne.
<i>Repeterbarhed</i>	<i>Repeterbarhed</i> er et mål for overensstemmelsen mellem to måleresultater, som er opnået uafhængigt af hinanden med samme målemetode, med det samme måleudstyr, håndteret af den samme operatør inden for et kort tidsinterval og på det samme materiale. Repeterbarheden angives med repeterbarhedsspredningen, σ_r .

Reproducerbarhed *Reproducerbarhed* er et mål for overensstemmelse mellem to måleresultater, som er opnået uafhængigt af hinanden, med den samme målemetode, med forskellige måleapparater, forskellige operatører, under forskellige, accepterede omstændigheder (tid, temperatur osv.), men på det samme materiale. Reproducerbarheden angives ved reproducerbarhedsspredningen, σ_R , fastlagt ved $\sigma_R^2 = \sigma_L^2 + \sigma_r^2$, hvor σ_L^2 betegner den variation, der er relateret til forskellige udstyr, måleforhold osv.

Figur 1. Definition af præcision (accuracy)



Reliabilitet Ud over måleudstyrets nøjagtighed er det relevant at anføre udstyrets reliabilitet (er ikke standardiseret).

Reliabilitet er defineret ved et måleudstyrs reproducerbarhed set i relation til den naturlige variation i den målte egenskab. Et normalt accepteret krav er, at målingerne højst indeholder 10% - 20% støj, det vil sige, at forholdet er: $\sigma_D^2 / (\sigma_D^2 + \sigma_R^2) > 0,8$, hvor σ_D^2 betegner "populationsvariansen". Men bemærk, der er ikke et formelt acceptniveau.

1.1 Hvordan bruges viden om nøjagtigheden?

Eksempel: måling af pH

pH i den kolde kammuskel måles med en reproducerbarhed estimeret til $s_R = 0,072$. Den naturlige pH-variation er ofte givet ved en spredning (s_D) ikke meget større end 0,1, og i så fald er reliabiliteten under 70% og dermed for ringe.

Præcisionen kan umiddelbart forbedres ved at bruge gennemsnittet af en dobbeltmåling som måleresultat. Reproducerbarheden af en dobbeltmåling kan beregnes ved $s_{R(\text{dobbel})}^2 = s_R^2 - 1/2 s_r^2$. I dette tilfælde udgør repeterbarheden den største del af reproducerbarheden, $s_r = 0,067$, hvorved reproducerbarheden for en dobbeltmåling bliver $s_{R(\text{dobbel})} = 0,054$, og reliabiliteten stiger til 80%.

Eksempel: max. forskel mellem målinger

Hvis den naturlige spredning er større end 0,1, kan enkeltmålinger være tilstrækkelige med hensyn reliabilitet. Men anvendes pH-måling til sortering, er det ikke sikkert, at en reliabilitet på 80% er tilstrækkelig.

En anden måde at udtrykke præcisionen på er ved at beregne den største difference (med 95% sikkerhed) mellem to målinger udført under "reproducerbarhedsbetingelser", dvs. man forestiller sig, at man på et tilfældigt tidspunkt beder to operatører hver tage et tilfæl-

digt (kalibreret) pH-meter og måle pH i en bestemt svinekam. Hvor stor forskel kan man risikere mellem de to målinger?

Denne forskel betegnes i ISO-standarden "R" og beregnes ved

$$R = 1.96 \times \sqrt{2} \times s_R = 0.2.$$

Eksempel: specifikation af krav til præcision

Det er ikke muligt at opstille generelle regler for eller krav til størrelsen af de enkelte parametre (korrekthed, præcision og reliabilitet). I forbindelse med vurdering af målemetoder til sortering af slagtesvin er der udarbejdet en metode til at vurdere værdien af en given præcision (Bayes beslutningsteori [8]). Metoden har erfaringsmæssigt været vanskelig at formidle og anvende i praksis, men er en elegant måde at sammenligne effekten af sortering på basis af måleudstyr med forskellig præcision. Metoden er medtaget her, i håb om at den kan inspirere til metoder, der lettere kan formidles.

Eksempel: konsekvens af præcisionsestimater

Eksempel på forsøgsplan, data og beregning er vist i bilaget. En konkret anvendelse er dokumentation af præcisionen af referencemetode til klassificering af svin, som er beskrevet i artiklen: The estimated accuracy of the EU reference dissection method for pig carcass classification [6]. Resultatet fik konkret indflydelse på EU-regulativet, idet referencemetoden baseret på manuelle dissektioner blev ændret, og præcisionen blev forbedret.

2 Guide to the Expression of Uncertainty in Measurements (GUM)

Guiden GUM [2] definerer måleusikkerhed som "en parameter associeret med et måleresultat, der karakteriserer spredningen af værdier, som med rimelighed kan tillægges måleresultatet".

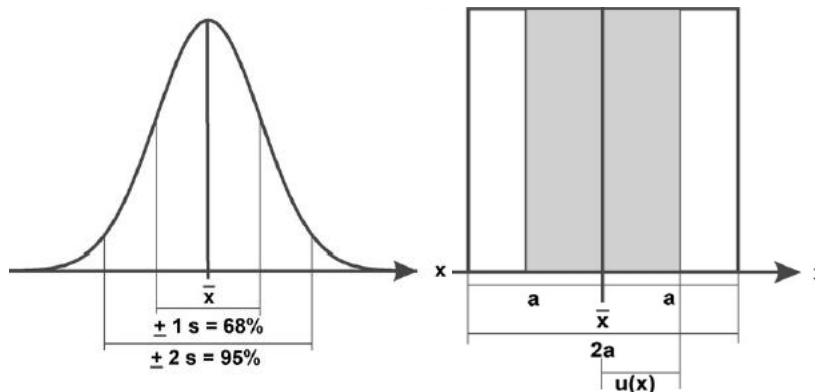
Det omfatter "nøjagtighed" defineret ovenfor. Men GUM giver også forslag til estimeringsmetoder, hvor måleresultaternes variation er vanskelig at fastlægge med forsøg.

