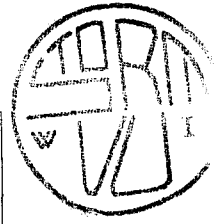


**Tentamen Toegepaste Statistiek
voor studenten
Informatiekunde en Kunstmatige Intelligentie**
vrije Universiteit, Divisie Wiskunde en Informatica
27 maart 2003, 13.30-16.30 uur

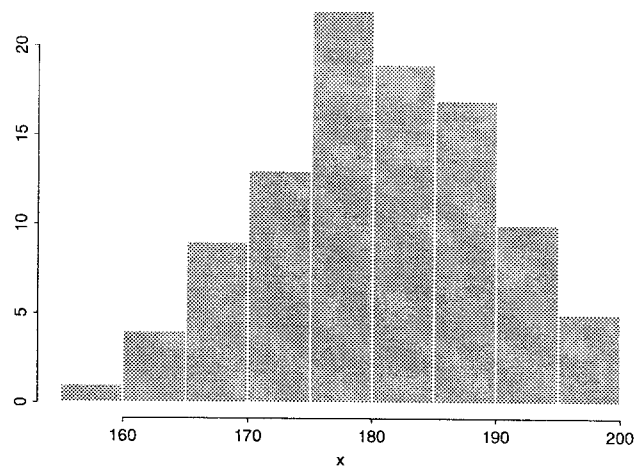


1.
 - a) Leg aan de hand van een voorbeeld uit wat 'cluster sampling' is.
 - b) Leg uit wat een 'placebo effect' is en geef aan wat er in de praktijk wordt gedaan om dit effect te ondervangen.
 - c) Geef een voorbeeld van een case-control studie en geef aan wat karakteristiek is voor zo'n studie.
 - d) Iemand beweert dat mensen die veel vis eten door de bank genomen langer worden dan mensen die dat niet doen. Om zijn bewering kracht bij te zetten wordt van vijftig aselekt gekozen viseters in de leeftijd 30-40 jaar de lengte opgemeten. Hetzelfde wordt gedaan voor vijftig niet-viseters. De gemiddelde lengte van de vijftig viseters was 1.76m. en die van de visweigeraars 1.72m. De conclusie luidt: viseters worden langer dan niet-viseters! Geef aan welk gevaar schuilt in de opzet van dit onderzoek. Geef verder uw oordeel over de getrokken conclusie.
2.
 - a) In een autoblad wordt een studie verricht naar de veiligheid van verschillende automerken en -types. De veiligheid wordt aangegeven met de getallen 0 t/m 3 (voor onveilig t/m veiligst). Geef het data type en de meetschaal aan.
 - b) Gedurende een maand wordt de gemiddelde snelheid gemeten van alle auto's die per uur langs de afslag WTC van de A10 suizen. Geef aan welk type data dit oplevert en op welke meetschaal deze worden gemeten.
 - c) Leg uit wat de volgende bewering betekent: 'De consumentenprijsindex in 2003, bij referentiejaar 1984, is 184.'
3. Een verzekeringsmaatschappij heeft aan 32 willekeurig gekozen volwassenen in de leeftijd 30-70 jaar gevraagd bij hoeveel auto ongelukken ze betrokken zijn geweest. Hieronder de resultaten van dit onderzoek.

0	1	0	3	2	1	0	2
1	1	1	0	2	0	4	1
2	0	0	1	0	2	1	3
1	3	0	0	1	0	5	4

- 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 lengen 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.
- 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. Hieronder staan de gewichten van 25 aselekt gekozen volwassenen.

67.2	72.8	57.6	74.5	69.5
71.1	54.0	96.4	64.1	56.5
85.3	69.2	89.0	57.9	69.4
76.6	48.1	70.6	74.5	69.2
72.1	82.2	58.1	71.8	88.8

De gewichten zijn in kilogram.

- a) Bereken op basis van relatieve frequenties een schatting voor de kans dat een willekeurige volwassene minder weegt dan 70.0 kg.

Uit eerder onderzoek is gebleken dat de populatieverdeling van deze gewichten bij benadering normaal is met populatiegemiddelde $\mu = 75$ en standaardafwijking 12.5.

- b) Geef op basis van deze bewering de kans dat een willekeurige volwassene minder weegt dan 70.0 kg. Tabel 1 aan het eind van dit tentamen kan bij het berekenen van deze kans van nut zijn.

