

Tentamen Toegepaste Statistiek voor studenten Informatiekunde
vrije Universiteit, Divisie Wiskunde en Informatica
28 maart 2002, 13.30-16.30u.

1.
 - a) Er zijn heel veel statistische studies geweest naar het effect van hoogspanningsmasten op het optreden van bepaalde soorten kanker. Op basis van de bijbehorende resultaten wordt onderzocht of er een verband is. Geef aan welk type studie hier wordt gedaan. Licht uw antwoord toe.
 - b) Men wil de 'kwaliteit' van horoscopen bestuderen. In het bijzonder wil men nagaan of een horoscoop die bij een persoon hoort (wat geboortedag betreft) meer over die persoon zegt dan een willekeurige andere horoscoop. Beschrijf (gebruikmakend van statistische terminologie) een proef die kan worden gedaan om aan de onderzoeksvraag tegemoet te komen.
 - c) Om na te gaan hoeveel uur een Amerikaan gemiddeld per nacht slaapt, is een onderzoek gedaan. Naar verhouding van het aantal inwoners van de betreffende staat, is uit iedere staat een groep inwoners aselekt gekozen. Deze personen zijn gevraagd hoeveel uur ze gemiddeld per nacht slapen. Hoe heet het type steekproeftrekken waar we hier mee te maken hebben? Beschrijf dat type ook in het algemeen.
 - d) Leg aan de hand van een voorbeeld uit wat het begrip *verstremgeling* of *confounding* inhoudt.
 - e) Een werknemer van een fietsbellenfabriek heeft metingen gedaan aan de diameter van het oor (dat waar je tegenaandrukt bij een fietsbel) en een kwaliteitskenmerk dat met de klank van de bel te maken heeft. Het blijkt dat een goede kwaliteit van de bel gepaard gaat met een (relatief) klein bel-oor (geen oorbell!). De verschillen in diameter zijn zeer klein, maar de samenhang is verrassend. De werknemer adviseert zijn superieuren de bel-oor-machine op een kleinere diameter af te stellen om de kwaliteit van het eindproduct te verbeteren. Geef aan of u het advies onderschrijft of niet. Licht uw antwoord toe, gebruikmakend van de statistische terminologie zoals die tijdens het college is gebezigd.
2.
 - a) Op een bepaalde plaats wordt het ozongehalte in de lucht dagelijks gemeten. Geef aan met welk type data we hier te maken hebben. Geef ook aan wat de meetschaal is.
 - b) In een middelgrote stad in de VS worden dagelijks aanvragen bij de sociale dienst ingediend. Over de periode van een jaar worden de aanvragers naar etniciteit ingedeeld. Welk type data levert dit op en wat is de meetschaal?
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. Geef aan welke visualisatie het meest geschikt zou zijn voor de volgende data. Motiveer uw antwoord.
- a) Van brandalarmen zijn er zo'n 40 verschillende merken. Deze worden in verschillende landen in verschillende hoeveelheden verkocht. De brancheorganisatie heeft informatie over deze aantallen voor de landen Nederland en België, en is met name ook geïnteresseerd in de verschillen in verkopen.
 - b) Een autofabrikant brengt al gedurende tien jaar vier types auto's op de markt. Hij is met name geïnteresseerd in het verloop (in de tijd) van het onderlinge marktaandeel van zijn producten.
 - c) Bij een personeelswerfingsbureau komen op een bepaalde vacature 120 sollicitaties binnen. Alle sollicitanten worden ingedeeld in klassen 1 t/m 5, afhankelijk van geschiktheid. Kandidaten uit de eerste klasse zijn het meest geschikt en die uit klasse 5 het minst. Men wil in één oogopslag een indruk krijgen van de geschiktheid van de sollicitanten.
5. Voor een steekproef van 22 in omloop zijnde exemplaren van de munteenheid 'quarter' in de VS, zijn hieronder de gewichten weergegeven.

