

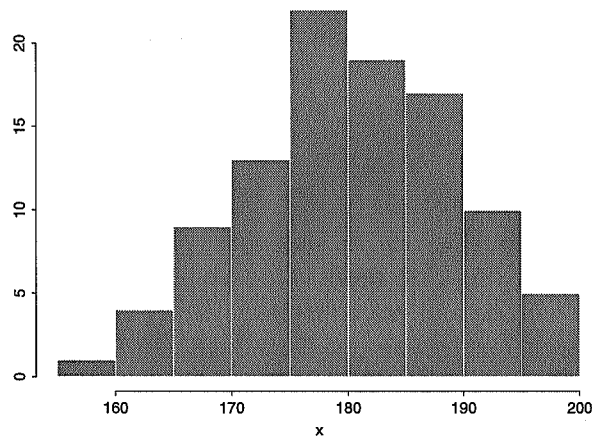
Tentamen Toegepaste Statistiek <i>voor studenten</i> Informatiekunde en Kunstmatige Intelligentie <i>vrije Universiteit, Faculteit Exacte Wetenschappen</i> 26 maart 2007, 8.45-10.45 uur

1.
 - a) Leg eventueel aan de hand van een voorbeeld uit wat 'stratified sampling' (gestratificeerd steekproeftrekken) is.
 - b) De directie van een omroep wil weten wat 'de Nederlander' van hun programma aanbod vindt. Om hier achter te komen, wordt een web-site geopend waar mensen hun mening over de programma's kunnen geven. Aan het bestaan van de website wordt in radioprogramma's van de betreffende omroep ruimschoots aandacht geschonken. Geef aan welke manier van steekproefnemen we hier te maken hebben en wat een gevaar is bij de gebezigde methode.
 - c) Leg uit wat causaliteit en correlatie zijn.
 - d) Gedurende de voorbije twintig jaar is veel onderzoek gedaan naar het zogenaamde 'Mozart effect'. Veel studies zijn verschenen in verschillende wetenschappelijke tijdschriften. Op basis van deze artikelen, wordt een nieuwe studie gedaan naar het bestaan van zo'n effect. Hoe wordt dit type studie genoemd?
2.
 - a) In maart 2007 is voor het eerst in Vlaanderen een snelheidslimiet van 70 km/h ingesteld vanwege smogvorming. Het smogniveau is bepaald op basis van metingen van het gehalte aan schadelijke stoffen in de lucht, uitgedrukt in p.p.m. (parts per million). Welk type data vormen deze metingen en wat is de meetschaal?
 - b) Een softwarefabrikant houdt een enquête onder haar klanten. Een aantal suggesties voor veranderingen worden gedaan en de klanten wordt gevraagd hoe wenselijk ze die veranderingen achten. Men geeft dit aan op een schaal van A (zeer wenselijk) t/m F (onwenselijk). Welk type data krijgt de softwarefabrikant terug en op welke schaal zijn die gemeten?
3. Bij een zeskamp is een onderdeel om van afstand een bal in een doel te schieten. Het aantal pogingen dat men nodig heeft om de eerste goal te scoren is voor de puntentelling van belang. Voor 25 deelnemers staat het aantal benodigde pogingen in de onderstaande tabel.

1	2	1	1	2
1	1	1	1	1
1	2	1	1	3
2	1	2	1	4
3	4	1	1	1

- a) Geef een frequentietabel van deze data.
- b) Teken een staafdiagram voor deze data en beschrijf de dataverdeling kwalitatief.

- c) Geef de modus en de mediaan van de data.
- d) Hieronder is een histogram weergegeven van een dataset bestaande uit lengten van honderd personen in een steekproef. Bereken op basis van dit histogram een benadering voor het percentage mensen in de populatie die langer zijn dan 1.90m.
- e) Geef aan wat de hoogte is van het *herschaalde* histogram van de data waarop het histogram is gebaseerd, en wel in cel 190-195.



Figuur 1: *Histogram van lengten*

4. Een fabrikant van AA drink doet onderzoek naar de preciese inhoud van de flesjes die 100ml zouden moeten bevatten. Van 25 aselekt gekozen flesjes uit de dagproductie wordt de inhoud (in ml) gemeten. Hieronder staan de resultaten van de metingen.

101.71	101.21	101.70	99.90	100.14
100.42	101.26	100.75	101.16	101.43
101.32	99.99	100.61	100.11	101.34
101.87	101.87	100.97	101.56	99.76
100.87	100.72	100.60	100.61	100.84

- a) Bereken op basis van relatieve frequenties een schatting voor de kans dat een willekeurig flesje uit de dagproductie minder dan 100.00 ml bevat.

Uit eerder onderzoek is gebleken dat de populatieverdeling van de inhoud van een flesje bij benadering normaal is met populatiegemiddelde $\mu = 101.00$ en standaardafwijking 0.50.

