

DE CENTRALE LIMIETSTELLING

VOOR EEN BREED PUBLIEK

LETO CAVÉ

Abstract

Hoe lang wacht je gemiddeld voordat je aan de beurt bent bij de koffieautomaat? Kan een handvol metingen daar al iets over zeggen? In dit artikel laten we zien hoe de centrale limietstelling precies die vraag beantwoordt – een stelling waarop vrijwel elke peiling, elk medisch onderzoek en elke kwaliteitscheck berust.

Doelgroep: Volwassenen, scholieren en studenten met interesse in wetenschap, geen wiskundige voorkennis vereist.

SLEUTELWOORDEN: Centrale limietstelling, Verdeling, Steekproefgemiddelde, Exponentiële verdeling, Normaalverdeling, Wachtijd

De centrale limietstelling

Je bent druk bezig en krijgt trek in een lekkere kop koffie. De ervaring bij de koffieautomaat is elke dag anders. Soms kun je direct doorlopen. Een andere dag staat er net een groep mensen voor je die het apparaat niet begrijpt, en sluit je achteraan aan in de rij.

Na een week frustratie vraag je je af: hoe lang sta ik hier nu gemiddeld heel het jaar te wachten? Je wilt hier niet dagenlang met een stopwatch gaan bivakkeren. Daarom kies je voor een slimme statistische aanpak: op een willekeurige dag meet je de wachttijd van 30 opeenvolgende personen. De resultaten van deze steekproef zijn te vinden in Tabel 1.

Tabel 1

Wachttijden bij de koffieautomaat

Observatie	1	2	3	4	5	...	29	30
Wachttijd (min)	0.5740	1.0391	0.6735	0.0542	0.9712	...	1.1577	1.7600

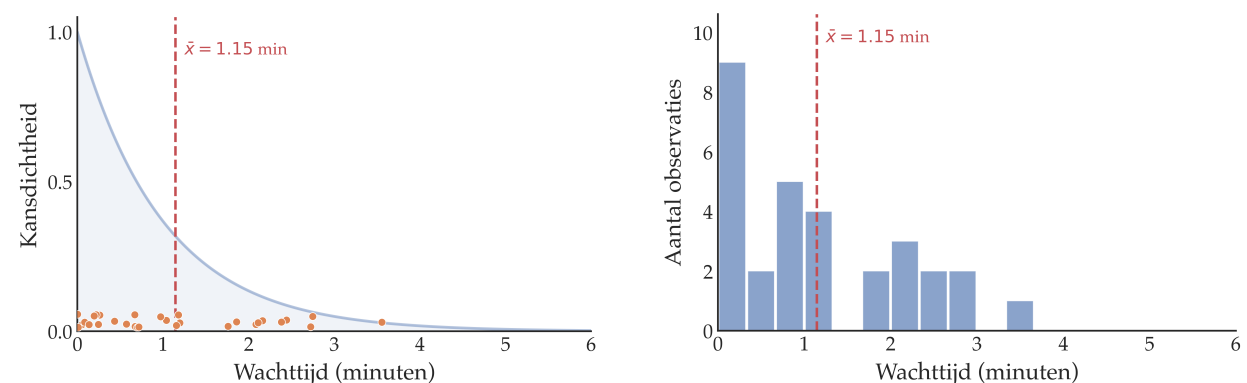
30 willekeurige metingen uit een exponentiële verdeling met $\lambda = 1$ (gemiddelde wachttijd van 1 minuut). Het gemiddelde van deze steekproef is $\bar{x} = \frac{1}{30} \sum_{i=1}^{30} x_i = 1.15$ minuten, en komt in de buurt van de werkelijke gemiddelde wachttijd $\mu = 1$ minuut.

We kunnen nu het gemiddelde van deze wachttijden berekenen. Dat doen we door alle metingen op te tellen en te delen door 30. De rekenmachine laat zien dat ons steekproefgemiddelde uitkomt op 1.15 minuten.

Ook merken we in ons notitieboekje op dat de individuele wachttijden verre van gelijkmatig zijn verdeeld. De meeste mensen hebben geluk en hoeven nauwelijks te wachten, terwijl een ongelukkige enkeling heel lang achteraan de rij aansluit. Als je al die losse wachttijden in een grafiek zet, zie je een scheve helling die heel steil begint en langzaam naar rechts uitdooft (zie Figuur 1). Statistici noemen dit een exponentiële verdeling.

Figuur 1

Steekproef van $n = 30$ wachttijden bij de koffieautomaat.



(a) Individuele wachttijden ($n = 30$) uitgezet tegen de daadwerkelijke verdeling.