5.60	5.63	5.58	5.56	5.66	5.58	5.57	5.59
5.67	5.61	5.84	5.73	5.53	5.58	5.52	5.65
5.57	5.71	5.59	5.53	5.63	5.68		

De gewichten zijn in gram.

- a) Bereken op basis van relatieve frequenties een schatting voor de kans dat een willekeurige quarter lichter is dan 5.55 gram.

Door de autoriteiten wordt beweerd de gewichten van de in omloop zijn de quarters een normale verdeling hebben met populatiegemiddelde $\mu = 5.60$ en standaardafwijking 0.08.

- b) Geef op basis van deze bewering de kans dat een willekeurige quarter lichter is dan 5.55 gram. Tabel 1 aan het eind van dit tentamen kan bij het berekenen van deze kans van nut zijn.
6. 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.
- Construeer een kansruimte voor dit experiment.
 - 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.
 - Beschouw de stochastische variabele X die de puntenopbrengst beschrijft bij het experiment. Construeer de kansfunctie van X netjes op basis van de formele definitie.
 - Bereken de verwachting EX van X .
 - 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.
7. De RVD is geïnteresseerd in het percentage Nederlanders dat de nieuwjaarstoespraak van de koningin op de radio beluistert. Om hierover informatie te krijgen, wordt onderzoek onder $n = 500$ aselekt gekozen Nederlanders gehouden. Van deze 500 personen bleken er 63 de nieuwjaarstoespraak te hebben beluisterd.
- Geef de gebruikelijke *puntschatting* $\hat{p}$ voor het onbekende percentage p Nederlanders dat naar de toespraak heeft geluisterd.

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

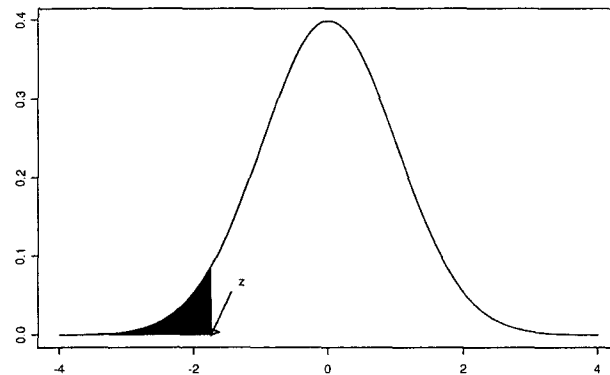
$$\sqrt{n} \frac{\hat{p} - p}{\sqrt{\hat{p}(1 - \hat{p})}}$$

bij benadering standaard normaal verdeeld.

- Leid op basis van deze uitspraak de *algemene uitdrukking* af voor een 95% betrouwbaarheidsinterval voor p .
 - Geef het 95% betrouwbaarheidsinterval voor p in de huidige situatie en ligt de interpretatie van dat interval toe.
8. In opgave 5 zijn de gewichten gegeven van 'quarters' voor een steekproef van omvang 22. De claim van de autoriteiten is dat het gemiddelde gewicht van in omloop zijnde quarters 5.60 gram is. In deze opgave willen we dit toetsen. Gegeven is dat

het gemiddelde gewicht van de 22 geselecteerde quarters 5.619 gram is en de standaardafwijking in deze getallen 0.075 is.

- Stel de nulhypothese H_0 en de alternatieve hypothese H_a op met betrekking tot het gemiddelde gewicht van alle in omloop zijnde quarters op.
- Bereken de standaardscore z (toetsingsgrootheid) gebaseerd op de data.
- Bereken de P -waarde behorend bij deze score (tabel 1 kan hierbij van nut zijn).
- Geef aan wat er gebeurt als er in de huidige context een *fout van type 2* wordt gemaakt.



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

z	Oppervlakte onder de curve <i>links van</i> z
-2.932	0.002
-1.960	0.025
-0.625	0.266
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 1*

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

Tabel 2: *te verdienen punten per onderdeel*

Cijfer=(aantal punten + 10)/10

Succes!