Fejltype A og B

GUM kalder estimater, som er fastlagt ved hjælp af forsøg med gentagne målinger, for Type A-estimater. Hvis denne type ikke er mulig at anvende, kan man anvende Type B-estimater, som er "bedste kvalificerede skøn". Det kan være estimater fra artikler, håndbøger, producentoplysninger m.v.

Det mest simple estimat opnås, hvis man kan antage, at mulige værdier tilhører et givet interval med sandsynlighed "1", og at alle værdier i intervallet er lige sandsynlige (ligefordeling). Usikkerheden (varians) estimeres ved: $u^2 = (\text{intervallængde})^2/3$. Flere eksempler kan ses i artikler samlet til GUM-referencen [2].

Figur 2. Fejltype B estimat



Eksempel

Et eksempel på anvendelse af fejltype A- og B-estimer kan ses i artiklen [A review of computed tomography and manual dissection for calibration](#) [5]. Her er indflydelse fra svidningsprocessen estimeret ved type B-estimat.

Kombineret standardafvigelse

Antag, at et måleresultat, Y, er en kombination af uafhængige målinger, $Y=f(X_1, X_2, \dots, X_N)$. Variansestimateret for Y opnås under antagelsen, at f kan tilnærmes ved hjælp af Taylors første ordens rækkeudvikling:

$$u_c^2(y) = \sum_{i=1}^N (c_i u(x_i))^2, \quad \text{hvor } c_i = \frac{\partial f}{\partial x_i}$$

Koefficienterne, c_i , kaldes sensitivitetskonstanter, der angiver effekten på variansen ved ændringer i x_i .

I bilaget findes en "regnemaskine" til bestemmelse af variansen på en kvotient.

Det er samme princip, der anvendes i et fejlbudget, se afsnit 5.

Antagelse om uafhængighed og lineær approksimation

Man kan ikke altid antage, at input-variablene, x 'erne, er uafhængige. Hvis måleresultatet for én variabel er delvis givet af resultatet fra en anden variabel, vil det påvirke variansen, som enten skal vurderes lidt højere (positiv korrelation) eller lidt lavere (negativ korrelation). Tilsvarende kan der være eksempler, hvor højere ordens Taylorudvikling er nødvendig. Der henvises til GUM-manualen [2] i disse tilfælde.

3 Faktorer med mulig indflydelse på måleresultatet (fiskebensdiagram)

Det er vigtigt for fastlæggelse af nøjagtigheden, at man kender de væsentligste årsager til mulige afvigende måleresultater. På baggrund af behov i store produktionsvirksomheder blev et årsag-virkningsdiagram udviklet.

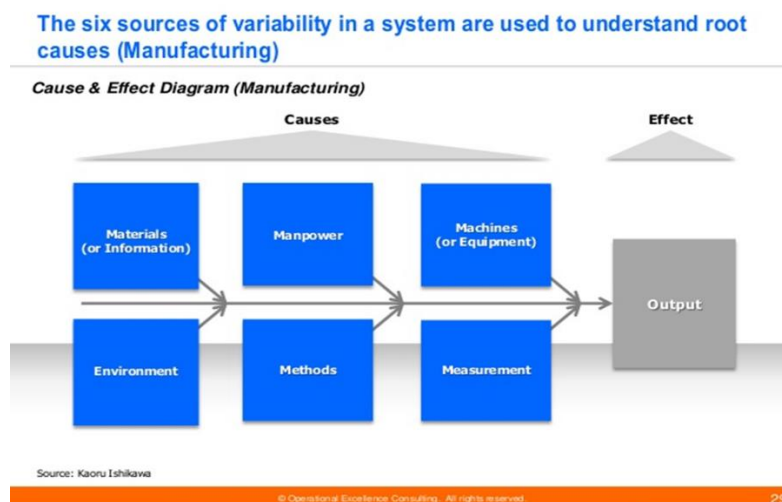
Figur 3. Ishikawa-
diagram

Alias:

Fiskebensdiagram

eller

Årsag-virknings-
diagram



Princippet er beskrevet af Kaoru Ishikawa (1968).

Hver årsag til fejl/afvigelser er kilde til variation. Årsagerne er ofte grupperet i kategorier, eksempelvis de "6 M'er":

