

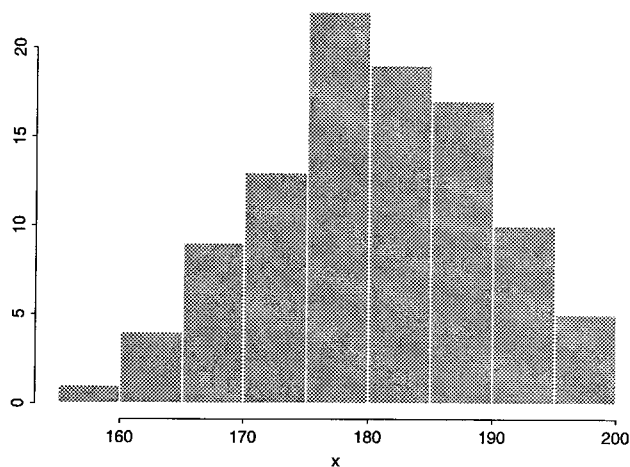
Tentamen Toegepaste Statistiek
voor studenten
Informatiekunde en Kunstmatige Intelligentie
vrije Universiteit, Faculteit Exacte Wetenschappen
26 maart 2004, 9.30-12.30 uur

1.
 - a) In *Metro* van 16 maart werd het volgende onderzoeksresultaat gemeld: *Tai Chi goed voor balans en bloedvaten*. Amerikaanse onderzoekers kwamen tot deze conclusie in *The Archives of International Medicine* door het bestuderen van 47 studies die over deze zaken zijn verschenen. Geef aan met welk type studie we hier te maken hebben en licht uw antwoord toe.
 - b) Om na te gaan hoeveel uur een Amerikaan gemiddeld per week sport, is een onderzoek gedaan. Naar verhouding van het aantal inwoners van de betreffende staat, is uit iedere staat een groep inwoners aselekt gekozen. Deze personen is gevraagd hoeveel uur ze gemiddeld per week sporten. Hoe heet het type steekproeftrekken waar we hier mee te maken hebben? Beschrijf dat type ook in het algemeen.
 - c) Er wordt beweerd dat kinderen die zich bezighouden met het schaakspel, meer talent hebben voor wiskunde dan kinderen die niet schaken. Op een bepaalde middelbare school wordt naar aanleiding van deze bewering in het eerste jaar verplicht schaakles gegeven. Welke (in de praktijk veel gemaakte) fout ligt hier op de loer?
 - d) In een medisch onderzoek wil men het effect van een nieuw soort zalf op een vorm van eczeem nagaan. Hiertoe worden 50 patienten met dit type aandoening geselecteerd. De helft krijgt de nieuwe zalf en de andere helft krijgt een zalf waarin de (vermoedelijk) werkende stof niet aanwezig is. Men voert een dubbelblinde studie uit, waarbij de patient iedere week in de kliniek komt om zich te laten onderzoeken. Leg uit wat zo'n dubbelblinde studie is en geef ook aan waarom het verstandig is om de studie dubbelblind uit te voeren. Gebruik in de beantwoording van deze vraag de geleerde terminologie.
2.
 - a) Wellicht is er bij dit tentamen een evaluatieformulier voor het vak bijgevoegd. Bij de meeste vragen die van toepassing zijn, wordt het antwoord verwacht op een schaal van --, -, +-, +, ++. Door het invullen van de lijst voorzien jullie het onderwijsbureau van data. Welk type data zijn dat en op welke meetschaal worden deze gegeven?
 - b) In Hilversum werd op vrijdag 19 maart een gifwolk gesignaleerd. Meteen trad een meetsysteem in werking, waarin in bepaalde omtrek rond de vuilnisbelt het gehalte van een giftige stof in de lucht werd gemeten, in parts per million (p.p.m.). Welk type data wordt hier verkregen? En op welke meetschaal?
 - c) Leg uit wat de volgende bewering betekent: 'De consumentenprijsindex in 2003, bij referentiejaar 1990, is 134.'

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

- Geef een frequentietabel van deze data.
- Teken een staafdiagram voor deze data *en* beschrijf de dataverdeling kwalitatief.
- Geef de modus en de mediaan van de data.
- 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 1.92m.
- 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 bierfabrikant doet onderzoek naar de preciese hoeveelheid van het kostelijke vocht die in de blikjes zit uit de productie van een bepaalde dag. Van 25 aselekt gekozen blikjes uit de dagproductie wordt de inhoud (in *cl.*) gemeten. Hieronder

staan de resultaten van de metingen.