- b) Geef op basis van deze bewering de kans dat een willekeurig flesje minder dan 100 ml bevat. Tabel 1 aan het eind van dit tentamen kan bij het berekenen van deze kans van nut zijn.
 - c) Geef op basis van deze bewering de kans dat de gemiddelde inhoud van 25 flesjes uit de dagproductie minder dan 100.90 gram is. Tabel 1 aan het eind van dit tentamen kan ook bij het berekenen van deze kans van nut zijn.
5. We hebben de beschikking over een kaartspel bestaande uit 52 kaarten. Van de kleuren *Harten*, *Klaver*, *Schoppen* en *Ruiten* hebben we de 'nummers'

$$2, 3, \dots, 10, V, H, B, A.$$

We trekken willekeurig een kaart uit het kaartspel.

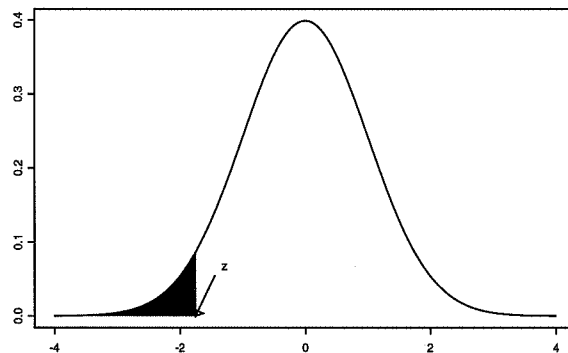
- a) Construeer een kansruimte voor dit experiment.
 - b) Bereken de kans dat een kaart met nummer 3 wordt getrokken. Doe dit door netjes de gebeurtenis te formuleren binnen de in a) geconstrueerde kansruimte.
 - c) Ga na of de gebeurtenissen 'nummer 3 wordt getrokken' en 'kleur *Harten* wordt getrokken' stochastisch onafhankelijk zijn. Doe dit door te controleren of deze twee gebeurtenissen aan de definitie van stochastische onafhankelijkheid voldoen.
6. Onder 2500 aselekt gekozen leden van voetbalverenigingen is een enquête gehouden. Een van de vragen was of Marco van Basten als bondscoach zou moeten opstappen. Van de 2500 mensen antwoordden er 1600 met nee.

- a) Geef de gebruikelijke *puntschatting* $\hat{p}$ voor het onbekende percentage p van alle leden van voetbalverenigingen die vinden dat Van Basten mag blijven.

In de onderdelen b) en c) van deze opgave mogen breuken en wortels in het antwoord blijven staan.

- b) Geef het 95% betrouwbaarheidsinterval voor p in de huidige situatie en licht de interpretatie van dat interval toe.
 - c) Als vooraf was geëist dat het 95% betrouwbaarheidsinterval smaller dan 0.02 had moeten zijn, hoe groot had men dan de steekproef uit de populatie van leden van voetbalverenigingen minimaal moeten kiezen?
7. In opgave 4 is voor 25 aselekt gekozen flesjes AA drink uit de dagproductie van een fabrikant de preciese inhoud gemeten. Het gemiddelde van die metingen is $\bar{x} = 100.87$ en de standaardafwijking 0.65. In deze opgave willen we op basis van die metingen toetsen of de gemiddelde inhoud van alle flesjes in de dagproductie 101.00 ml kan zijn, zoals uit eerder onderzoek zou zijn gebleken.

- Stel de nulhypothese H_0 en de alternatieve hypothese H_a op met betrekking tot de gemiddelde inhoud van de flesjes uit de dagproductie.
- Bereken de standaardscore z (toetsingsgrootheid) gebaseerd op de data.
- Bereken de P -waarde behorend bij deze score (tabel 1 kan hierbij van nut zijn).
- Wordt H_0 ten gunste van H_a verworpen bij een significantieniveau $\alpha = 0.05$? Licht uw antwoord toe.



Figuur 2: standaard normale curve corresponderend met tabel 1

z	Oppervlakte onder de curve links van z
-2.932	0.002
-2.000	0.023
-1.960	0.025
-1.000	0.159
-0.200	0.421
0.253	0.600
1.188	0.883
1.645	0.950
1.875	0.970
2.500	0.994

Tabel 1: oppervlakten onder de standaard normale curve corresponderend met figuur 2

1a: 3	1d: 2	3a: 3	3d: 3	4b: 3	5b: 3	6b: 3	7b: 3
1b: 4	2a: 3	3b: 3	3e: 3	4c: 4	5c: 3	6c: 3	7c: 3
1c: 3	2b: 3	3c: 3	4a: 3	5a: 3	6a: 3	7a: 3	7d: 2

Tabel 2: te verdienen punten per onderdeel

Rekenmachine is niet nodig en niet toegestaan tijdens dit tentamen

Cijfer=(aantal punten + 8)/8

Succes!