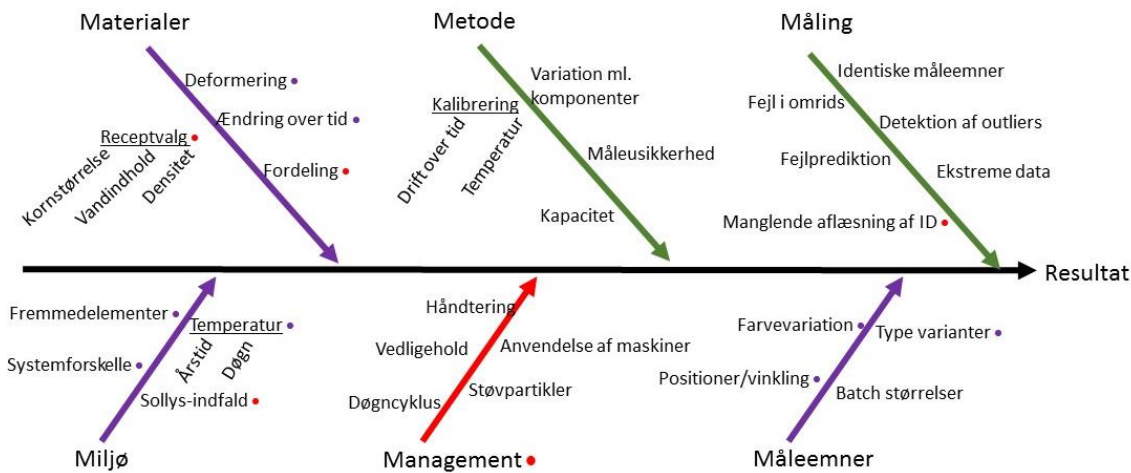
- **M**anpower: Enhver, som er involveret i processen.
- **M**etode: Hvordan processen bliver udført og specifikke krav for at gøre det, så som regler, procedurer og love.
- **M**askiner: Alle udstyr, computere, komponenter, værktøjer osv., som er nødvendig for at gennemføre opgaven.
- **M**aterialer: Råvarer, hjælpestoffer osv., som er nødvendig for at producere det endelige produkt.
- **M**ålinger: Data genereret fra processen med henblik på at evaluere kvaliteten
- **M**iljø: Betingelser som tid, sted, temperatur og kultur der, hvor processen skal fungere.

De nævnte punkter giver ikke nødvendigvis mening i en konkret problemstilling. Til gengæld kan andre årsagsgrupper være relevante.

Eksempel

Som et led i udviklingsarbejdet med visionmåling af foder mængden i en kostald er der udarbejdet et årsag-virkningsdiagram med henblik på at synliggøre faktorer, der kan have indflydelse på måleresultatet. En synliggørelse af faktorer, hvor effekten kan løses softwaremæssigt hhv. skal løses i form af vejledning og begrænsninger (mærket med røde prikker), var et vigtigt led i dialogen med kunden. Det udarbejdede diagram fik en struktur tilsvarende diagrammet i figur 4.

Figur 4. Eksempel på årsag-virkningsdiagram



4 Produktrisiko (FMEA)

Udviklingsarbejde: I udviklingsarbejder kan analysen af faktorer med indflydelse på målingerne bruges til at identificere input til en Failure Mode and Effects Analysis, FMEA. Det er en produktrisikovurdering dvs. en metodologi, hvor formålet er at forudse og foregribe fejl i designfasen ved at identificere mulige fejl i design- eller fremstillingsprocessen. Se [4], som er et link til et regneark til FMEA.

"Nødvendig støj" Hvis indflydelsen fra en faktor ikke kan fjernes, kan man blive nødt til at leve med denne usikkerhedskilde, og indflydelsen skal inkluderes i reproducerbarheden.

Toleranceområde Nogle faktorer indflydelse på måleresultatet kan begrænses ved at fastlægge et toleranceområde. For eksempel kan man opnå en acceptabel måleusikkerhed, hvis blot temperaturen er i et givet interval.

5 Robusthed og fejlbudget

Robusthedstest Ud over de faktorer, som har en kendt indflydelse på målingerne, og som elimineres eller reduceres ved opstilling af toleranceintervaller, kan der være faktorer, hvor man ikke forventer nogen betydende indflydelse på resultatet. For at kunne dokumentere dette gennemføres en robusthedstest, hvor mulige påvirkningsfaktorer indgår.

Test med reduceret design Det er normalt tilstrækkeligt indledningsvis at udføre robusthedstesten ved hjælp af et forsøg med "reduceret design". Se bilag. Faktorerne afprøves i deres yderværdier (lav/høj) under antagelse om ingen/minimale vekselvirkningseffekter. I et udviklingsforløb kan samme design anvendes i et screeningsfor-

søg med henblik på at identificere de faktorer, der har indflydelse på måleresultatet.

Yderligere test

Hvis analysen peger på faktorer med uventet effekt, anbefales en ny test med fokus på disse faktorer. Der anvendes flere og kombinerede niveauer med mulighed for at teste for vekselvirkning mellem faktorerne.

Fejlbudget

Til validering og dokumentation af det endelige måleudstyr udarbejdes et "fejlbudget". Det er en tabel, hvor de forskellige årsager til variation er listet sammen med de estimerede bidrag til måleusikkerheden. Betegnelsen er ikke et opslagsord i standarderne. I GUM findes et eksempel med betegnelsen "Summary of data..." i tabel H.10. Beskrivelse og eksempler kan ses i [3] og [5].

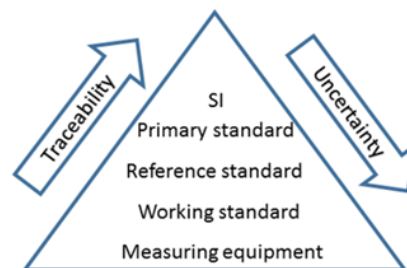
6 Reference og sporbarhed

Beskrivelse af reference

For at kunne dokumentere korrektheden skal dokumentationen omfatte en beskrivelse af referencemetode og -materialer. Hvis referencen er afhængig af *miljøforhold* og *anvendelser*, skal beskrivelsen være opdelt i henhold til dette. Hvis referencen er resultatet af en anden målemetode, skal dennes præcision oplyses.

Sporbarhed

Sporbarhed er en egenskab ved et måleresultat, hvorved resultatet kan relateres til en reference (SI-enhed) gennem en dokumenteret ubrudt kæde af kalibreringer, der hver bidrager til måleusikkerheden.



