De logica van de onzekerheid (2) Simon van der Salm

Onzekere tijden

In deze tijd van Corona blijkt maar al te duidelijk dat er slechts één zekerheid bestaat, namelijk… dat niets zeker is. Moderne, westerse mensen, opgegroeid in een gevaar- en risicoloze samenleving (geloven ze maar al te graag) kunnen heel slecht overweg met onzekerheid. En dat terwijl de statistiek heel handige hulpmiddelen biedt om de grootte van die onzekerheid in te schatten.

Onzekerheidsinterval bij verkiezingen

Ik vat nog even samen. Elke waarneming van de waarde van een grootheid is omgeven door een onzekerheidsinterval. Chris noemt dat hierboven *bandbreedte*, een zeer toepasselijk woord. Ik heb dat in MIR84 uitgelegd aan de hand van verkiezingspeilingen. Als de peiling zegt dat bijvoorbeeld 30 zetels zullen gaan naar politieke partij X, dan zit daarom heen een onzekerheidsinterval van bijvoorbeeld ± 2 zetels. Je hoort dus eigenlijk te zeggen 30 ± 2 zetels. Die ±2 zetels is afhankelijk van die 30! Voor een veel kleiner aantal dan 30 geldt een kleiner onzekerheidsinterval. Je kunt dat gewoon uitrekenen, maar voor journalisten geldt het getal 2 gemakshalve voor alle partijen. Zo stond de partij BBB, in de afgelopen maand maart, in de exit poll van 21:00 uur, vlak na de verkiezingen op de verkiezingsdag, op 1 zetel in de peilingen, eh… ± 2, en niemand in de TV-uitzending vond dat een numeriek hilarische schatting.

Betrouwbaarheid

Maar er is ook nog sprake van betrouwbaarheid. In de meeste (vooral sociaalwetenschappelijke) toepassingen is die 95%. Dat is de zogenaamde *2 sigma betrouwbaarheid*. Die leidt tot een vrij grote kans op feilbaarheid, maar in de niet-natuurkundige disciplines is, als gevolg van allerlei onzekere maatschappelijke parameters, streven naar een grotere betrouwbaarheid tot mislukken gedoemd.

In de industrie hanteert men steeds vaker ruim 99%, de 3 sigma betrouwbaarheid. Er zijn kwaliteitssystemen die zelfs uitgaan van 6 sigma betrouwbaarheid: slechts bij 3,4 van de miljoen producten (of diensten) mag iets fout gaan.

De 2 sigma betrouwbaarheid betekent dat 95% van de onzekerheidsintervallen ook daadwerkelijk de werkelijke waarde van de meetgrootheid bedekt. Dus als de exit polls zeggen 30 ± 2 zetels wil dat niet met zekerheid zeggen dat de werkelijke verkiezingsuitslag in dat interval zal liggen. Als er in de werkelijke verkiezingsuitslag 26 zetels voor partij X worden geteld, is dat niet een aanwijzing dat de statistici hun werk niet goed hebben gedaan, maar gewoon een gevolg van de 5% feilbaarheid van het onzekerheidsinterval.

Enfin, we zeggen dus 30 ± 2 zetels met een betrouwbaarheid van 95%. Sorry, maar beter en zekerder heeft de wetenschap niet te bieden.

Dat er wel wat wiskundig werk achter het bepalen van het onzekerheidsinterval zit, moet u maar voor lief nemen. Op grond daarvan is in de jaren zestig, door o.a. Amerikaanse producent Pickett, een aantal statistische rekenlinialen ontwikkeld, o.a. de Pickett 525 StatRule. Zie mijn artikel daarover in de *Journal of the Oughtred Society, Spring 2017*. Zie [2].

Figuur 1 toont de uitgebreidere Pickett 6-T voor statistische kwaliteitscontrole. Een betere afbeelding kunt u vinden op de website van het Sliderulemuseum.