(b) Verdeling van 30 observaties over intervallen van 30 seconden.

Note. Bij $n = 30$ observaties ligt $\bar{x}$ al aanzienlijk dicht bij $\mu = 1$ minuut.

Maar hoe betrouwbaar is één meting? Als we dit experiment morgen rond een willekeurig tijdstip herhalen meten we 30 andere personen en vinden we ongetwijfeld een net iets ander gemiddelde.

Nu volgt een uiterst belangrijk resultaat met een zeer verrassende wending. Stel je voor dat we het experiment 100 000 keer herhalen, maar dat we telkens de grootte van de groep variëren: eerst één persoon, dan

twee, dan vijf, en uiteindelijk dertig. Elke keer berekenen we het groepsgemiddelde en al die honderdduizenden gemiddelden houden we bij. Ons papieren notitieboekje hebben we inmiddels maar thuisgelaten; we zijn overgestapt op Excel om deze gigantische hoeveelheid data soepel te ordenen. Zetten we de resultaten naast elkaar, dan tekent zich het volgende patroon af:

Tabel 2

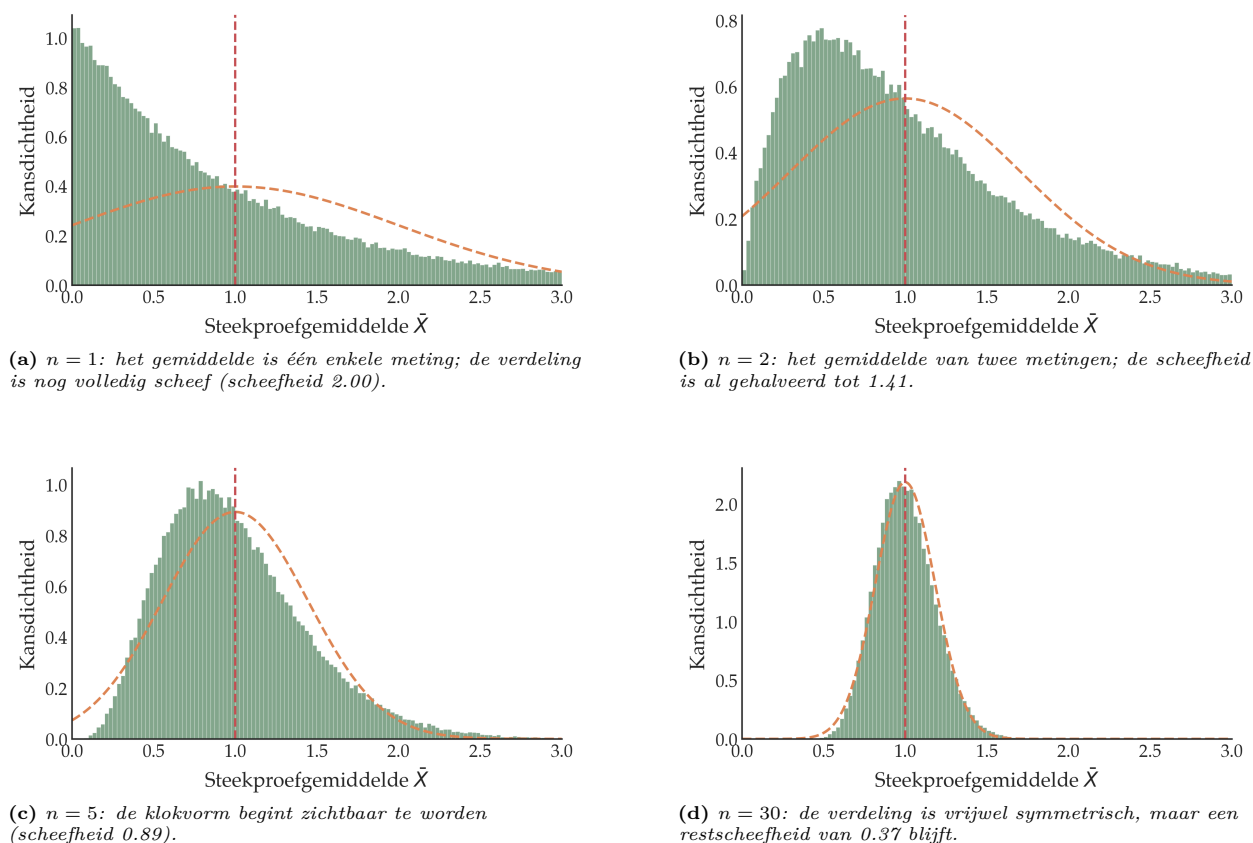
Naarmate de groep groter wordt, wordt de verdeling van het gemiddelde zowel smaller als symmetrischer.

