

Taalonderzoek

In 2013 was er een onderzoek naar de woordenschat van mensen in Nederland en Vlaanderen. Iedereen kon meedoen met het onderzoek door een test te doen op internet. Bij deze test kreeg een deelnemer 100 willekeurig gekozen woorden te zien uit een lijst van 50 000 bestaande Nederlandse woorden en 20 000 door de onderzoekers verzonnen 'nepwoorden'. Van elk woord moest worden aangegeven of het een bestaand woord is of niet. Het aantal nepwoorden in een test is (bij benadering) binomiaal verdeeld.

Marieke heeft de test gedaan. In haar test zaten 37 nepwoorden.

- 4p **6** Bereken de kans dat in een test van 100 woorden 37 of meer nepwoorden voorkomen.

Na afloop van de test wordt een score toegekend. Hiervoor wordt berekend:

- het percentage bestaande woorden dat de deelnemer (terecht) als bestaand aanmerkt; dit percentage, afgerond op gehele, noemt men A;
- het percentage nepwoorden dat de deelnemer (ten onrechte) als bestaand aanmerkt; dit percentage, afgerond op gehele, noemt men B.

Vervolgens geldt: $\text{score} = A - B$.

Bij haar test van totaal 100 woorden heeft Marieke van de bestaande woorden in de test er 56 herkend. Van de 37 nepwoorden heeft ze er 5 ten onrechte als bestaand bestempeld.

- 3p **7** Laat met een berekening zien dat Marieke een score van 75 had voor de test.

Eind oktober 2013 was de test 572 146 keer gemaakt door 368 798 verschillende deelnemers. Er waren dus (flink wat) deelnemers die de test meer dan eens gedaan hadden. Uit onderzoek bleek dat de deelnemers in drie groepen verdeeld konden worden:

- de proevers: deze deelnemers maakten de test één keer;
- de ambitieuzen: deze deelnemers maakten de test 2–10 keer;
- de doorzetters: deze deelnemers maakten de test meer dan 10 keer.

De verdeling van deze groepen over het totaal aantal deelnemers was: proevers 76%, ambitieuzen 21% en doorzetters 3%.

Uit deze gegevens blijkt dat het gemiddeld aantal keren dat de ambitieuzen de test deden, minder dan 3 was.

- 4p **8** Bereken hoe groot dit gemiddeld aantal keren ten hoogste kan zijn. Rond je antwoord af op één decimaal.

Voor een vervolgonderzoek kiest men willekeurig 15 van deze 368 798 deelnemers.

- 3p **9** Bereken de kans dat deze groep 2, 3 of 4 ambitieuzen bevat.