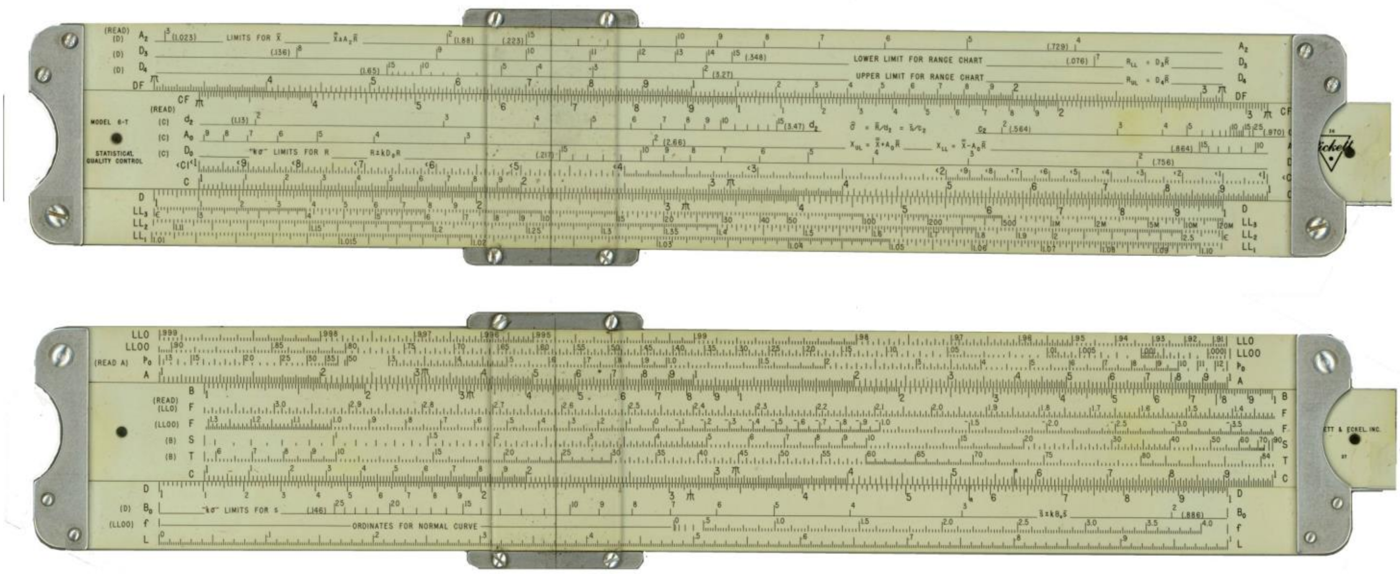

Fig. 1. De Pickett 6-T Statistical Quality Control. Bron: https://www.sliderulemuseum.com/Pickett/Pickett_Model_6-T_Statistical_OrionMillerCollection.jpg.

Foutbalkjes
Het onzekerheidsinterval wordt wel aangeduid met *foutbalkjes*. Zie figuur 2 met een voorbeeld uit de natuurkunde.

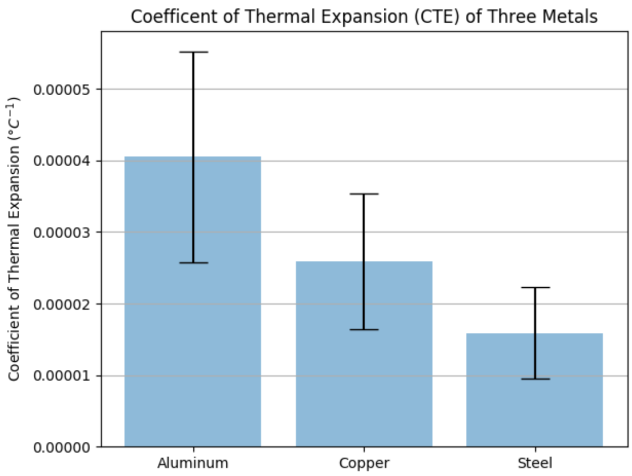

Fig. 2. Foutbalkjes om de onzekerheidsintervallen van drie empirische waarden aan te duiden. Merk op, dat de foutbalkjes verschillende lengtes hebben. Hoe groter de meetwaarde, hoe groter het balkje.
Bron: https://pythonforundergradengineers.com/python-matplotlib-error-bars.html

Figuur 3 laat een oneindige serie, *verticale* onzekerheidsintervallen (gearceerde gebied, een zogenaamde pluim) van een tijdreeks zien. Merk ook hier op, dat de grootte van een onzekerheidsinterval groter is, naarmate de empirische waarde groter is. Onzekerheidsintervallen hebben dus geen constante lengte!

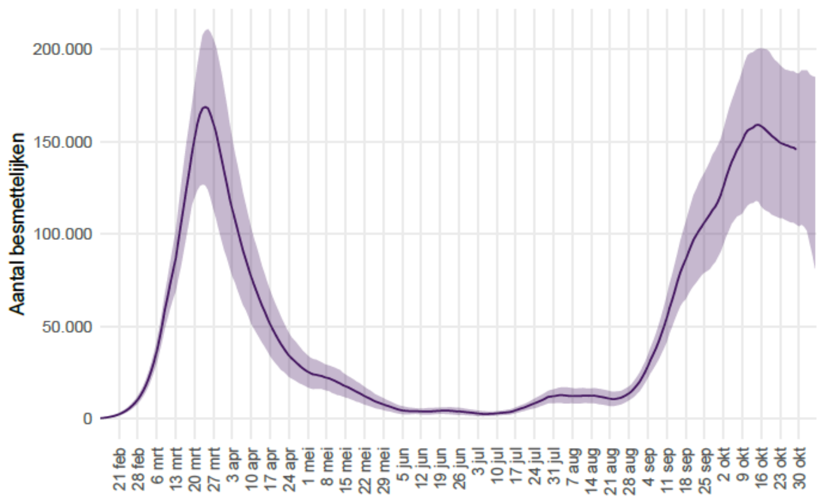

Fig. 3. Het aantal Covid-besmettingen per week. Ieder weekgetal heeft zijn eigen onzekerheidsinterval.
Bron: https://www.ergomag.eu/vaker-en-langer-verzuim-tijdens-tweede-golf/

Eisen aan leveranciers
Tegenwoordig worden er strikte kwaliteitseisen gesteld aan leveranciers die bijvoorbeeld halffabricaten maken voor andere producenten. Zie [3]. Als je bijvoorbeeld metalen boutjes levert aan de auto-industrie, dan moet je kunnen bewijzen dat de lengte ervan conformeert aan de specificaties die de afnemer oplegt. Zelfs hoe getallen moeten worden afgerond en uitgedrukt in significante cijfers is in zulke confirmaties in detail beschreven. Een fraai voorbeeld is het document E29-13 van de ASTM, de *American Society for Testing Materials*. Zie [1]. Chris wees er al op: zulke dichtgetimmerde specificaties zijn vooral bedoeld om allerlei juridische moeilijkheden/aansprakelijkheden te voorkomen.