33.25	33.16	33.42	33.16	33.34
33.31	33.30	33.21	33.52	33.17
33.05	32.91	33.50	33.23	33.53
33.35	33.20	33.09	33.18	33.17
33.27	33.28	33.47	33.32	33.04

- a) Bereken op basis van relatieve frequenties een schatting voor de kans dat een willekeurig blikje uit de dagproductie minder dan 33.00 *cl.* bier bevat.

Uit eerder onderzoek is gebleken dat de populatieverdeling van de inhoud van een bierblikje bij benadering normaal is met populatiegemiddelde $\mu = 33.20$ en standaardafwijking 0.20.

- b) Geef op basis van deze bewering de kans dat een willekeurig blikje minder dan 33.00 *cl.* bier 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 blikjes uit de dagproductie lager dan 33.10 *cl.* ligt. 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 5 wordt getrokken. Doe dit door netjes de gebeurtenis te formuleren binnen de in a) geconstrueerde kansruimte.
- c) Ga na of de gebeurtenissen ‘nummer 5 wordt getrokken’ en ‘kleur *Klaver* wordt getrokken’ stochastisch onafhankelijk zijn. Doe dit door te controleren of deze twee gebeurtenissen aan de definitie van stochastische onafhankelijkheid voldoen.
6. Binnen een groot bedrijf wil men nagaan welk percentage van de werknemers behoefte heeft aan een faciliteit om in de middagpauze te kunnen sporten. Hiertoe wordt aan 100 aselekt gekozen werknemers van het bedrijf de vraag voorgelegd of men behoefte heeft aan zo’n faciliteit. Van de 100 werknemers gaven er 36 aan behoefte te hebben aan zo’n faciliteit.
- a) Geef de gebruikelijke *puntschatting* $\hat{p}$ voor het onbekende percentage p van geïnteresseerde werknemers.

De Centrale Limietstelling is een stelling die in de huidige situatie het volgende zegt. Bij een grote steekproefomvang n , is de grootte

$$\sqrt{n} \frac{\hat{p} - p}{\sqrt{\hat{p}(1 - \hat{p})}}$$

bij benadering standaard normaal verdeeld. 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) Hoe groot moet de steekproef uit de werknemers bij benadering zijn om een 95% betrouwbaarheidsinterval ter breedte 0.02 te krijgen?
7. In de onderstaande tabel zijn de gewichten gegeven van ‘quarters’ voor een steekproef van omvang 25.

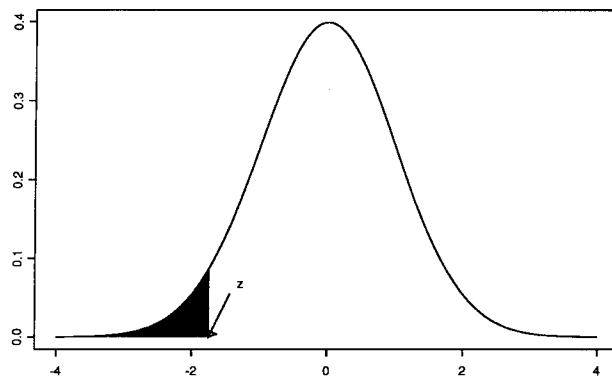
5.60 5.63 5.58 5.56 5.66 5.58 5.57 5.59
 5.67 5.61 5.84 5.73 5.53 5.58 5.52 5.62
 5.57 5.71 5.59 5.53 5.63 5.68 5.62 5.65

De gewichten zijn in gram. De claim van de autoriteiten is dat het gemiddelde gewicht van in omloop zijnde quarters 5.60 gram is. In deze opgave willen we dit toetsen. Gegeven is dat het gemiddelde gewicht van de 25 geselecteerde quarters 5.620 gram is en de standaardafwijking in deze getallen 0.075 is.

- a) Stel de nulhypothese H_0 en de alternatieve hypothese H_a op met betrekking tot het gemiddelde gewicht van alle in omloop zijnde quarters op.
- b) Bereken de standaardscore z (toetsingsgrootte) gebaseerd op de data.
- c) Bereken de P -waarde behorend bij deze score (tabel 1 kan hierbij van nut zijn).
- d) Geef aan wat er gebeurt als er in de huidige context een *fout van type 2* wordt gemaakt.

z	Oppervlakte onder de curve <i>links van</i> z
-2.932	0.002
-2.000	0.023
-1.960	0.025
-1.000	0.159
-0.400	0.345
0.253	0.600
1.333	0.909
1.645	0.950
1.875	0.970
2.500	0.994

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



Figuur 2: *standaard normale curve corresponderend met tabel 1*

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

Tabel 2: *te verdienen punten per onderdeel*

Rekenmachine is niet nodig en niet toegestaan tijdens dit tentamen

Cijfer=(aantal punten + 10)/10

Succes!