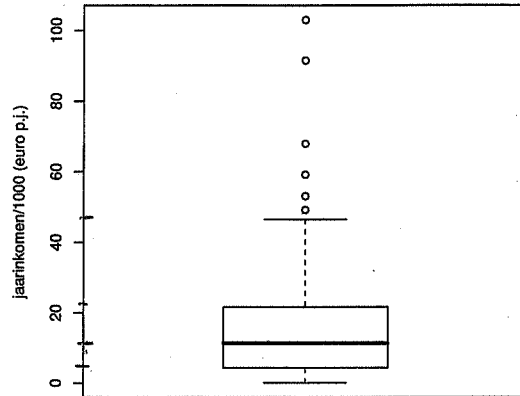

Tentamen Toegepaste Statistiek
voor studenten IMM en AI
Vrije Universiteit, Faculteit Exacte Wetenschappen
25 maart 2008, 8.45-10.45 uur
Een rekenmachine is bij dit tentamen niet toegestaan.

1.
 - a) Uit een enquête onder alle deelraadsleden van een grote stad, van wie bijna de helft reageerde, blijkt dat 70% van de onderzochte raadsleden niet betrokken is bij door hun deelraad gesubsidieerde organisaties. Geef aan wat voor type steekproefnemen dit is en wat hierbij een probleem zou kunnen zijn.
 - b) Leg, eventueel aan de hand van een voorbeeld, uit wat 'cluster trekken' (cluster sampling) is.
 - c) Welke typen studies kunnen we onderscheiden?
 - d) Leg, eventueel aan de hand van een voorbeeld, uit wat een case-control studie is.
 - e) Beargumenteer met welk type studie onderzocht zou kunnen worden of het hoge gehalte aan nicotine in de elektronische sigaret hartkloppingen veroorzaakt.
2. Geef datatype en meetschaal van onderstaande metingen. Geef tevens aan welke visualisatie het meest geschikt is voor de data en waarom.
 - a) Onderzoek van het IOC heeft aangetoond dat de luchtvervuiling in Peking geen directe bedreiging vormt voor de gezondheid van atleten die deelnemen aan de Olympische Spelen. De luchtvervuiling werd gemeten als het gehalte aan schadelijke stoffen in de lucht, uitgedrukt in p.p.m. (parts per million).
 - b) Na een tv-debat tussen presidentskandidaten A, B en C kunnen kijkers per SMS aangeven op wie van de drie kandidaten ze denken te gaan stemmen. Met 1, 2 of 3 geven ze te kennen op, respectievelijk, kandidaat A, B of C te gaan stemmen.
3.
 - a) In Figuur 1 staat een boxplot van 200 jaarinkomens. Wat kun je aflezen uit deze grafiek met betrekking tot lokatie en schaal van deze dataset? (De whiskers verbinden de box in dit geval met de meest extreme datapunten die niet meer dan 1.5 maal de interkwartiel afstand van de box afliggen.)
 - b) Schets een histogram dat bij deze data zou kunnen passen en becommentarieer de vorm van de dataverdeling.
 - c) Welk van de twee lokatiematen zal voor deze dataset groter zijn, het steekproefgemiddelde of de steekproefmediaan en waarom?



Figuur 1: *Boxplot van jaarinkomens.*

4. Hieronder staan van 25 jaren de gemiddelde temperaturen in de maand juli gemeten in De Bilt.

15.9	18.1	16.8	14.7	14.9
13.7	14.2	15.3	14.1	18.1
15.7	17.2	17.0	15.4	17.8
19.3	17.0	15.4	15.9	15.8
16.3	18.9	20.1	17.2	18.2

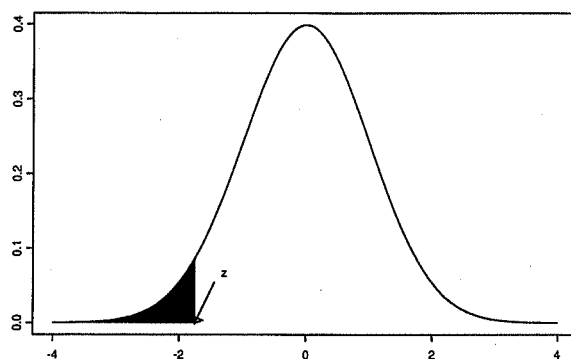
- a) Bereken op basis van relatieve frequentie een schatting voor de kans dat de gemiddelde temperatuur in juli in De Bilt boven de 18.0 graden ligt.

Uit voorgaande jaren is gebleken dat de populatieverdeling van de gemiddelde temperatuur in juli in De Bilt bij benadering normaal is met verwachting (populatiegemiddelde) $\mu = 16.5$ en standaardafwijking 1.5.