I mange sammenhænge er sporbarheden vigtig med henblik på at forstå og sikre måleresultatets nøjagtighed. Det gælder i tilfælde, hvor måleresultater fra flere måleudstyr skal kunne sammenlignes, eller hvor måleresultatet skal anvendes som dokumentation overfor kunder eller myndigheder.

7 Måleresultater, som er enten "OK" eller "ikke OK"

Prævalens

Prævalens er et begreb, som anvendes indenfor epidemiologi og udtrykker andelen af en befolkning med en bestemt tilstand ud af den samlede population. Tilstanden er typisk en sygdom, som man testes "positiv" over for, det vil sige at "Ikke OK" = "positiv", hvilket kan være forvirrende. I fødevarerammen bliver prævalensen af forskellige sygdomme bestemt af veterinærkontrollen på slagterierne.

Forekomst af fremmedlegemer

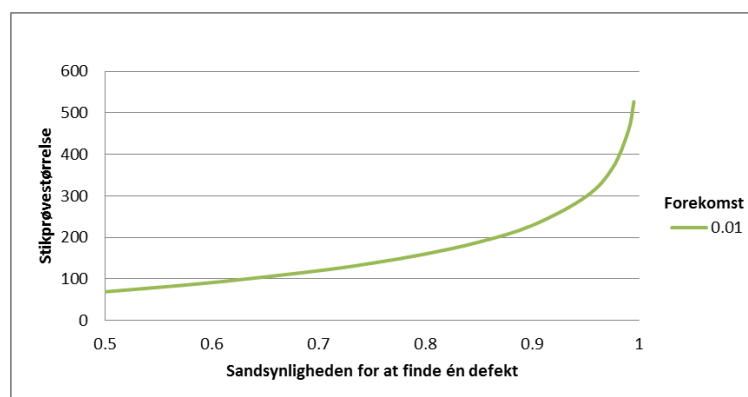
De samme overvejelser kan anvendes til at beskrive problemstillinger, hvor der skal påvises tilstedeværelse af eventuelle fremmedlegemer. Ud over høj sandsynlighed for at detektere tilstedeværelse af fremmedlegemer ønskes en lav frekvens af "falsk positive" (unødvendig frasortering af OK produkter) og "falsk negative" (fremmedlegemer, der ikke detekteres). Udfordringen er ofte, at fremmedlegemerne er vanskelige at detektere og forekommer med lille hyppighed, hvorfor der skal gennemføres mange kontrolmålinger, hvis man vil sikre, at en fremmedlegemedetektor fungerer korrekt.

Dimensionering af stikprøver

Hvis man med 90% sikkerhed vil kunne sikre, at der højst forekommer 1% fejl, skal stikprøven være større end 230 produkter, se figuren nedenfor.

Hvis fejlhyppigheden er 0,1% skal stikprøven være større end 2300. Hvis man vil være næsten 100% sikker, skal stikprøven være større end 5300. Se regnearket [7].

Figur 5. Fastlæggelse af stikprøvestørrelse ved en given forventet forekomst som funktion af ønsket sikkerhed



Hyppighed og usikkerhed

I en stikprøve på n emner, hvor x emner har en given egenskab, kan x antages binomialfordelt (n, p) , hvor p er sandsynligheden for egenskaben. Hyppigheden x/n er et estimat for p og $\sqrt{\frac{x}{n}(1-\frac{x}{n})}$ er estimat af usikkerheden på estimatet af p .

Estimering af "falsk negative" hhv. "positive emner"

Antag, at man med stor sandsynlighed (99%) vil teste, at hyppigheden af "falsk negative" henholdsvis "falsk positive" ved måling med en detektor højst er 5%.

Der skal gennemføres to forsøg. I det ene måles 100 emner, som man med sikkerhed ved er "positive"/"ej OK" og repræsentative for den egenskab, man ønsker at detektere.

I det andet forsøg måles 100 emner, som man med sikkerhed ved er "negative"/"OK". I begge tilfælde forventes højst én fejlmåling. Stikprøvestørrelserne skal naturligvis justeres, hvis forholdet mellem forekomst og sikkerhed er et andet.

Beregning af usikkerhed på prævalens I regnearket [10] er en "regnemaskine" til beregning af usikkerheden på måling af en egenskabs prævalens. Det forudsætter, at man kender niveauet for falsk positive og falsk negative eventuelt fastlagt ved forsøg, se ovenfor.

8 Produktrisikovurdering for OK/ej OK – målesystemer

Ishikawa-diagram og FMEA-analyse I udviklingen af detektionsmetoder er der også behov for at analysere, hvilke faktorer der kan have indflydelse på måleresultatet tillige med en produktrisikovurdering, FMEA. Fremgangsmåden er den samme som beskrevet ovenfor.

9 Stikprøvestørrelse

Hvor stor skal stikprøven være? Det hyppigste spørgsmål, en statistiker får, er: "Hvor stor skal stikprøven være?", og svaret er lige "irriterende" hver gang. "Det afhænger af, hvad du vil vise". Der findes ikke et entydigt svar. Ofte kan hovedspørgsmålet reduceres til et spørgsmål, om der er forskel på to grupper/behandlinger/metoder. I det tilfælde er det "nok" at kende variationen (standardafvigelsen) indenfor en gruppe. Men den viden er ofte ikke kendt. Derfor må man enten basere stikprøveplanen på et skøn og en vurdering af, om en sikker forskel i størrelsesordenen "en konstant x standardafvigelsen", fx en hel eller halv standardafvigelse, er det, man ønsker at opnå.