5. We gooien twee keer met een zuivere dobbelsteen.
 - a) Construeer een kansruimte voor dit experiment.
 - b) Bereken de kans dat het aantal ogen in de tweede worp minstens zo groot is als het aantal in de eerste worp. Doe dit door netjes de gebeurtenis te formuleren binnen de in a) geconstrueerde kansruimte.
 - c) Beschouw de stochastische variabele X die de het aantal ogen weergeeft op de dobbelsteen met het minst aantal ogen. Construeer de kansfunctie van X netjes op basis van de formele definitie.
 - d) Bereken de verwachting EX van X .
6. Een chipfabrikant wil een grote partij van zijn chips aan een afnemer verkopen. Hij is verplicht een indicatie te geven voor het percentage defecte exemplaren in de partij. Hiertoe neemt hij een steekproef van omvang 100, en bepaalt of de exemplaren in deze steekproef deugdelijk zijn. Vier exemplaren blijken defect.
 - a) Geef de gebruikelijke *puntschatting* $\hat{p}$ voor het onbekende percentage p chips in de populatie

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

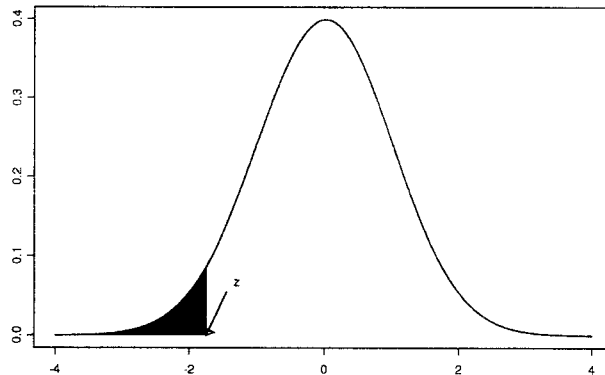
$$\sqrt{n} \frac{\hat{p} - p}{\sqrt{\hat{p}(1 - \hat{p})}}$$

bij benadering standaard normaal verdeeld.

- b) Leid op basis van deze uitspraak de *algemene uitdrukking* af voor een 95% betrouwbaarheidsinterval voor p .
 - c) Geef het 95% betrouwbaarheidsinterval voor p in de huidige situatie en ligt de interpretatie van dat interval toe.
 - d) Hoe groot moet de steekproef uit de partij bij benadering zijn om een 95% betrouwbaarheidsinterval ter breedte 0.005 te krijgen?
7. In het verleden was de gemiddelde 'levensduur' van een type batterij 9.2 uur. De fabrikant heeft het productieproces bijgesteld en wil nagaan of deze aanpassing een positief effect op de gemiddelde levensduur van het product heeft.
 - a) Stel de nulhypothese H_0 en de alternatieve hypothese H_a op met betrekking tot de gemiddelde levensduur van de batterijen die met het nieuwe procedé zijn zijn vervaardigd.

De fabrikant neemt de proef op de som en bepaalt de levensduur van 25 batterijen in een aselechte steekproef. De gemiddelde levensduur van de batterijen in deze steekproef bedraagt 9.4 uur, en de standaardafwijking bedraagt 0.5 uur.

- b) Bereken de standardscore z (toetsingsgrootheid) gebaseerd op de data.
 c) Bereken de P -waarde behorend bij deze score (tabel 1 kan hierbij van nut zijn).
 d) Wordt H_0 ten gunste van H_a verworpen bij een significantieniveau $\alpha = 0.01$?
 Licht uw antwoord toe.



Figuur 2: standaard normale curve corresponderend met tabel 1

z	Oppervlakte onder de curve <i>links van</i> z
-2.932	0.002
-2.000	0.023
-1.960	0.025
-0.625	0.266
-0.400	0.345
0.253	0.600
1.188	0.883
1.645	0.950
1.960	0.975

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

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

Tabel 2: te verdienen punten per onderdeel

Rekenmachine is niet nodig en niet toegestaan tijdens dit tentamen

$$\text{Cijfer} = (\text{aantal punten} + 10) / 10$$

Succes!

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
1.188	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: 3	2c: 3	3c: 3	4a: 3	5a: 4	6a: 3	6d: 3	7c: 4
1b: 5	2a: 3	3a: 3	3d: 5	4b: 4	5b: 3	6b: 3	7a: 4	7d: 3
1c: 4	2b: 3	3b: 4	3e: 3	4c: 4	5c: 3	6c: 3	7b: 4	

Tabel 2: te verdienen punten per onderdeel

Rekenmachine is niet nodig en niet toegestaan tijdens dit tentamen

Cijfer=(aantal punten + 10)/10

Succes!