Significante cijfers
De onzekerheid van via empirie bepaalde waarden (direct gemeten of indirect berekend) hangt samen met het aantal significante (betekenisvolle) cijfers dat gebruikt mag worden. De onzekerheid wordt meestal uitgedrukt in één cijfer, maximaal in twee cijfers, omdat de onzekerheid in de onzekerheid zo

groot is dat het geen zin heeft om meer cijfers te gebruiken. Dus een uitdrukking als 2,34 ± 0,05678 V is dus onzin. Correct is 2,34 ± 0,06. D.w.z. het meest rechter cijfer 4 van de drie significante cijfers 2,34 is onzeker.

Absolute limieten

Chris wijst hierboven op een eigenaardigheid waarmee je als producent te maken kunt krijgen. Soms stelt de afnemer een absolute limiet voor een parameter, zoals: de lengte van een boutje dat je levert aan een autofabrikant mag beslist niet groter zijn dan 2,5 cm. Elke meetwaarde (niet afgerond) of elke berekende waarde die ook maar enigszins groter is dan 2,500… conformeert dan niet aan de specificaties en wordt afgekeurd.

Als, zoals in het voorbeeld van Chris, de absolute minimum limiet 1,0 m bedraagt, mag je dan een lengte als 0,98 m nog goedkeuren? In het vak van Chris kan ik me voorstellen dat men hoofdschuddend naar zulke discussies luistert. Maar in de ziekenhuizen is een maximale lekstroom van 100 μA toch echt een zinvolle, absolute, maximum limiet.

Afrondingsmethode

Als zo'n absolute methode als hierboven niet noodzakelijk is kan de afnemer grenzen formuleren als afrondingen volgens een bepaalde procedure. Het ASTM-document [1] beschrijft in detail dergelijke procedures. Een beperkt aantal cijfers wordt dan als significant beschouwd in het licht van de confirmatie met de specificaties. Men rondt dan bijvoorbeeld een waargenomen of berekende druk af op de dichtstbijzijnde kPa, een weerstand op de dichtstbijzijnde 10 ohm of een relatieve meetafwijking op de dichtstbijzijnde 0,1%, enzovoorts. In de civiele techniek van Chris lijkt me dat een zinvolle methode. In zo'n geval gaat men er van uit dat de bandbreedte maximaal ± 0,5 maal de decimale plaatswaarde is van het eerste weggelaten cijfer. Dus in het geval van Chris is het onzekerheidsinterval 1,00 ± 0,05. Merk de extra 0 op! De waarde 0,98 valt daarbinnen.

Statistisch onzekerheidsinterval

Veel afnemers eisen van de leverancier bovendien een statistische analyse van de kwaliteit van het productieproces. Dat is tegenwoordig niet meer zo'n probleem als je honderden of duizenden 'identieke' producten maakt en het meet- en rekenwerk (Shewhart's analyse) continu door computers laat doen. Zie [3]. Bijvoorbeeld het produceren van boutjes voor de auto-industrie. Dan is de standaardafwijking *s* van de één of andere waarnemingsreeks een maat voor de precisie. Daar komt het woord *sigma* in de kwaliteitscontrole vandaan. Stel bijvoorbeeld dat je een lengte van 1,45729 cm hebt berekend en je weet dat *s* = 0,0052 cm. Dan formuleer je de waarde als 1,457 cm, dus afgerond op de dichtstbijzijnde 0,001 cm, omdat het meest significante cijfer van *s* op de plaats van de duizendsten staat. Het onzekerheidsinterval is dan 1,457 ± 0,005 cm. Dat 1 sigma interval heeft een betrouwbaarheid van 'slechts' ongeveer 68%. Zie [1]. Met een vermenigvuldigingsfactor (bedekkingsfactor) kun je het onzekerheidsinterval verlengen, waardoor – paradoxaal - de meetonzekerheid weliswaar toeneemt, maar de betrouwbaarheid wordt vergroot.

Referenties

1. ASTM E29-13, *Standard Practise for Using Significant Digits in Test Data to Determine Conformance with Specifications*.
2. Salm, S.A.M. van der, *The Pickett N-525 StatRule Statistical Scales*, Journal of the Oughtred Society, Volume 26, Number 1, Spring 2017.
3. Montgomery, D, C, *Statistical Quality Control*, 6th edition, John Wiley and Sons Inc, 2009.