Sikkerhed eller styrken af stikprøven En stikprøve skal naturligvis være repræsentativ for det fænomen, man ønsker at undersøge. Hvis stikprøven kun består af fem testemner i hver gruppe, så er risikoen for at få et andet resultat ved fem nye test-emner temmelig stor. Erfaringsmæssigt giver 20 test-emner i hver gruppe ca. 80% sikkerhed (styrken af testen), for at man opnår den samme konklusion ved en gentagelse af et eksperiment, som består i at påvise en forskel i størrelsesordenen "én standardafvigelse" mellem to grupper/behandlinger/metoder.

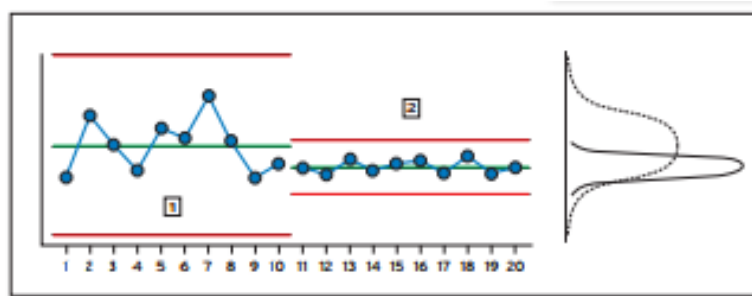
Fastlæggelse af stikprøvestørrelse I regnearket [11] er en "regnemaskine" til bestemmelse af stikprøvestørrelse som beskrevet ovenfor. I samme regneark er et eksempel med "én gruppe". Har man mulighed for at udtage en stikprøve, hvor hvert testemne kan deles i identiske emner med hver sin behandling, opnås en mere sikker test. I det tilfælde analyseres differencen mellem behandlingsresultaterne. Det svarer til, at man tester, at middelværdien i én gruppe har en bestemt værdi, fx 0.

10 Proceskontrol, SPC – kontrolkort

Statistisk proceskontrol, SPC

Analyse og forståelse af variation i processer kan give en statistisk valid indsigt i forandringer eller mangel på samme i den proces, der følges. Der er dog en række forudsætninger for, at analyserne giver et korrekt billede. Mange målinger følger en såkaldt normalfordeling, hvor de fleste målinger ligger lige omkring gennemsnittet. Når målinger vises i et SPC-diagram, vises målingerne i rækkefølge efter tidspunkt eller sekvens af målingerne. Figur 6 viser to perioder med målinger. I begge perioder viser processen alene "normal variation". Men i første periode er variationen stor. I anden periode er det lykkedes at få bedre kontrol over processen (mindre variation). Det ses ved, at afstanden mellem kontrolgrænserne er mindsket.

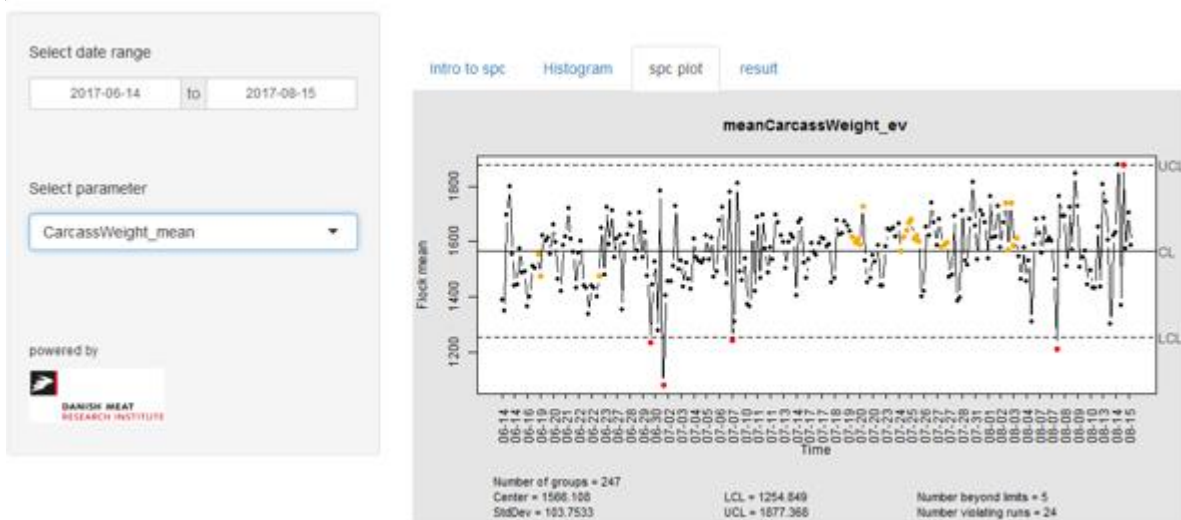
Figur 6, SPC-diagram/kontrolkort



SPC-software

Der findes en række softwareprogrammer til bestemmelse af SPC-diagrammer/kontrolkort, se [12]. Instituttet har i RK-regi udviklet software (R + Rshiny), som enten kan tilpasses en konkret proces eller indbygges i instituttets egne udstyr.

