

Naam

Toetsende Statistiek
Rijksuniversiteit Groningen

Lente
Docent: John Nerbonne

Tentamen

di. 22 juni om 14 uur
tentamenhal

Belangrijke instructies

1. Schrijf uw naam & studentnummer hierboven, schrijf voorletters op alle bladzijden.
2. U heeft 2 uur de tijd.
3. Geef op alle vragen een antwoord.
4. Let op dat sommige vragen zwaarder tellen dan andere (er zijn in totaal 100 punten).

Overzicht

1. Terminologie, basisbegrippen (40%)
2. Interpretaties, Grafieken en Tabellen (25%)
3. Keuze van Toets (35%)

1 Basisbegrippen, 'W'/'O' voor waar/onwaar

1. Het significantie-niveau $p = 0.04$ betekent dat er 4% kans is dat de alternatieve hypothese waar is (gezien de steekproef). W O
2. De p -waarde in een statistische toets houdt rekening met factoren zoals systematische meetfouten (vertekeningen). W O
3. U bestudeert de invloed van leeftijd op kans van herstel van afasie die als gevolg van een beroerte optreedt. De verklarende variabele (`leeftijd`) lijkt sterk niet-normaal verdeeld te zijn. Stelling: het is zinvol om de variabele `leeftijd` logaritmisches te transformeren voordat men de analyse uitvoert. W O
4. Als men `herstel` in een getal kan meten, zou men enkelvoudige lineaire regressie in het zojuist beschreven geval mogen toepassen. W O
5. Als `herstel` alleen categorisch (wel/niet hersteld) gemeten kan worden, is enkelvoudige lineaire regressie goed toepasbaar. W O

- | | | | |
|-----|--|---|---|
| 6. | Als men de regressie analyse in (3 t/m 5) boven uitvoert dan zou men met 1 vrijheidsgraad rekenen omdat er maar één verklarende variabele is. | W | O |
| 7. | U gebruikt ANOVA om de verwerkingstijden tussen vier implementaties (van één algoritme) te vergelijken. Nadat U de tijden van 800 proeflopen (200 per implementatie) heeft verzameld, blijkt het dat de populatie niet normaal is, maar (rechts) scheef. Sommige looptijden zijn dus heel lang. Stelling: Men mag ANOVA verder gebruiken omdat de aantallen groot genoeg zijn. | W | O |
| 8. | Dezelfde situatie als boven, alleen heeft U maar 40 proeflopen (10 per implementatie). Stelling: ANOVA is niet geschikt voor aantallen van 40 waar de populatie vermoedelijk niet normaal verdeeld is. | W | O |
| 9. | Bij ANOVA moet men aannemen dat de standaarddeviaties in alle groepen ongeveer gelijk zijn: geen standaarddeviatie mag drie keer groter dan een andere zijn. | W | O |
| 10. | De nulhypothese in een MANOVA heeft altijd de vorm “Er is geen verschil tussen de (gemiddelden van de) populaties waarvan de steekproeven zijn gekozen.” | W | O |
| 11. | Om na te gaan of er significante verschillen zijn tussen drie groepen van scholieren, past men de <i>t</i> -toets op alle drie paren (1-2, 2-3, 1-3) toe. | W | O |
| 12. | Door te meten hoeveel een mens in een gesprek in een groep aan het woord is, kan men zijn 'spraakzaamheid' meten. Als men de verhouding in spraakzaamheid tussen mannen en vrouwen zou willen weten, zou men echtgenoten voor hun spraakzaamheid kunnen toetsen. Een regressie analyse zou moeten aantonen of er een verschil is tussen mannen en vrouwen in 'spraakzaamheid'. | W | O |
| 13. | U heeft gestandaardiseerde toetsresultaten om het woordenschatniveau van drie groepen voor ieder 25 deelnemers te gaan meten. De drie groepen hebben drie verschillende voorbereidende cursussen gehad. U voert een ANOVA uit. Stelling: U berekent de <i>p</i> -waarde op basis van $F(3, 75)$ vrijheidsgraden. | W | O |
| 14. | Enkelvoudige ANOVA is nog steeds toepasbaar als één van de subgroepen een sd heeft die anderhalf keer zo groot is als een andere (bijv. 7,5 ms. vs. 5,0 ms. in een reactietijd experiment). Dit verschil is klein genoeg om de toets door te laten gaan. | W | O |
| 15. | U bestudeert het effect van sexe enerzijds en burgerlijke staat anderzijds op de acceptatie van web-sites voor “e-shopping” (winkelen). Omdat mogelijke interactie bestaat tussen deze factoren (sexe en burgerlijke staat), is een meervoudige ANOVA de aangewezen opzet voor de analyse van gegevens. | W | O |

2 Interpretaties, Grafieken en Tabellen

2.1 Grafieken (15%)

Het residuendiagram is afkomstig van een kleinste-kwadraatlijn die gehaalde opleidingsjaren vergelijkt met het inkomstnivo van de ouders (over de laatste eeuw). De x as geeft de tijd aan en de y residuen.



