



## KPI'S MET WEIBULL

### Het schatten van een kansverdeling uit een rij met data.

Hoe vaak komt het niet voor dat er gegevens beschikbaar zijn om kpi's mee te berekenen? Een setje data met levertijden, doorlooptijden, bewerkingstijden, of misschien wel de termijn waarop een onderdeel stuk gaat of correctief onderhoud vraagt. Vaak worden hierop gemiddelden gerapporteerd en is de 'hogere legerleiding' (on)gerust voor een maand.

Veel interessanter is het om de bijpassende kansverdeling te schatten en eens echt een goede doorkijk te bieden op de te verwachten uitkomsten van het proces. Zo verkrijgen we een verwachting voor de volgende maand of de volgende procesprestatie.

#### Praktijkcase

Ter verduidelijking pakken we een praktijkcase:

Een meubelbedrijf levert, jawel, meubelen aan haar klanten. Als de verkoper in de showroom op het punt staat een mooie bank of kast aan een klant te slijten rijst de vraag: wanneer kan ik mijn meubel eigenlijk verwachten?

De verkoper heeft nu een aantal keuzes:

1. Praat er wat gemakkelijk over: "*mevrouw, het komt goed, echt...*"
2. Geef aan wat de gemiddelde levertijd is: "*meestal duurt het een week of 8.*"
3. Geef aan wat bij welke kans  $p$  er  $x$  weken of minder gewacht moet worden op het meubel.

Het lijkt duidelijk dat optie 3 de beste informatie biedt en de juiste verwachtingen wekt. Menig verkoper vindt het echter lastig dit zelf te berekenen.

Hoe kunnen we een goede inschatting geven over de verwachte leverweek en de kans dat deze gehaald wordt (of eerder)? Hoe benutten we historische data uit processen om daarmee een kansverdeling te schatten en die vervolgens te gebruiken om het proces richting klant te evalueren?

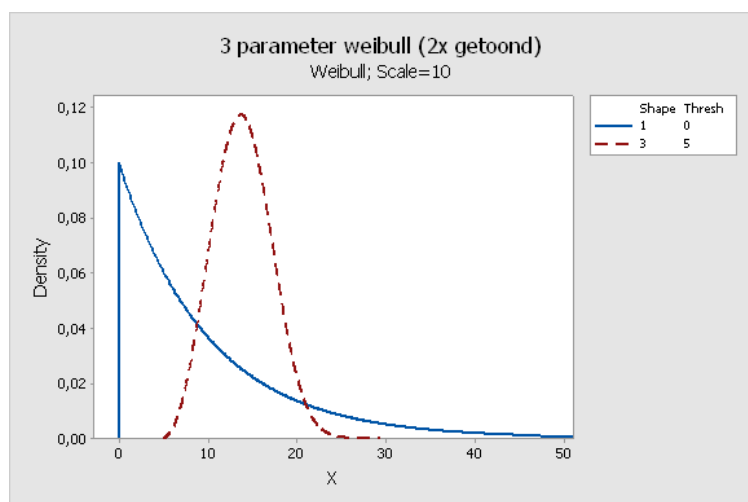
Het begint allemaal met een rijtje leveringsdata van de leverancier (zeg vanaf 50 datapunten lang). Hieruit moet dan een kansverdeling worden geschat waarmee we de kans en het aantal weken wachttijd op een meubel kunnen bepalen.

We realiseren ons dat de 50 leveringsweken in ons voorbeeld een steekproef zijn, welke (hopelijk) representatief zijn voor de levertermijnen van de leverancier. Het is daarom beter om een kansverdeling te schatten in plaats van simpelweg die leverweek te zoeken waarbij het percentage  $p$  gekoppeld kan worden aan  $x$  aantal leverweken of minder. Zeker indien het percentage hoog ligt ( $>80\%$ ) en de onderliggende dataset klein is ( $<50$ ).

Merk op: een al te kleine dataset ( $n < 30$ ) wordt een risico bij het bepalen van een passende kansverdeling en het schatten van haar parameters.

