# Labo 04 – Continue kansverdeling

## Doelstellingen

Reken met kansen en kansverdelingen wordt toegepast in alle domeinen van ons leven, gaande van het uitrekenen van risico’s, het voorspellen van het weer, inschatten van de impact van een reclamecampagne op de omzetcijfers van een bedrijf, modelleren van de verspreiding van een (computer) virus, etc.

Bij deze opdracht komt de normaalverdeling aan bod.

De normaalverdeling is de meest gekende en gebruikte kansverdeling in de statistiek. Het is een continue klokvormige kansverdeling die symmetrisch verloopt rond het gemiddelde. Ze is hoog in het midden, en wordt naar lage en hoge waarden steeds kleiner zonder ooit echt nul te worden.

De bedoeling van deze opdracht is om vlot overweg te kunnen met deze kansverdeling bij het uitrekenen van kansen en het kunnen bepalen van de verwachte waarden van kansverdelingen.

## Inhoud

* Kunnen nagaan of een reeks gegevens al dan niet normaal verdeeld zijn.
* Rekenen met en interpreteren van de normaalverdeling.

## Algemene toelichting

### Indienen en uitwerking van de opdracht

Bestanden die ingediend moeten worden bij deze opdracht:   
**Labo\_04\_Continue\_Kansverdelingen\_Opdracht.xlsx  
Labo\_04\_Continue\_Kansverdelingen\_Opdracht.ipynb**

**Zorg dat je naam duidelijk zichtbaar in de bestandsnaam vermeld wordt!**

De uitwerkingen en de antwoorden vul je in bovenstaande bestanden aan.  
  
Extra uitleg:

Video normal distribution function excel: <https://www.youtube.com/watch?v=B1PY6hTtpmc>

(Redelijk goede uitleg, enkel ze spreekt over het gebruik van de PMF, wat we nooit zullen gebruiken bij de normaalverdeling (zie theorie))

Documentatie: <https://support.microsoft.com/en-us/office/norm-dist-function-edb1cc14-a21c-4e53-839d-8082074c9f8d>

## 

## Opdracht DEEL 1: Excel

#### Vraag 1 – Bepaal onderstaande kansen met behulp van Excel

|  |  |
| --- | --- |
| **Normaalverdeling** | **Via Excel** |
| P(*Z* < -1,67) |  |
| P(*Z* > 2,56 ) |  |
| P(1,34 < *Z* < 2,34) |  |
| *X~N(50;5)* P(*X<60*) |  |
| *X~N(5;1)* P(*X* > 2,5) |  |
| *X~N(120;6)* P(114 < *X* < 126) |  |

#### Vraag 2 – Spanningsomvormer

Een omvormer voor zonnepanelen zorgt ervoor dat de opgewekte gelijkstroom door de zonnepanelen wordt omgezet in bruikbare wisselstroom voor ons huishouden.

De MTTF (Mean Time To Failure = gemiddelde levensduur) van een bepaald type omvormer bedraagt 10 jaar met een standaardafwijking van 3 jaar. De garantieperiode op deze omvormer is 5 jaar.

1. Wat is de kans dat een omvormer binnen de garantieperiode stuk gaat?
2. Stel dat de zonnepanelen zelf na 20 jaar worden vernieuwd. Wat is de kans dat gedurende deze periode de omvormer nooit vervangen dient te worden?
3. Wat is de kans dat de omvormer stuk gaat tussen de 5 jaar en de 20 jaar en je dus niet op de garantie kunt terugvallen?
4. Na hoeveel jaar moet je de omvormer vervangen om ervoor te zorgen dat je minder dan 10% kans hebt dat je met een kapotte omvormer komt te zitten?

#### Vraag 3 – Riemen

Een kledingfabrikant wil een nieuwe riem op de markt brengen. Er is een variant voor mannen en een variant voor vrouwen. Elke variant komt in smal, medium en large.

