

Tentamen Toegepaste Statistiek <i>voor studenten</i> Informatiekunde en Kunstmatige Intelligentie <i>vrije Universiteit, Faculteit Exacte Wetenschappen</i> 26 augustus 2004, 13.30-16.30 uur
--

1.
 - a) Leg aan de hand van een voorbeeld uit wat 'convenience sampling' is.
 - b) Leg aan de hand van een voorbeeld uit wat 'confounding' (verstrengeling) is.
 - c) Er zijn veel publicaties waarin de uitwerking van een bepaalde kleurtoevoeging aan voedsel wordt onderzocht. Een onderzoeker duikt in deze publicaties om na te gaan of de kleurtoevoeging schadelijk is. Welk type studie is dit?
 - d) Bij een tv-programma worden kijkers opgeroepen telefonisch hun mening te geven over een politiek onderwerp, om de mening van 'het Nederlandse volk' te peilen. Geef, gebruikmakend van de geleerde terminologie, aan welke gevaren er aan dit type onderzoek kleven.
2.
 - a) Een vliegtuigspotter noteert de merken van vliegtuigen die hij gedurende een dag op een landingsbaan ziet landen. Welk type data levert dit op, en op welke meetschaal?
 - b) Een bestand bevat de netto inkomens van alle inwoners van Amsterdam, afgerond op gehele euro's. Geef het type van, alsmede de meetschaal waarop deze data zijn gegeven aan.
 - c) Geef aan welk gevaar er kleeft aan het gebruik van tijd-plots van aantallen, waarbij de schaal op de y -as niet bij nul begint.
3. Binnen de filmwereld worden prijzen uitgereikt. De meest prestigieuze is de *academy award*. Hieronder zijn de leeftijden van de 34 meest recente mannelijke winnaars van die prijs weergegeven.

32	37	36	32	51	53	33	61	35
45	55	39	76	37	42	40	32	60
38	56	48	48	40	43	62	43	42
44	41	56	39	46	31	47		

- a) Geef een tabel met 4 kolommen. In de eerste de *klassen*, lopend van 20-29 tot en met 80-89; in de tweede de bijbehorende frequentie, de derde de relatieve frequentie en in de vierde kolom de cumulatieve relatieve frequentie.
- b) Schets op basis hiervan een histogram voor de data en beschrijf de vorm van de dataverdeling kwalitatief.
- c) Geef de modale klasse aan en bereken de mediaan van de data.

4. 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 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.* 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 gooien twee keer met een zuivere munt. De afzonderlijke uitkomsten worden gecodeerd met -1 en +1. Denk aan een spelletje waarbij iemand die kop gooit 1 punt krijgt en iemand die munt gooit 1 punt inlevert.
- a) Construeer een kansruimte voor dit experiment.
- b) Bereken de kans dat de score in de tweede worp minstens zo gunstig is als 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 puntenopbrengst beschrijft bij het experiment. Construeer de kansfunctie van X netjes op basis van de formele definitie.
- d) Bereken de verwachting EX van X .
- e) Maak de gebeurtenissen die worden omschreven als 'eerste worp is +1' en 'tweede worp is -1' expliciet binnen de kansruimte en laat (volgens de formele definitie van onafhankelijkheid) zien dat deze gebeurtenissen *onafhankelijk* zijn.
6. Bij een groot adviesbureau wil men de animo voor deelname aan een PC-privé project inschatten. Hiertoe wordt aan 100 aselekt gekozen werknemers van het adviesbureau de vraag voorgelegd of men wil meedoen aan het project. Van deze 100 werknemers gaven er 16 aan mee te willen doen.

- 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 grootheid

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

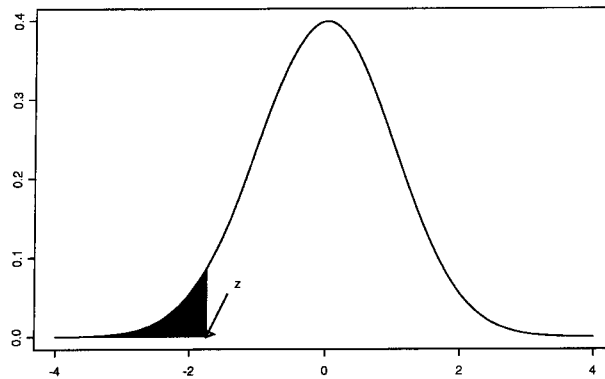
bij benadering standaard normaal verdeeld.

- 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 werknemers van het bedrijf 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 standaardscore 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.

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



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

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

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