Van de verschillende kansverdelingen die in aanmerking komen (exponentieel, normaal, weibull, en nog een paar) belichten we de 3 parameter weibull verdeling en wel om de volgende reden:

- deze verdeling kent vele vormen, waarmee we de proceskarakteristiek goed kunnen volgen, kijk daarvoor eens naar afbeelding 1:



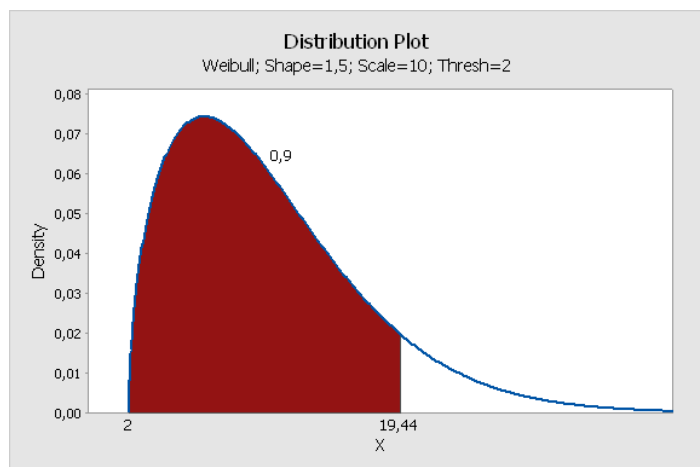
Afbeelding 1

- Ook kunnen we de kansverdeling mooi opschuiven naar links of rechts dmv de threshold. Zo kan een deel van een proces worden bekeken (bijvoorbeeld na 3 weken voorbewerking).

Om de zaken uit te rekenen moeten we weten:

- hoe de verdeling eruit ziet;
- welke kans  $p$  we hanteren
- voor welke event  $x$

Bijvoorbeeld: een 3 parameter weibull met kans 90% dat een levering binnen  $x$  weken plaatsvindt: zie afbeelding 2.



Afbeelding 2

Lees maar af: 19 en een halve week (19.44 op de x-as) of minder zal het wel duren voor 9 van de 10 leveringen... (0.9 rood oppervlak onder de blauwe lijn).

Laten we de 3 parameter weibull verdeling eens precies bekijken, en dan vooral de onbepaalde integraal van deze verdeling:

vergelijking (1):

$$F(x)_{(\lambda, \theta, k)} = 1 - e^{-\left(\frac{x-\theta}{\lambda}\right)^k}$$

waarbij:

$x$  het event is (het aantal weken of dagen doorlooptijd)  $x \in [0, \infty]$

$\theta$  de threshold parameter is (minimum vereiste (of geobserveerde) doorlooptijd in dezelfde grootheid als  $x$ ).  $\theta \in [\mathbb{R}]$

$\lambda$  de scale parameter is (63.2 percentielpunt in dataset, waarbij 36.8% van de datapunten groter zijn)  $\lambda \in [0, \infty]$

$k$  de shape parameter is (de vormfactor van de in (1) genoemde kansverdeling).  $k \in [0, \infty]$

Daar vergelijking (1) de kans  $p$  weergeeft op een event  $x$  (de kans  $p$  dat een meubel binnen  $x$  weken wordt geleverd) dient deze te worden omgeschreven in een vorm zodat bij een kans  $p$  het verwachte aantal leverweken wordt verkregen:

vergelijking (2):

$$p = 1 - e^{-\left(\frac{x-\theta}{\lambda}\right)^k} \Leftrightarrow$$

$$p = 1 - e^{-a} \vee a = \left(\frac{x-\theta}{\lambda}\right)^k \Leftrightarrow$$

$$1 - p = e^{-a} \Leftrightarrow$$

$$\ln(1 - p) = -a$$

$x$  oplossen uit  $a$ :

$$a = \left(\frac{x-\theta}{\lambda}\right)^k \Leftrightarrow$$

$$a^{1/k} = \left(\frac{x-\theta}{\lambda}\right) \Leftrightarrow$$

$$x - \theta = \lambda a^{1/k} \Leftrightarrow$$

$$x = \theta + \lambda a^{1/k}$$

door substitutie van  $a$ :

$$x = \theta + \lambda [ \ln(1 - p) ]^{1/k}$$

Nu hebben we de 3 parameter weibull in een voor ons doel bruikbare vorm; we kunnen immers bij een gegeven kans  $p$  en een goed beschreven threshold, scale en shape, het verwachte aantal leverweken  $x$  berekenen.