Elk soort riem is qua lengte als volgt verstelbaar:

|  |  |  |
| --- | --- | --- |
|  | **Mannen** | **Vrouwen** |
| **Smal** | Van 80 cm tot 92 cm | Van 78 cm tot 90 cm |
| **Medium** | Van 90 cm tot 105 cm | Van 85 cm tot 100 cm |
| **Large** | Van 102 cm tot 115 cm | Van 95 cm tot 112 cm |

Om een inschatting te kunnen maken van de productieaantallen laat de fabrikant een onderzoek uitvoeren naar de heupomtrek bij mannen en vrouwen. De gegevens kan je terugvinden onder het Excel werkblad “riemen”.

1. Ga na of deze data normaal verdeeld is. Doe dit door de kansverdeling te plotten en door na te gaan hoeveel procent van de gegevens binnen de 1, 2 en 3 standaardafwijkingen is gelegen. Vergelijk met de werkelijke normaalverdeling die hoort bij het berekende gemiddelde en standaardafwijking. Doe de analyse apart voor mannen en vrouwen.
2. Vergelijk de maten van de mannen en de vrouwen in termen van gemiddelde en spreiding.
3. **In de veronderstelling dat de data normaal verdeeld is,** bereken telkens hoeveel procent van de mannen maat smal, medium en large kan dragen. Doe dit ook voor vrouwen. Vul daarbij onderstaande tabel aan.

|  |  |  |
| --- | --- | --- |
|  | Mannen | Vrouwen |
| Smal | % | % |
| Medium | % | % |
| Large | % | % |

1. Bij hoeveel procent van de mannen past zowel smal als medium?
2. Bij hoeveel procent van de vrouwen past zowel medium als large?
3. Bij hoeveel procent van de mannen past geen enkele maat?
4. Bij hoeveel procent van de vrouwen past geen enkele maat?

Ligt deze data in lijn met de bevindingen van bovenvermelde Amerikaanse studie? Met andere woorden volgt de data ongeveer dezelfde kansverdeling?

#### Vraag 4 – Hematocriet

De hematocriet (afkorting: HT of HCT) of hematocrietwaarde is het volume van het bloed dat door de rode bloedcellen (erytrocyten) wordt ingenomen, weergegeven als een fractie van het totale bloedvolume. De hematocriet wordt bepaald door een buisje bloed te centrifugeren. Onderin verzamelen zich dan de rode bloedcellen, daarbovenop de witte bloedcellen (leukocyten) en bovenin de buis het plasma.

De hematocriet wordt op kunstmatige wijze verhoogd door injectie van erytropoëtine (epo). Bij het wielrennen en enkele andere sporten wordt een startverbod opgelegd met een te hoge hematocriet. De grens is daarbij 0,50, maar voor sporters die van nature een hoge hematocriet hebben geldt een uitzondering. Dit startverbod is formeel niet bedoeld als schorsing, maar voor de eigen gezondheid van de sport(st)er. Een te hoge hematocriet leidt namelijk tot dik en stroperig bloed, waardoor verstoppingen in haarvaten (met alle gevolgen van dien) kunnen ontstaan.

Bij wielrenners is de HCT waarde normaal verdeeld met een gemiddelde van 44,5 en een standaardafwijking van 2,9.

1. Wat is de kans dat een wielrenner van nature een HCT heeft die lager ligt dan 40?
2. Wat is de kans dat een wielrenner van nature een HCT heeft tussen de 40 en de 50?
3. Wat is de kans dat een wielrenner onterecht een startverbod opgelegd krijgt? (van nature een HCT groter dan 50)
4. Bepaal het HCT gehalte waarvoor geldt dat maar 2% van de wielrenners een hoger HCT gehalte hebben?
5. In het tabblad van hematocriet vind je 120 gemeten hematocrietwaarden bij wielrenners. Liggen deze waarden in lijn met de bovenstaande veronderstellingen? Met andere woorden volgt de data dezelfde kansverdeling? Controleer aan de hand van Excel door de kansverdeling te plotten en er de theoretische verdeling bij te tekenen. Ga ook na of ongeveer 68% van de HCT waarden binnen 1 standaardafwijking van het gemiddelde liggen en ongeveer 95 procent van de HCT waarden binnen de 2 standaardafwijkingen van het gemiddelde.

## Opdracht DEEL 2: Python

De dataset en de opgave zijn te vinden in ***Labo\_04\_Continue\_Kansverdelingen\_Opdracht.ipynb***.

Beantwoord de vragen in het document.