

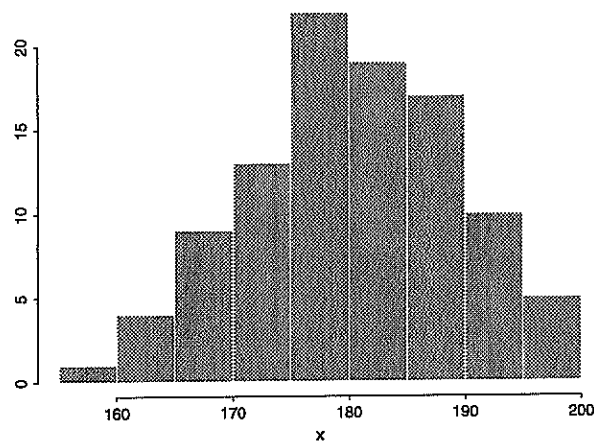
|                                                                                                                                                                                                                                                           |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p style="text-align: center;"><b>Tentamen Toegepaste Statistiek</b><br/><i>voor studenten</i><br/><b>Informatiekunde en Kunstmatige Intelligentie</b><br/><i>vrije Universiteit, Faculteit Exacte Wetenschappen</i><br/>4 juni 2007, 18.30-20.30 uur</p> |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

1.
  - a) Leg aan de hand van een voorbeeld uit wat 'systematic sampling' 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) 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) Gedurende de voorbije twintig jaar is veel onderzoek gedaan naar de bewering die onder c) is gegeven. Veel studies zijn verschenen in verschillende wetenschappelijke tijdschriften. Op basis van deze artikelen, wordt een nieuwe studie gedaan naar het waarheidsgehalte van de bewering. Hoe wordt dit type studie genoemd?
2.
  - a) In een bosperceel leeft een specht die iedere avond actief begint te hameren op een boom. Een (gepensioneerde) vogelliefhebber in de buurt houdt gedurende een jaar iedere dag bij hoe laat het hameren begint. Geef voor de resulterende dataset het data type en de meetschaal aan.
  - b) Bij een wetenschappelijk tijdschrift worden aangeboden artikelen aan drie referenten toegestuurd met bijgevoegd een evaluatieformulier. Onder andere op het gebied 'originaliteit' moeten de referenten het hun toegestuurde artikel beoordelen. De mate van originaliteit moet worden uitgedrukt in de cijfers 1 (zeer origineel) t/m 5 (niet origineel). De hoofdredacteur van het tijdschrift krijgt zo per artikel drie van zulke scores binnen. Geef aan wat het type van deze data is en ook wat de meetschaal is.
  - 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 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.
- 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*

- Een fabrikant van cola doet onderzoek naar de preciese hoeveelheid cola die in de blikjes zit die uit zijn productie komen. Van 25 aselekt gekozen blikjes uit de dag-productie 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.* cola bevat.

Uit eerder onderzoek is gebleken dat de populatieverdeling van de inhoud van een colablikje 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.* cola 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 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. Bij een groot ministerie wil men de animo voor deelname aan een PC-privé project inschatten. Hiertoe wordt aan 100 aselekt gekozen ambtenaren van het ministerie de vraag voorgelegd of men wil meedoen aan het project. Van de 100 ambtenaren gaven er 16 aan mee te willen doen.

- a) Geef de gebruikelijke *puntschatting*  $\hat{p}$  voor het onbekende percentage  $p$  van geïnteresseerde ambtenaren.

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.

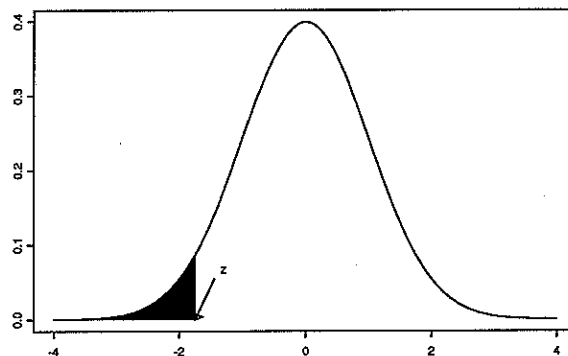
- b) Leid op basis van de bijbehorende normale benadering de *algemene uitdrukking* af voor een 95% betrouwbaarheidsinterval voor  $p$ .

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

- c) Geef het 95% betrouwbaarheidsinterval voor  $p$  in de huidige situatie en licht de interpretatie van dat interval toe.
- d) Hoe groot moet de steekproef uit de ambtenaren op het ministerie bij benadering zijn om een 95% betrouwbaarheidsinterval ter breedte 0.02 te krijgen?

7. In opgave 4 is voor 25 aselekt gekozen colablikjes uit de dagproductie van een fabrikant de preciese inhoud gemeten. Het gemiddelde van die metingen is  $\bar{x} = 33.26$  en de standaardafwijking 0.16. In deze opgave willen we op basis van die metingen toetsen of de gemiddelde inhoud van alle blikjes in de dagproductie 33.20 cl kan zijn, zoals uit eerder onderzoek zou blijken.

- a) Stel de nulhypothese  $H_0$  en de alternatieve hypothese  $H_a$  op met betrekking tot de gemiddelde inhoud van de colablikjes uit de dagproductie.
- b) Bereken de standdaarscore  $z$  (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  $\alpha = 0.05$ ? 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                                           |
| -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

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

Succes!