Vervolg vraag: "hoe kunnen we de parameters goed schatten?"

$\theta$  (threshold parameter) kan worden gegeven als het geobserveerde minimum van de dataset; of worden ingesteld als een minimum waarde welke een fysieke grens voorstelt (bijvoorbeeld bij 5 weken transport over zee).

$\lambda$  (scale parameter) kan worden gevonden door de dataset te ordenen van laag naar hoog en vervolgens dat punt (door interpolatie) te vinden dat het 63.2 percentiel voorstelt. (63,2% van de datapunten is kleiner of gelijk, 36,8% is groter) Deze volgt rechtstreeks uit de definitie van de kansverdeling (1).

$k$  (shape parameter) kan worden gevonden met behulp van de mediaan uit de dataset, immers:

vergelijking (3):

$$\text{median} = \lambda (\ln(2))^{1/k} \Leftrightarrow$$

$$\frac{\text{median}}{\lambda} = (\ln(2))^{1/k} \Leftrightarrow$$

$$\frac{\log(\frac{\text{median}}{\lambda})}{\log(\ln(2))} = \frac{1}{k} \Leftrightarrow$$

$$k = \frac{\log(\ln(2))}{\log(\frac{\text{median}}{\lambda})}; \{\theta = 0\}$$

Indien de threshold parameter  $\theta$  wordt geïntegreerd in vergelijking (3) resulteert

vergelijking (4):

$$k = \frac{\log(\ln(2))}{\log\left(\frac{(\text{median} - \theta)}{(\lambda - \theta)}\right)}$$

Waar waren we gebleven...

We hebben een kansverdeling (3 parameter weibull) met een threshold, scale en shape factor. We hebben daarnaast een kans  $p$  dat er binnen  $x$  weken geleverd kan worden.

Tijd voor een praktisch voorbeeld. Zie hiervoor de file welke u [hier](#) kunt downloaden. Een praktisch voorbeeld met 300 datapunten (leverweken; x-en) opgeslagen en van klein (kort) naar groot (lang) gesorteerd in de genoemde excel-file.

Uit de kolom data valt af te lezen:

$\theta$  (threshold) 3,334 (minimale datapunt uit de set)

$\lambda$  (scale) 21,88 (via datapunten nr 189 en 190)

median 18,325 (via datapunten 150 en 151)

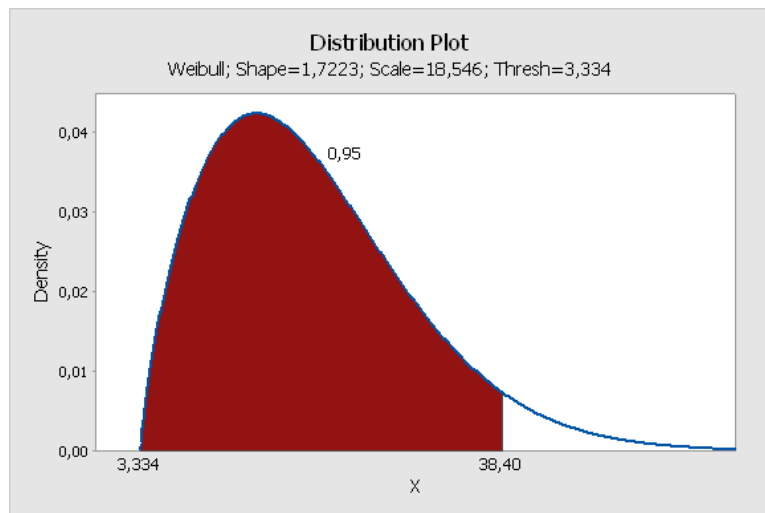
$$k = \frac{\log(\ln(2))}{\left(\frac{(18,325 - 3,334)}{(21,88 - 3,334)}\right)} = 1,7223$$

