



*Statistical Control of Shewhart Control Charts*

R. Goedhart

# Samenvatting

---

De effecten van het schatten van parameters in statistische procesbeheersing en de bijbehorende correcties voor regelkaarten (*control charts*) zijn het onderwerp van dit proefschrift. Aangezien de beschikbaarheid van data toeneemt in de huidige tijden, neemt ook de behoefte toe voor methoden om grote datastromen te monitoren. In het vakgebied SPM (*statistical process monitoring*, statistische procesbeheersing) staat het detecteren van mogelijke veranderingen binnen een proces door het monitoren van procesdata centraal. Een belangrijk hulpmiddel voor deze detectie is de regelkaart. Dit proefschrift richt zich op een specifiek type regelkaart, namelijk de *Shewhart* regelkaart.

Elk proces is onderhevig aan variatie. Denk bijvoorbeeld aan je eigen lichaam, waarin je hartslag, bloeddruk, en vele andere karakteristieken variëren over de tijd. Het is vaak erg lastig om de precieze oorzaak van deze verschillen aan te wijzen, aangezien het om combinaties van vele kleine invloeden gaat. Specifieke gebeurtenissen of verstoringen kunnen echter het onderliggende proces veranderen, wat een ander soort variatie tot gevolg heeft. Een virus of ziekte kan bijvoorbeeld invloed hebben op bepaalde karakteristieken in ons lichaam. Het vroeg detecteren van deze gebeurtenissen of verstoringen maakt het mogelijk om tijdig tegenmaatregelen te nemen om de invloed te beperken.

Om het detecteren van veranderingen in een onderliggend proces mogelijk te maken, moet eerst een onderscheid gemaakt worden tussen twee soorten variatie, variatie door normale oorzaken (*common cause variation*) en variatie door bijzondere oorzaken (*special cause variation*). Eerstgenoemde is variatie die inherent is aan het proces, terwijl laatstgenoemde veroorzaakt wordt door speciale gebeurtenissen. Wanneer een proces slechts variatie door normale oorzaken bevat, wordt er gesproken van een *in-control* (beheerst) proces. Als er speciale oorzaken aanwezig zijn, dan wordt gesproken van een *out-of-control* proces. De eerste stap voor het detecteren van mogelijke out-of-control situaties, is het bepalen van de in-control situatie van het proces. Ver-

volgens worden voor de Shewhart regelkaart twee regelgrenzen (*control limits*) geconstrueerd, de UCL (*upper control limit*) en de LCL (*lower control limit*), waarmee bepaald kan worden of een proces in-control is of niet.

Oorspronkelijk werd de in-control situatie van een proces bekend verondersteld. Dit gebeurde door de aanname dat het gemiddelde en de standaardafwijking van een proces bekend zijn bij het opstellen van de UCL en LCL. In dat geval is de prestatie van de regelkaart, in termen van de FAR (*false alarm rate*, de kans op een onterecht out-of-control signaal) of de ARL (*average run length*, het gemiddelde aantal waarnemingen tot een out-of-control signaal), een bekende constante waarde. In de praktijk zijn het gemiddelde en de standaardafwijking van een proces echter onbekend, evenals de werkelijke verdeling van de data, en moeten deze worden geschat. Het schatten van de in-control situatie wordt *Phase I* genoemd.

Het schatten van de in-control situatie in Phase I wordt over het algemeen gedaan met behulp van een representatieve steekproef. Aangezien verschillende gebruikers verschillende steekproeven zullen verzamelen, betekent dit dat zij ook verschillende schattingen verkrijgen. Dit heeft tot gevolg dat de geschatte UCL en LCL, en de bijbehorende regelkaart prestatie, ook verschillen tussen de gebruikers. Het verzamelen van grotere steekproeven reduceert deze variatie tussen verschillende gebruikers, maar de benodigde steekproef omvangen voor een adequate regelkaart prestatie zijn vaak niet beschikbaar. Om die reden richt dit proefschrift zich op aanpassingen van de regelgrenzen ter compensatie van het effect van het schatten van parameters in Phase I.