Figur 7, brugergrænseflade til DMRI SPC-program



11 Referencer og links

- [1] ISO 5725 del 1-4. Accuracy (trueness and precision) of measurement methods and results (Biblioteket, udlånt til EVO/DBN)
- [2] Guide to the Expression of Uncertainty in Measurements (GUM)
<Y:\Organization\D05\C028 Målesystemer\ISO standarder, håndbøger, vejledninger\GUM inkl. artikler om usikkerhed>
- [3] Beskrivelse og eksempel på usikkerhedsbudget
<http://www.muelaner.com/uncertainty-budget/>
- [4] FMEA regneark med eksempel
<Y:\Organization\D05\C028 Målesystemer\ISO standarder, håndbøger, vejledninger\Produkt risikoanalyse FMEA.xlsx>
- [5] Eksempel på type A og B variansestimater og fejlbudget
[A review of computed tomography and manual dissection for calibration](#). Meat Science 123 (2017) 35–44
- [6] Eksempel på bestemmelse af nøjagtighed
[The estimated accuracy of the EU reference dissection method for pig carcass classification](#). Meat Science (2006) 22-28
- [7] Regneark til bestemmelse af stikprøvestørrelse for 0-1-variable
<Y:\Organization\D05\C028 Målesystemer\ISO standarder, håndbøger, vejledninger\Stikprøvestørrelse - 0 fejl.xlsx>
- [8] Værdisætning af måleusikkerhed
<Y:\Organization\D05\C028 Målesystemer\ISO standarder, håndbøger, vejledninger\Værdi af sortering.docx>
- [9] Eksempel på udtagning af identiske prøver
<Y:\Organization\D05\C028 Målesystemer\ISO standarder, håndbøger, vejledninger\Jensen K 2016 From fatted calf to food data How to accomplish a representative sampling.pdf>
- [10] Usikkerhed på prævalens
<Y:\Organization\D05\C028 Målesystemer\ISO standarder, håndbøger, vejledninger\Prævalens estimat inkl. usikkerhed.xlsx>
- [11] Stikprøvestørrelse – normalfordelte observationer
<Y:\Organization\D05\C028 Målesystemer\ISO standarder, håndbøger, vejledninger\Stikprøvestørrelse normalfordelte stikprøver.xlsx>
- [12] Proceskontrol. Reference og vejledning til DMRI SPC-program under udarbejdelse.

BILAG – Eksempler

Forsøgsplan til fastlæggelse af præcision

Generelt

Den generelle forsøgsplan er givet ved, at der skal udføres et antal uafhængige gentagelser (repetierbarhed) på en række måleemner, der repræsenterer det relevante måleområde. Alle forhold, der repræsenterer en normal brugssituation, skal indgå, det vil sige flere kopier af det relevante udstyr, eventuelt flere operatører, forskellige tider (indenfor dag hhv. uge) med videre (reproducerbarhed).

Et eksempel på anvendelse er beskrevet i artiklen: The estimated accuracy of the EU reference dissection method for pig carcass classification [6].

Dimensionering af forsøg

Der kan kun udstikkes nogle meget simple, minimale retningslinjer for forsøgets omfang:

- Uafhængige gentagelser: min. 2 pr. måleemne/faktor
- Min. 2 udstyr/komponent
- Hvis operatørbetjent: min. 2 operatører
- Min. 20 måleemner repræsenterende hele måleområdet.

Afvigelse fra idealsituationen

Ideelt skal der opnås mindst 2 måleresultater for alle kombinationer af måleudstyr/operatører osv. ved måling på de samme måleemner. Dette kan rent praktisk være umuligt. Man kan eventuelt udnytte, at højre og venstre side på en slagtekrop kan antages at være ens. Alternativt må man opdele forsøget i blokke, men på en sådan måde at det fortsat er muligt at beregne reproducerbarheden.

Uafhængige målinger

Det er endvidere af fundamental betydning, at målingerne optages uafhængigt af hinanden, det vil sige, at resultatet af én måling ikke må påvirke den næste måling. Hvis målingen er afhængig af operatørens subjektive vurdering, skal målingerne tilrettelægges, så operatørens hukommelse distraheres mest muligt mellem gentagne målinger.

Bestemmelse af korrekthed

Forsøgsplanen er i princippet af samme type som ovenfor, men hvor måleemnerne er erstattet af referencematerialer, eller hvor der for hvert måleemne fastlægges en referenceværdi.

Den gennemsnitlige afvigelse og spredningen på afvigelserne mellem måleresultater og reference beregnes, og overensstemmelsen med kravene, sådan som de er formuleret ovenfor (under *Reference*), undersøges.

Bestemmelse af reliabilitet – "signal-støj-forhold"

Naturlig variation Hvis den naturligt forekommende variation af den målte egenskab, σ^2_D , ikke kendes, kan *præcisions*-forsøget eventuelt tilrettelægges, så måleemnerne udgør en tilfældig repræsentativ stikprøve. Dermed bliver det muligt at beregne et estimat, s^2_D , af "populationsvariansen".

Reliabiliteten estimeres ved $s^2_D/(s^2_D+s^2_R)$, og denne størrelse bør – som tommelfingerregel – være mindst 0,8-0,9.

Eksempel – bestemmelse af repeterbarhed, reproducerbarhed og reliabilitet

I eksemplet indgår to udstyr (A og B), fire måleemner, som er målt med begge udstyr til to forskellige tider, og hver gang er der målt to gange.

		Udstyr	
		A	B
Tid 1		5.3992	5.73041
		5.53899	5.76617
Tid 2		5.80821	5.60094
		5.85806	5.61357

Emne 1-4

$$\text{Måleresultat}_{ijk} = \alpha + U_i + T_j + E_k + \varepsilon_{ijk}$$

- Beregning af**
- **Repeterbarhed** Data antages at følge en normalfordeling med middelværdi " α " og en varians beskrevet af fire stokastiske led, $U \approx$ udstyr A og B, $T \approx$ tid 1 og 2, $E \approx$ emne 1,2,3,4 og $\varepsilon \approx$ gentagelse 1 og 2. Variablene antages indbyrdes uafhængige og normalfordelte med middelværdi 0 og varianser σ_U^2 , σ_T^2 , σ_E^2 og σ^2 . Repeterbarheden, σ_r^2 , estimeres med σ^2 ($=0,06221$), reproducerbarheden, σ_R^2 , med $\sigma_U^2 + \sigma_T^2 + \sigma^2$ ($0,08677 + 0,0106 + 0,06221 = 0,15958$). De fire emner er ikke tilstrækkelige til at dække hele variationsområdet, men for eksemplets skyld bestemmes reliabiliteten ved $\sigma_E^2/(\sigma_E^2 + \sigma_R^2)$ ($0,09959/(0,09959+0,15958) = 0,38$). Reliabiliteten er i dette tilfælde 38%. Da tommelfingerreglen er "80%", er målemetoden derfor ikke tilstrækkelig præcis til at differentiere mellem måleemnerne.