1. Ziet men in deze grafiek of mensen meer opleidingsjaren in de loop van deze eeuw halen? Waarin is dit te zien?
2. Wat zijn de eenheden op de y as?
3. Leg uit wat de hoge waarde in de jaren '50 betekenen.
4. Leg uit wat de lage waarde aan het begin van de eeuw betekent.

Tabellen (5 pt.)

1. U bestudeert de aantrekkelijkheid van vier soorten web-sites voor jongeren. Van elke groep worden 15 sites aselekt gekozen. De nulhypothese voorspelt geen verschil in de groepen, wat u d.m.v. ANOVA kunt toetsen. De hierbij gebruikte F -statistiek is afhankelijk van twee vrijheidsgraden, en wordt gemeld bijv. als $F(df_1, df_2)$. Deze moet men weten om de kritieke waarden na te kunnen kijken. Wat zijn de vrijheidsgraden in dit geval?

Interpretatie

1. U wilt d.m.v. een logistische regressie een schatting maken van het belang van inkomst als voorspeller voor de beslissing om wel of niet mee te doen bij een aanbieding voor een multimedia PC dat op een door u beheerde webstek staat. U heeft data over 450 klanten en de logistische regressie laat zien dat u de hypothese dat inkomst geen invloed heeft met met $p < 0.005$ kunt afwijzen. Is dit voldoende? Als niet, wat moet u bekijken om een schatting van de invloed van inkomst op de koopbeslissing te hebben?

3 Toetsingen

Keuze van toets (15pt) Het is moeilijk om taalvaardigheden terug te krijgen na hersenletsel, bijv. door een beroerte. U vergelijkt twee programma's voor rehabilitatie, waarvan het eerste zich vooral concentreert op alledaagse communicatie, en het tweede op het beheersen van taalkundige regels. Bij beide programma's doen mannen en vrouwen mee en uw wilt ook kijken of de programma's verschillen per sexe. Patiënten worden aselekt aan het ene of het andere programma toegewezen, en hun verbetering na drie maanden vergeleken op basis van een standaard toets. Er volgt een samenvatting van de resultaten:

| Gem. Verbetering | mannen | vrouwen |
|------------------|--------------|-------------|
| Therapie | | |
| Aldagstaal | 5.52 (n=11) | 7.10 (n=9) |
| Taalkundig | 11.05 (n=10) | 10.81 (n=8) |

Het tweede (taalkundige) programma deed het bij deze patienten duidelijk beter. Maar misschien is het verschil aan de kans te wijten. Is er hier statistisch significante evidentie dat het tweede programma beter is? Analyseer het resultaat als een hypothese toets.

1. Welke toets is hier van toepassing? Geef redenen voor uw keuze.
2. Formuleer H_0 en H_a .
3. Hoeveel vrijheidsgraden zijn er bij de vraag over welke soort programma beter is?
4. Als de waarschijnlijkheid van de toetsstatistiek $p = 0.04$ is, wat valt te concluderen?
5. Hoe gaat u na of de programma's ook per sexe verschillen?

Keuze van toets (20pt) U onderzoekt of stemhoogte verstaanbaarheid beïnvloedt. De toonhoogte van de stem meet men in frequentie, vibraties per seconde (Hz), en is dus een numerieke variabele. Verstaanbaarheid is ook in een numerieke variabele weer te geven, bijv. aantal verstaanbare woorden per honderd. Welke statistische analyse is geschikt om na te gaan of de numerieke variabele ‘toonhoogte van de stem’ invloed heeft op een tweede numerieke variabele ‘verstaanbaarheid’?

1. Welke statistische toets zou toegepast moeten worden?
2. Wat zijn de verdere voorwaarden waaraan de gegevens zouden moeten voldoen?
3. Formuleer H_0 en H_a
4. Als aan één of meerdere voorwaarden niet voldaan kan worden, zijn er dan terugvalopties (in de keuze van toets)?