Hoofdstuk 2 gaat over Shewhart regelkaarten voor locatie (bijvoorbeeld gemiddelde) voor normaal verdeelde data. Hier worden correcties gegeven zodat een gewenste regelkaart prestatie wordt verkregen door gebruikers in verwachting. Dit criterium wordt ook wel het *bias criterion* genoemd, en relateert zich aan de onconditionele (verwachte) prestatie van de regelkaart. De motivatie achter deze correcties is dat wanneer parameters worden geschat, de verwachte regelkaart prestatie niet gelijk is aan de nominale (ontwerp) waarde. De afleidingen in dit hoofdstuk zijn te gebruiken voor verscheidene prestatie indicatoren, zoals de FAR of de ARL. De gerelateerde correcties voor regelkaarten voor spreiding (zoals de standaardafwijking) zijn te vinden in Diko et al. (2017). Een nadeel van sturen op verwachting is echter dat er nog steeds een relatief groot aantal gebruikers kan zijn met een ontoereikende regelkaart prestatie.

---

Hoofdstuk 3 gaat ook over Shewhart regelkaarten voor locatie voor normaal verdeelde data, maar hier in combinatie met het zogeheten *exceedance probability criterion*. Dit criterium richt zich op de conditionele prestatie van de regelkaart, en houdt in dat tenminste een specifieke conditionele prestatie bereikt wordt met een vooraf gedefinieerde kans. In dit hoofdstuk worden de benodigde aanpassingen aan de regelgrenzen gegeven om dit te bereiken. Ook wordt er een vergelijk gemaakt met andere vergelijkbare methoden. Daarnaast wordt de out-of-control prestatie van de regelkaart onderzocht voor verschillende scenario's. Gezien het belang van de conditionele prestatie in de evaluatie van regelkaarten, richten ook de volgende hoofdstukken zich op dit criterium.

Hoofdstuk 4 gaat over aangepaste regelgrenzen voor regelkaarten voor spreiding. Wanneer data normaal verdeeld zijn, dan volgen veel schatters van spreiding een chi-kwadraat verdeling. In dat geval zijn de symmetrische regelgrenzen uit de voorgaande hoofdstukken niet geschikt. Daarom worden in dit hoofdstuk aangepaste regelgrenzen afgeleid die rekening houden met de chi-kwadraat verdeling van de schatter voor spreiding. Via een zelfde aanpak is het mogelijk om aangepaste regelgrenzen te bepalen voor het monitoren van de locatie van data die niet normaal verdeeld zijn. Echter, de in-control verdeling van de data is over het algemeen onbekend, en moet geschat worden samen met de bijbehorende parameters.

Hoofdstuk 5 gaat daarom over nonparametrische regelgrenzen, evenals methoden die de theorie voor normale data toepasbaar maken. Een voordeel van nonparametrische regelkaarten is dat ze niet afhankelijk zijn van de verdeling van de data, en daardoor ook kunnen worden toegepast voor het monitoren van verschillende statistieken. Ze kunnen dus toegepast worden bij regelkaarten voor zowel locatie als spreiding. Het nadeel is dat de benodigde steekproef omvang om een specifieke regelkaart prestatie te behalen groter is. Als deze steekproef omvang niet beschikbaar is, moeten er meer restricties op de dataset geplaatst worden, of moet een mindere regelkaart prestatie worden geaccepteerd. Wanneer geaggregeerde steekproef statistieken gebruikt worden in plaats van individuele waarnemingen, is het een redelijk alternatief om een hogere FAR als minimum prestatie te accepteren.

Samenvattend bevat dit proefschrift verschillende aangepaste regelgrenzen voor Shewhart regelkaarten om te corrigeren voor het effect van het schatten van parameters. De correcties zijn afgeleid voor zowel onconditionele (in

verwachting) als conditonele prestatie indicatoren. We raden aan om bij het opstellen van de regelgrenzen te concentreren op de conditionele prestaties. Indien genoeg data beschikbaar zijn, adviseren wij om nonparametrische regelgrenzen te gebruiken, aangezien deze een betere algemene prestatie leveren voor verschillende verdelingen. Wanneer de steekproef omvang beperkt is, moet er gekozen worden tussen het aanscherpen van de restricties op de data, of het accepteren van een mindere regelkaart prestatie. De bevindingen en afwegingen in dit proefschrift zijn ook relevant in vele andere statistische toepassingen.