Kommentar I eksemplet er kun medtaget hovedvirkninger. Det kan være relevant at medtage vekselvirkninger som stokastiske reproducerbarheds-effekter, hvilket vil reducere repeterbarheden.

Forsøgsplan til analyse af robusthed

Formålet er at kunne dokumentere, at forventede påvirkningsfaktorer kun påvirker måleresultatet inden for et acceptabelt niveau. Alternativt at kunne udpege de faktorer, som har størst indflydelse på måleresultatet.

Hvis blot 5 faktorer på 2 niveauer skal analyseres, inklusive vekselvirkninger, skal der udføres 2^5 forsøg med mindst 2 gentagelser/måleemner for at få alle aspekter belyst. Derfor anbefales at starte med en screening af faktorerens betydning.

Test med reduceret design

Det er normalt tilstrækkeligt indledningsvis at udføre robusthedstesten ved hjælp af et forsøg med "reduceret design". Hvis analysen peger på faktorer med uventet effekt, anbefales en ny test med fokus på disse faktorer. Der anvendes flere niveauer og gentagelser, så analysen blandt andet kan omfatte test af eventuelle vekselvirkninger mellem faktorerne.

Youden-design

7 faktorer og 8 forsøg

Erfaringerne har vist, at det i første omgang er tilstrækkeligt at få identificeret hovedvirkningerne af op til 7 faktorer på to niveauer i et "Youden-design" med kun 8 forsøg. Hvert forsøg har mindst to gentagelser. Forsøgene er fastlagt ved bestemte kombinationer af faktorniveauer, som beskrevet nedenfor. Det er vigtigt at anvende præcis disse kombinationer. Derved kommer hvert niveau af hver faktor til at optræde præcis 4 gange og således, at de øvrige faktorer alle forekommer præcis to gange hver på hvert niveau. Designet er i "balance", og data kan analyseres med en almindelig variansanalyse kun med hovedvirkninger.

Til forsøget anvendes et antal (mindst to) måleemner eller referencematerialer, som skal indgå i alle delforsøg. Alternativt kan afvigelse fra en referenceværdi anvendes, se eksempel nedenfor. Hvis faktorerne ingen indflydelse har, vil man således få samme resultat i alle delforsøg. I variansanalysen vurderes en faktors indflydelse i forhold til gentagelsesvariationen.

Hvis det ikke er muligt at definere 7 relevante faktorer, men kun 6 eller 5, da indsættes dummy-faktorer. Ved færre faktorer bør et andet design overvejes.

Eksempel – Faktorer med indflydelse på måling af fedtindhold i trimmings

De syv faktorer

Et måleudstyr til analyse af fedtindhold i trimmings ønskes dokumenteret. Med henblik på at undersøge om CT-skanning kan bruges som en hurtig referencemetode, blev 8 prøver af trimmings (4 med oksekød, 4 med svinekød) udtaget. To af prøverne af hver type havde højt fedtindhold, de andre to havde lavt indhold. Halvdelen af prøverne blev frosset og optøet, den anden halvdel var frisk. Den ene halvdel af prøverne blev vakuumpakket, den anden halvdel pakket uden vakuum. Skanning blev foretaget med enten høj eller lav opløsning, og billedbehandlingen blev baseret på hele prøven eller uden randpixels (svarende til 2/3 af volumen). Håndteringen blev udført i overensstemmelse med beskrivelsen nedenfor:

Faktor		Niveau 1		Niveau 2
Trimming	A	Coarse	a	Fine
Temp.	B	Defrosted	b	Fresh
Fat_level	C	Low	c	High
Type	D	Beef	d	Pork
Resolution	E	10 mm	e	1 mm
Volumen	F	Total	f	2/3
Air	G	+vacuum	g	-vacuum

Youden-design

1. delforsøg	A	B	C	D	E	F	G	Eksempel: Niveau A 2 x b, 2 x B, 2 x c, 2 x C osv. Niveau a 2 x b, 2 x B, 2 x c, 2 x C osv.
2. delforsøg	A	B	c	D	e	f	g	
3. delforsøg	A	b	C	d	E	f	g	
4. delforsøg	A	b	c	d	e	F	G	
5. delforsøg	a	B	C	d	e	F	g	
6. delforsøg	a	B	c	d	E	f	G	
7. delforsøg	a	b	C	D	e	f	G	
8. delforsøg	a	b	c	D	E	F	g	

Måleresultater

Hver prøve blev analyseret i laboratoriet, og måleresultatet var differencen mellem målt fedtindhold med CT-skanner og laboratoriets resultat, se nedenfor:

