
Tentamen Toegepaste Statistiek

voor studenten IMM en AI

Vrije Universiteit, Faculteit Exacte Wetenschappen

2 juni 2008, 18.30-20.30 uur

Een rekenmachine is bij dit tentamen niet toegestaan.

1.
 - a) Uit de resultaten van meer dan tienduizend enquêtes onder huishoudens, bedrijven en boeren in dertig kleine dorpen in Europa blijkt o.a. dat meer dan de helft van de dorpsbewoners in Nederland hun boodschappen in het eigen dorp doet. Geef aan wat voor type steekproefnemen dit zou kunnen zijn geweest. Denk je dat de resultaten van deze studie representatief zijn? Licht je antwoord toe.
 - b) Leg, eventueel aan de hand van een voorbeeld, uit wat 'cluster trekken' (cluster sampling) is.
 - c) Welke typen studies kunnen we onderscheiden?
 - d) Uit een studie onder 1282 ouderen met en zonder depressie bleek dat depressieve mensen gemiddeld veertien procent minder vitamine D in hun bloed hebben dan mensen zonder depressie. Welk type studie is dit? Licht je antwoord kort toe.
2. Geef datatype en meetschaal van onderstaande metingen. Geef tevens aan welke visualisatie het meest geschikt is voor de data en waarom.
 - a) In een call center worden continu telefoontjes voor een aantal verschillende bedrijven afgehandeld. Elk telefoontje wordt bij binnenkomst meteen geënclassificeerd naar bedrijf.
 - b) Een bestand bevat de netto inkomens van alle inwoners van Amsterdam, afgerond op gehele euro's.
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	3	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.

4. Een bierbrouwerij doet onderzoek naar de precieze hoeveelheid bier die in de blikjes zit die uit haar productie komen. 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 frequentie een schatting voor de kans dat een willekeurig blikje uit de dagproductie minder dan 32.98 cl bier bevat.

Uit eerder onderzoek is gebleken dat de populatieverdeling van de inhoud van een bierblikje bij benadering normaal is met verwachting (populatiegemiddelde) $\mu = 33.18$ en standaardafwijking $\sigma = 0.20$.

- b) Geef op basis van deze bewering een schatting voor de kans dat een willekeurig blikje minder dan 32.98 cl bier bevat. Je kunt hiervoor Tabel 1 aan het eind van dit tentamen gebruiken.
- c) Geef, ook op basis van deze bewering, de kans dat de gemiddelde inhoud van 25 blikjes uit de dagproductie lager dan 33.08 cl is. 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?
5. Iemand gooit twee keer met een onzuiver muntstuk. De kans op kop is 0.4 en de kans op munt is 0.6.
(Bij onderstaande onderdelen mag je in je antwoord een breuk of produkt laten staan.)

- a) Construeer een kansruimte voor dit experiment.

Als kop bovenkomt krijgt de persoon die gooit 1 euro, bij munt krijgt ze niets.

- b) Bepaal 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.
- c) Beschouw de stochastische variabele X die het bij het experiment totaal verkregen bedrag beschrijft. Construeer de kansfunctie $p(x)$ van X netjes op basis van de formele definitie.
- d) Bereken de verwachting EX van X ; geef niet alleen de uitkomst, maar laat zien hoe je hier aan komt.

6. Voor een bepaalde ziekte bestaat een diagnostisch middel. Als een persoon de ziekte heeft, geeft het middel dit in 99% van de gevallen aan. Als een persoon de ziekte niet heeft, geeft het middel dit in 95% van de gevallen aan. De ziekte is bij 1 op de 1000 mensen aanwezig en een willekeurig persoon heeft de ziekte onafhankelijk van een andere persoon. Geef bij onderstaande onderdelen niet alleen het antwoord, maar geef ook kort aan welke rekenregel of eigenschap van kansen je bij de berekening gebruikt.

Voor de berekeningen kan het handig zijn de volgende definities te gebruiken: zij

A de gebeurtenis dat het middel aangeeft dat een persoon de ziekte heeft;

B_1 de gebeurtenis dat een persoon de ziekte heeft;

B_2 de gebeurtenis dat een persoon de ziekte niet heeft.

(*Je mag in je antwoord een breuk of produkt laten staan.*)

- Wat is de kans dat twee willekeurige personen ziek zijn?
 - Wat is de kans dat tenminste één van twee willekeurige personen ziek is?
 - Wat is de kans dat het middel aangeeft dat een persoon de ziekte heeft?
 - Wat is de kans dat een persoon de ziekte daadwerkelijk heeft, gegeven dat het middel aangeeft dat de persoon de ziekte heeft?
7. Bij een groot bedrijf wil men de animo voor deelname aan een fietsproject inschatten. Hiertoe wordt aan 100 aselekt gekozen werknemers van het bedrijf de vraag voorgelegd of men wil meedoen aan het project. Van de 100 werknemers gaven er 16 aan mee te willen doen.

- 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 grootheid

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

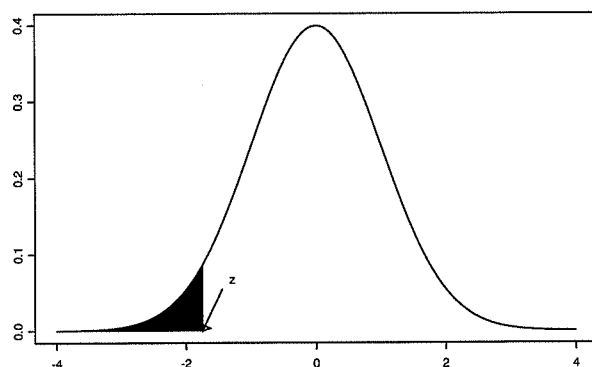
bij benadering standaard normaal verdeeld.

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

(*In onderdeel c) van deze opgave mogen breuken en wortels in het antwoord blijven staan.*)

- Geef op basis hiervan het 95% betrouwbaarheidsinterval voor p uitgaande van de gegeven data set, en licht de interpretatie van dat interval toe.

8. In opgave 4 is voor 25 aselekt gekozen bierblikjes uit de dagproductie van een brouwerij de precieze inhoud gemeten. Het gemiddelde van die metingen is $\bar{x} = 33.26$ en de standaardafwijking $s_n = 0.16$. In deze opgave willen we op basis van die metingen toetsen of de gemiddelde inhoud van alle blikjes in de dagproductie 33.18 cl kan zijn, zoals uit eerder onderzoek zou zijn gebleken.
- a) Stel de nulhypothese H_0 en de alternatieve hypothese H_a op met betrekking tot de gemiddelde inhoud van de bierblikjes uit de dagproductie.
 - b) Bereken de standaardscore z (toetsingsgrootheid) gebaseerd op de data.
 - c) Bereken de P -waarde behorend bij deze score. (Je kunt hiervoor weer Tabel 1 gebruiken.)
 - d) Wordt H_0 ten gunste van H_a verworpen bij een significantieniveau $\alpha = 0.05$? Licht je antwoord toe.



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

Tabel 2: *Aantal punten per onderdeel.*

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