De kans  $p$  om binnen  $x$  weken een compleet meubel uit te leveren bij de klant:

$$x = 3,334 + (21,88 - 3,334) \cdot \left(\ln(1 - 0,95)\right)^{1/1,7223} = 38,4 \text{ weken}$$

Het 95% punt uit de data (datapunt 285) kent een waarde van 39 weken.

Minitab\_analyse komt tot dezelfde voorspelling: 38,4 weken: zie afbeelding 3.



Afbeelding 3

### Final check

Hoe komen we erachter of de door ons gekozen 3-parameter Weibull met de gevonden parameters past bij de data van ons proces? Daarvoor moeten we eens wat punten berekenen voor  $x$ , klopt dat dan met onze dataset? We nemen voor  $p$  de set  $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ . We gaan voor elk van deze  $p$ -waarden de bijpassende event  $x$  berekenen (aantal weken) en observeren of een en ander past.

Zoals uit onderstaande tabel (zie bijgevoegde excel-file 2e tabblad) is af te lezen:

klopt dit verhaal heel aardig...

$p$	verwacht aantal weken	geobserveerd aantal weken:	% afwijking schatting tov geobserveerd:
0,1	8,355096021	7,5371	10,9%
0,2	11,09693565	10,3793	6,9%
0,3	13,52674709	13,2909	1,8%
0,4	15,89047813	15,8012	0,6%
0,5	18,32499652	18,546	1,2%
0,6	20,96212278	21,3614	1,9%
0,7	23,99055173	23,3211	2,9%
0,8	27,78233154	27,8839	0,4%
0,9	33,43351825	34,8234	4,0%

De statistisch juiste manier om vast te stellen of de verdeling past is een Goodness of fit test te doen (bijvoorbeeld via de Anderson Darling test, of een Maximum Likelihood test) welke een  $p$ -waarde oplevert met de kans dat de verdeling past. De bovenstaande tabel geeft een goede indicatie, maar niet meer dan dat.

### Afdronk

De 3 parameter weibull verdeling biedt een prachtig perspectief op het modelleren van gebeurtenissen met een continue CTQ (tijd bijvoorbeeld). Het kan allemaal in Excel en met een klein beetje rekenwerk kunnen we de hogere legerleiding voorzien van verwachte prestaties en zelfs scenario's doorrekenen (wat is de kans dat?).

Deze methode is ook nog eens robuuster omdat de gehele dataset wordt meegenomen in het bepalen van de parameters, waarna uitspraken worden gedaan over events die wellicht niet precies voorhanden zijn in de dataset zelf. Een afdoende ( $n \geq 30$ ) en representatieve steekproef blijven natuurlijk essentiële randvoorwaarden, ook moet het proceskarakter redelijk overeenkomen met de gekozen kansverdeling.

### Meer weten

Heeft u vragen en/of wilt u hier meer over weten? Neem dan contact met op Melvin Harteveld. Bel naar 023 – 707 81 15, of stuur een mail naar [melvin.harteveld@projectsone.nl](mailto:melvin.harteveld@projectsone.nl)



### ProjectsOne. Geen rapporten, wel resultaten

ProjectsOne is opgericht door enthousiaste, ervaren professionals met uitgebreide kennis van verbeterprojecten. We geven trainingen, doen coachingtrajecten en begeleiden implementaties. We kiezen altijd voor een openen directe benadering. Zonder dikke rapporten maar met meetbare uitkomsten in euro's, klant- en personeelstevredenheid. Zo bieden we u precies de ondersteuning die u kunt gebruiken om uw organisatie naar een hoger niveau te tillen.

ProjectsOne  
De corridor 12L  
3621 BZ Breukelen

www.projectsone.nl  
info@projectsone.nl  
t +31(0)23 707 8115