A	B	C	D	E	F	G	Gen- tagelse	Måle- resultat
Trimming	Temp.	Fat_level	Type	Resolution	Volumen	Air		
Coarse	Defrosted	Low	Beef	10 mm	Total	+vacuum	1	-0,21746
Coarse	Defrosted	Low	Beef	10 mm	Total	+vacuum	2	1,733
Coarse	Defrosted	High	Beef	1 mm	2/3	-vacuum	1	2,74408
Coarse	Defrosted	High	Beef	1 mm	2/3	-vacuum	2	3,14056
Coarse	Fresh	Low	Pork	10 mm	2/3	-vacuum	1	1,25512
Coarse	Fresh	Low	Pork	10 mm	2/3	-vacuum	2	1,07945
Coarse	Fresh	High	Pork	1 mm	Total	+vacuum	1	-2,91237
Coarse	Fresh	High	Pork	1 mm	Total	+vacuum	2	-1,71545
Fine	Defrosted	Low	Pork	1 mm	Total	-vacuum	1	2,67569
Fine	Defrosted	Low	Pork	1 mm	Total	-vacuum	2	3,27708
Fine	Defrosted	High	Pork	10 mm	2/3	+vacuum	1	-1,82605
Fine	Defrosted	High	Pork	10 mm	2/3	+vacuum	2	-1,31293
Fine	Fresh	Low	Beef	1 mm	2/3	+vacuum	1	1,96764
Fine	Fresh	Low	Beef	1 mm	2/3	+vacuum	2	1,65452
Fine	Fresh	High	Beef	10 mm	Total	-vacuum	1	2,189
Fine	Fresh	High	Beef	10 mm	Total	-vacuum	2	1,79091

Dataanalyse

Måleresultaterne blev analyseret med en almindelig variansanalyse med hovedvirkninger, se tabellen nedenfor. Residual-std.dev. (RMSE) blev bestemt til 0,627817, og tre af faktorerne havde signifikant betydning. Fedt-niveau, type (okse/gris) og vakuum/ikke vakuum havde betydning for måleresultatet.

Konklusion

Det betyder, at CT-skanning kun kan anvendes som referencemå-
 tode, hvis metoden specificeres til enten okse- eller svinekød. I
 måleprotokollen skal det besluttes, om kødet skal vakuumpakkes
 eller ej. Endelig kræves en kalibrering mod laboratoriebestemt
 fedtindhold, da selve fedtniveauet har betydning.

Source	DF	Sum of Squares	Mean Square	RMSE	F Value	Pr > F
Model	7	53.96075050	7.70867864		19.56	0.0002
Error	8	3.15323381	0.39415423	0.627817		
Corrected Total	15	57.11398430				

Source	DF	Mean Square	F Value	Pr > F	Signifikans
Trimming	1	1.76154611	4.47	0.0674	
Temp.	1	1.50378839	3.82	0.0866	
Fat_level	1	8.01918752	20.35	0.0020	**
Type	1	13.10749166	33.25	0.0004	***
Resolution	1	2.35676996	5.98	0.0402	
Volumen	1	0.22136507	0.56	0.4751	
Air	1	26.99060178	68.48	<.0001	***

Eksempel: Beregning af varians på kvotient (uafhængige input-variable)

“Regnemaskine” til
 bestemmelse af vari-
 ans på kvotient af to
 uafhængige stokasti-
 ske variable

Baseret på Taylors første ordens rækkeudvikling kan variansesti-
 matet for kvotienten $G(x, y) = z = \frac{x}{y}$ beskrives ved:

$$s^2(G(x, y)) = \left(\frac{\partial G}{\partial x}\right)^2_{\bar{z}} \times s^2(x) + \left(\frac{\partial G}{\partial y}\right)^2_{\bar{z}} \times s^2(y)$$

$$= \left(\frac{1}{y}\right)^2 \times s^2(x) + \left(-\frac{x}{y^2}\right)^2 \times s^2(y)$$

Regneark: Indtast gennemsnit og spredning for tæller (x) og nævner (y) i de
 lysegrønne felter. Resultat i det mørkegrønne felt.

x	s(x)	s²(x)	y	s(y)	s²(y)	
522	38.1	1451.61	1716	70.2	4928.04	
$1/y^2$	$(1/y^2)s^2(x)$	$(-x/y^2)^2$	$(-x/y^2)^2s^2(y)$		$s^2(z)$	
3.39598E-07	0.000492964	3.14248E-08	0.000154862		0.000647827	
					s(z)	
					0.025452439	

Eksempel: Fejlbudget for CT-skanning

Fremgangsmåde ved fastlæggelse af nedenstående fejlbudget er beskrevet i artiklen: "A review of computed tomography and manual dissection for calibration" [5]. Der indgår eksempler på variansestimater af både type A og B, jf. GUM-vejledningen [2].

Computed tomography – segmentation methods		
Error sources	Uncertainty estimate	Variance contribution
Scanner variation	Type B: $a = 0.5$	0.08
Measuring protocol • Voxel size: $1 \times 1 \times 10 \text{ mm}^3$ vs. $1 \times 1 \times 1 \text{ mm}^3$	Type A: Std.dev. = 0.33 Christensen et al. (2006)	0.11
Measuring protocol • Volt/Amp: 100 kV/110 mA vs. 140 kV/80 mA	Type A: Std.dev. = 0.68 Christensen et al. (2006)	0.46
Measuring protocol • Rind side up vs. rind side down	Type A: Std.dev. = 0.32 Christensen et al. (2006)	0.10
Type of scanning method	Type A: Std.dev. = 0.48 Based on Table 4	0.23
Type of segmentation	Type A Based on Table 5	0.18
Traceability	Type A: Std.dev. = 0.54	0.29
Density estimates	Type A: Std.dev. = 0.49 Based on Table 6	0.24
Total variance contribution assuming independency		$1.69 = 1.30^2$