Groepsgrootte n	Spreading $1/\sqrt{n}$	Scheefheid $2/\sqrt{n}$
1	1.000	2.000
2	0.707	1.414
5	0.447	0.894
30	0.183	0.365

Wanneer we al deze berekende groepsgemiddelden in een staafgrafiek zetten, gebeurt er iets opmerkelijks. Je zou misschien verwachten dat de grafiek van al die groepsgemiddelden er net zo scheef uit zal zien als de losse wachttijden uit Figuur 1. De werkelijkheid is echter wonderbaarlijk: naarmate we het gemiddelde over grotere groepen nemen, verdwijnt de scheve vorm en zien we een vrijwel symmetrische klokform, ook wel de normaalverdeling genoemd.

Figuur 2

De verdeling van het steekproefgemiddelde bij toenemende groepsgrootte n , telkens gebaseerd op $M = 100\,000$ herhalingen. De stippellijn is de normale verdeling $\mathcal{N}(1, 1/n)$.



In Figuur 2 wordt direct duidelijk dat het gemiddelde bij grotere groepen veruit het vaakst heel dicht

rond de 1 minuut ligt. Bij een groep van dertig personen is een gemiddelde van 3 minuten of van slechts 0.5 minuut een extreme uitzondering geworden; meet je maar één persoon, dan komt zo iets nog regelmatig voor.

Dit is de essentie van de centrale limietstelling. Het maakt niet uit hoe verdeeld, scheef of onvoorspelbaar de oorspronkelijke, individuele wachttijden op een dag zijn; zodra je het gemiddelde neemt van een groepje mensen, nadert de verdeling van de steekproefgemiddelden een symmetrische klokvorm. Meer herhalingen simuleren (groter M) maakt het beeld alleen scherper: het laat ons nauwkeuriger zien welke verdeling er is, maar verandert die verdeling niet. Grotere groepen nemen (groter n) is wat de verdeling zelf verandert: de klokvorm wordt zowel smaller als symmetrischer rond het werkelijke gemiddelde van 1 minuut.

Omdat we nu weten dat groepsgemiddelden zich in deze stabiele klokvorm organiseren, kunnen we met één enkele steekproef van 30 mensen al met een redelijk grote precisie voorspellen waar het werkelijke gemiddelde ligt. Onze eerste schatting van een wachttijd tot een kopje koffie van 1.15 minuten was dus zo gek nog niet.

Een onmisbare stelling in de wetenschap

Dit koffie-experiment stelt bloot waarom de centrale limietstelling bij vrijwel elk onderzoek gebruikt wordt. Of het nu gaat om een politieke peiling onder duizend stemmers, de kwaliteitscontrole van miljoenen microchips in een fabriek, of het testen van een nieuwe medicijn bij een selecte groep proefpersonen: de werkelijkheid is vaak veel te groot of te complex om in haar geheel te meten.

Dankzij de centrale limietstelling hoeft dat gelukkig ook niet. We kunnen erop vertrouwen dat de gemiddelden van onze steekproeven zich volgens de voorspelbare wetten van de klokvorm gedragen. Door de prettige wiskundige eigenschappen van deze klokvorm kunnen wetenschappers met grote precisie de betrouwbaarheid van hun resultaten bepalen. Merk wel op dat bij $n = 30$ een lichte scheefheid overblijft — de vuistregel “ $n \geq 30$ ” is dan ook een vuistregel en geen stelling. Daar schuilt echter nog wel een addertje onder het gras, maar dat avontuur is voor een volgende dag.

Bronnen

- [1] Joseph C. Watkins. *An Introduction to the Science of Statistics: From Theory to Implementation*. Topic 11: The Central Limit Theorem. University of Arizona. URL: https://archive.math.arizona.edu/jwatkins/K_clt.pdf (visited on 09/04/2026).
- [2] Remco van der Hofstad. *Handout limietstellingen Kansrekening 2WS20*. Jan. 11, 2018. URL: https://rhofstad.win.tue.nl/hand_out_limit_theorems.pdf.
- [3] Shaun Turney. *Centrale Limietstelling (Central Limit Theorem) | Formule*. Dutch. Scribbr. Aug. 16, 2022. URL: <https://www.scribbr.nl/statistiek/centrale-limietstelling/>.
- [4] Jan van de Craats. *Kansrekening en Statistiek*. Dutch. Collegedictaat. Koninklijke Militaire Academie, Aug. 2002. URL: https://staff.fnwi.uva.nl/j.vandecraats/Kans_Stat.pdf.