- b) Geef op basis van deze bewering een schatting voor de kans dat de gemiddelde temperatuur in juli in De Bilt boven de 18.0 graden ligt. Je kunt hiervoor Tabel 1 aan het eind van dit tentamen gebruiken.
- c) Geef, ook op basis van deze bewering, de kans dat het gemiddelde van 25 metingen van de gemiddelde temperatuur in juli in De Bilt boven de 16.8 graden ligt. Ook hiervoor kun je Tabel 1 aan het eind van dit tentamen gebruiken.
- d) De schattingen van de kans onder a) en b) verschillen nogal. Wat zou hiervan de oorzaak kunnen zijn?
- e) Wat valt op aan de uitkomsten bij b) en c)? Hoe verklaar je dit?

5. Iemand gooit twee keer met een zuiver muntstuk. Als munt bovenkomt krijgt ze 2 euro, bij kop krijgt ze niets.
- Construeer een kansruimte voor dit experiment.
 - Bereken de kans dat het aantal verkregen euro's in de tweede worp gelijk is aan dat van de eerste worp. Doe dit door netjes de gebeurtenis te formuleren binnen de in a) geconstrueerde kansruimte.
 - Beschouw de stochastische variabele X die het totaal verkregen bedrag beschrijft bij het experiment. Construeer de kansfunctie van X netjes op basis van de formele definitie.
 - Bereken de verwachting EX van X ; geeft niet alleen de uitkomst, maar laat zien hoe je hier aan komt.
6. Bij een productieproces komen van tijd tot tijd defecte produkten van de lopende band. Elk produkt is, onafhankelijk van een ander produkt, met een kans van 1% defect. Van de dagproductie worden drie produkten aselekt gekozen. Geef bij onderstaande onderdelen niet alleen het antwoord, maar geef ook kort aan welke rekenregel of eigenschap van kansen je bij de berekening gebruikt.
- Wat is de kans dat alle drie de produkten defect zijn.
 - Beschrijf het complement van bovenstaande gebeurtenis en geef de kans op deze gebeurtenis.
 - Wat is de kans dat tenminste één van de drie produkten defect is.
- Defecte produkten worden door twee technici gerepareerd. Persoon 1 behandelt 60% van de defecte produkten en persoon 2 behandelt 40%. Bij persoon 1 gaat in 5% van de gevallen de reparatie niet goed, bij persoon 2 in 10% van de gevallen.
- Als een produkt uit de dagproductie defect blijkt te zijn na een incorrect uitgevoerde reparatie, wat is dan de kans dat deze reparatie is uitgevoerd door persoon 2?
7. In opgave 4 is voor 25 jaren de gemiddelde temperatuur in de maand juli gemeten. Het gemiddelde van die metingen is $\bar{x} = 16.52$ en de standaardafwijking is $s_n = 1.69$. In deze opgave bepalen we op basis van die metingen een betrouwbaarheidsinterval voor het populatiegemiddelde μ van de gemiddelde temperatuur in juli. Op basis van de Centrale Limietstelling weten we dat het gemiddelde van 25 metingen van de gemiddelde temperatuur in juli in De Bilt met een kans van ongeveer 95% tussen de waarden $\mu - 2s_n/\sqrt{25}$ en $\mu + 2s_n/\sqrt{25}$ ligt.
- Leid uit deze kansuitspraak de *algemene uitdrukking* af voor een 95% betrouwbaarheidsinterval voor de onbekende waarde μ .

- b) Geef op basis van deze algemene uitdrukking het 95% betrouwbaarheidsinterval voor μ uitgaande van de gegeven dataset.
 - c) Wat is de interpretatie van het in b) bepaalde interval?
8. Van een aselechte steekproef van 100 personen zijn 12 personen linkshandig. We willen op basis van deze steekproef de bewering toetsen dat 10% van de populatie linkshandig is.
- a) Stel de nulhypothese H_0 en de alternatieve hypothese H_a met betrekking tot het percentage p van linkshandigen op.
 - b) Bereken de schatting $\hat{p}$ voor p .
 - c) Bereken de standaardscore z (toetsingsgroottheid) voor $\hat{p}$.
 - d) Bereken de P -waarde behorend bij deze score. Je kunt hiervoor weer Tabel 1 gebruiken.
 - e) Wat is de conclusie van de toets?



Figuur 2: *Standaard normale curve corresponderend met Tabel 1.*

z	Oppervlakte onder de curve <i>links van</i> z
-2.000	0.023
-1.960	0.025
-1.000	0.159
-0.050	0.480
0.000	0.500
0.255	0.601
0.650	0.742
0.700	0.758
1.645	0.950
2.500	0.994

Tabel 1: *Oppervlakten onder de standaard normale curve corresponderend met Figuur 2.*

1a: 3	2a: 5	3a: 4	4a: 3	5a: 4	6a: 2	7a: 3	8a: 2
1b: 3	2b: 5	3b: 3	4b: 4	5b: 3	6b: 2	7b: 3	8b: 2
1c: 2		3c: 3	4c: 4	5c: 3	6c: 2	7c: 2	8c: 3
1d: 3			4d: 2	5d: 3	6d: 4		8d: 2
1e: 3			4e: 1				8e: 2

Tabel 2: *Aantal punten per onderdeel.*

Cijfer=(aantal punten + 10)/10

