

Toets Statistiek voor INF en BIT (Module 6 -201400256)

Vrijdag 16 januari 2014, 13.45-16.00 uur

Deze toets bestaat uit 4 opgaven, een formuleblad en de $N(0,1)$ -, t -, χ^2 -, binomiale en Shapiro-Wilk-tabellen. Een gewone rekenmachine is toegestaan, een programmeerbare (GR) niet.

1. Uit een aselechte steekproef van 24 medisch specialisten bleek dat het gemiddelde bruto (full time) jaarinkomen 178 000 is en de standaardafwijking 54 000.
De 5-getallen-samenvatting luidt: 112 000, $Q_1 = 145 000$, $m = 160 000$, $Q_3 = 190 000$, 394 000.
(Alle gegevens zijn van enige jaren geleden).
- Wat kun je op grond van deze informatie over de vorm van het histogram van jaarinkomens zeggen?
 - Bepaal de z-scores van de hoogste en de laagste jaarinkomens.
 - Welk percentage van de specialisten heeft in dit onderzoek een jaarsalaris hoger dan 190 000 Euro (dit is de norm voor maximumsalarissen in de semioverheid: 130% van een ministersalaris)?
Als we voor dat percentage een 95%-betrouwbaarheidsinterval zouden willen bepalen met een breedte van hooguit 8%, hoeveel specialisten moeten we dan opnemen in een (nieuwe) steekproef?
 - Bepaal een 95%-betrouwbaarheidsinterval voor het verwachte jaarinkomen van een specialist op grond van bovenstaande steekproefgegevens, uitgaande van een normale verdeling voor de jaarinkomens.
 - Is de veronderstelling van normaliteit van de inkomensverdeling noodzakelijk voor het in d. bepaalde interval? Motiveer je antwoord.
 - Bij het uitvoeren van de toets van Shapiro-Wilk vonden we 0.908 als waarde van de toetsingsgrootte. Wat betekent deze waarde voor de normaliteitsveronderstelling? Geef alléén de hypothesen, het kritieke gebied en je conclusie voor de betreffende toets met $\alpha = 10\%$
2. Een verzekeringsmaatschappij kan zich voor de reparatie van schade aan auto's wenden tot twee garages. Het vermoeden bestaat dat garage 1 duurder is dan garage 2. Om te onderzoeken of dit vermoeden waar is, laat men van zeven willekeurig gekozen auto's de schade taxeren door zowel garage 1 als garage 2. De getaxeerde bedragen (in honderden Euro's) zijn als volgt:

	Bedragen							\bar{x}	s^2
Garage 1	15.2	20.4	19.0	22.6	6.0	12.6	10.6	15.2	34.80
Garage 2	14.6	18.2	16.8	23.0	5.4	11.6	9.8	14.2	33.99
Verschil	0.6	2.2	2.2	-0.4	0.6	1.0	0.8	1.0	0.87

In de kolom \bar{x} staan steekproefgemiddelden; de steekproefvarianties staan vermeld in de kolom s^2 .

- Voer een toets uit om te onderzoeken of het vermoeden dat garage 1 structureel hoger taxeert dan garage 2 waar is. Neem onbetrouwbaarheidsdrempel 5% en ga uit van normale verdelingen. Voer de toetsingsprocedure uit met behulp van het kritieke gebied, uitgaande van normale verdeling(en).
- Bepaal de overschrijdingskans bij de toets van onderdeel a. en vermeld voor welke waarden van α de nulhypothese verworpen zou worden. (Schat m.b.v. de tabellen, een interval mag ook).
- Voer nu een non-parametrische toets uit om het vermoeden (dat garage 1 hoger taxeert) opnieuw te toetsen, weer met $\alpha = 5\%$.

3. Bij het beoordelen van de bruikbaarheid en de geldigheid van een vragenlijst, testen onderzoekers de vragenlijst vaak met verschillende onafhankelijke steekproeven van respondenten. Kennis van de verschillen en overeenkomsten tussen de steekproeven en de bijbehorende populaties is belangrijk om de geldigheid van de vragenlijst te kunnen beoordelen. Een nieuwe vragenlijst werd getest op de twee onafhankelijke steekproeven die in de tabel staan vermeld.

	Managers en professionals	Part time MBA studenten
Steekproefomvang	162	109
Percentage mannen	95.0	68.9

- Bereken het 95%-betrouwbaarheidsinterval voor het verschil in (populatie)percentages mannen tussen 'managers en professionals' en 'part time MBA studenten'.
 - Geef een interpretatie, in voor iedereen begrijpelijke bewoordingen, van het in a. bepaalde interval.
 - Zou je uit het in a. bepaalde interval concluderen dat de percentages (statistisch) verschillen? Motiveer je antwoord.
 - Bepaal een 90%-betrouwbaarheidsinterval voor het **verwachte aantal mannen** in een groep van 2000 part time MBA-studenten, op grond van bovenstaande beperkte steekproef van 109 part time MBA-studenten.
4. Een zekere vitamine zou van invloed zijn op het voorkomen van verkoudheid. Bij een experiment kregen 100 proefpersonen de vitamine en 100 een placebo. Aan alle 200 werd verteld dat zij vitamine kregen. Dit zijn de resultaten:

	Minder verkoudheden	Meer verkoudheden	Geen verschil
Controle groep	39	21	40
Behandelde groep	51	20	29

- Als we willen toetsen of vitamines van invloed zijn op verkoudheid, moeten we dan op deze gegevens een toets op onafhankelijkheid of juist een toets op de homogeniteit toepassen?
- Pas de in a. gekozen toets toe m.b.v. de toetsingsprocedure en gebruik $\alpha = 5\%$.

Normering: toetscijfer = $1 + \frac{\text{aantal punten}}{45} \times 9$, afgerond op 1 decimaal

1						2			3				4		Tot
a	b	c	d	e	f	a	b	c	a	b	c	d	a	b	
2	2	3	4	1	3	6	2	5	3	1	2	4	2	5	45