2^e deeltentamen Statistiek
27 juni 2012

I Schrijf je naam en studentnummer op elk vel dat je inlevert.

II Het gebruik van het boek van J.A. Rice, aantekeningen, handouts en zakrekenmachines is toegestaan.

III U mag in ieder onderdeel de conclusies van voorgaande onderdelen gebruiken, ook als u die (nog) niet bewezen hebt.

IV Motiveer steeds uw antwoord door duidelijk aan te geven welke argumenten en welke resultaten u gebruikt om een bepaalde conclusie te trekken.

V U heeft 3 uur de tijd voor het tentamen.

VI Achter elke vraag staat het aantal punten dat met de vraag te behalen is. De puntenverdeling is: 1 - 20, 2 - 25, 3 - 22, 4 - 20, 5-13.

VII Veel succes!

Opgave 1 Stel dat er een onderzoek is gedaan naar de associatie van 20 genen en een bepaalde ziekte. De volgende p-waarden voor de associaties zijn gevonden: $p_i = \frac{1}{2}e^{-\frac{1}{2}(i-1)}$ voor $i \in \{1, \ldots, 20\}$. Welke genen beoordeelt u als statistisch significant geassocieerd als u zoveel mogelijk statistisch significante associaties wilt vinden en

a 10 pt) de kans dat u minstens 1 gen ten onrechte als geassocieerd benoemt kleiner moet zijn dan 0,05.

b 10 pt) u accepteert dat het verwachte aantal genen dat ten onrechte als significant geassocieerd is gekwalificeerd maximaal 5% is van het aantal genen dat als significant geassocieerd is gekwalificeerd, d.w.z., maximaal 5% van de positieve resultaten is fout-positief.

Opgave 2 In een experiment is de bloeddruk gemeten van 10 proefpersonen waarvan de body mass index (BMI) bekend is. De onderzoeksvraag is of er een relatie is tussen bloeddruk en BMI. De resultaten van het experiment staan in de tabel hieronder:

a 7pt) Vind de regressielijn voor de bloeddruk als functie van de BMI.

b 9pt) Bepaal 95% betrouwbaarheidsintervallen voor het snijpunt (β_0) en de helling (β_1).

c 9pt) Bepaal een 95% betrouwbaarheidsinterval voor de verwachte bloeddruk van een persoon met een BMI van 23.

individu i	BMI x_i	x_i^2	bloedruk (in mmHg) y_i	RSS $y_i - \beta_0 - \beta_1 x_i$	$\to (\ldots)^2$
1	18	324	110	-3,696	13,6
2	18	324	114	-0,304	0,09
3	20	400	115	0,094	0,008
4	21	441	121	5,498	30,23
5	22	484	117	0,884	0,781
6	24	576	112	-5,326	28,36
7	24	576	120	2,674	7,07
8	25	625	119	1,069	1,142
9	26	676	117	-1,536	2,35
10	30	900 +	121	-2,762	7,62

Opgave 3 Er is een onderzoek gedaan met 40 personen met depressieve klachten. 20 personen hebben een antidepressivum gekregen, de 20 andere personen een placebo. Na een maand is de deelnemers gevraagd om op een schaal van één tot vijf aan te geven hoe de klachten zijn t.o.v. een maand eerder. De betekenis van de schaal is als volgt: de klachten zijn I: sterk verminderd, II: verminderd, III: gelijk gebleven, IV: verergerd, V: sterk verergerd. De resultaten van het experiment staan in de tabel:

Behandeling	minder Uitkomst				verergerd	
	I	II	III	IV	V	
Antidepressivum	6	5	4	3	2	X
Placebo	4	4	4	5	3	Y

De onderzoeksvraag is of het antidepressivum de depressieve klachten vermindert. De onderzoekers gaan uit van een significantieniveau (α) van 0,05.

a 6 pt) Geef de nulhypothese en de alternatieve hypothese.

b 8 pt) Bereken de Mann Whitney statistiek (U). Als twee patiënten dezelfde uitkomst hebben en een patiënt het antidepressivum heeft gekregen en de ander het placebo, schrijf dan 50% van het gewicht van dat paar toe aan de placebo-arm (alsof de patiënt met het placebo een beter resultaat had dan de patiënt met het antidepressivum) en 50% van het gewicht van dat paar aan de behandeling (alsof de patiënt met de behandeling een beter resultaat heeft geboekt).

c 8 pt) Gebruik de Mann Whitney statistiek om te bepalen of het antidepressivum als werkzaam beschouwd mag worden op basis van dit experiment. (Indien u onderdeel b niet heeft berekend, neem dan aan dat $U = 238$.)

Opgave 4 Zij $C = (X, Y, Z)^T \sim N(\mu_C, \Sigma_{C,C})$ met

$$\mu_C = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \text{ en } \Sigma_{C,C} = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 3 & -1 \\ -1 & -1 & 4 \end{pmatrix}. \text{ Zij } D = \begin{pmatrix} U \\ V \end{pmatrix} \text{ met } U = 2X + 3Y - Z$$

en $V = X + Z$.

a 7pt) Bereken de covariantiematrix Σ_{DD}.

b 3pt) Zijn U en V onafhankelijk?

c 5pt) Bereken de correlatiecoëfficiënt tussen X en Y.

d 5pt) Stel dat je van een steekproef van C van grootte 1 weet dat X de waarde 2 heeft aangenomen. Wat is de verwachte waarde voor Y?

Opgave 5 Stel dat we n datapunten hebben van de vorm: $y_i, x_{i,1}, x_{i,2}, \dots, x_{i,m}$ voor $i \in \{1, n\}$. Stel dat bij een multivariate lineaire regressie analyse met de kleinste kwadratenmethode we de vector $\hat{\beta} := (\beta_0, \beta_1 \dots, \beta_m)^T$ vinden voor de coëfficiënten van de regressie-analyse. Stel dat we de verklarende variabelen $x_{i,j}$ als volgt transformeren: $u_{i,j} = c_j + a_j x_{i,j}$ ($a_j \neq 0 \quad \forall j \in \{1, m\}$) of in vector notatie: $u_i := (u_{i,1}, \dots u_{i,m})^T = c + Ax_i$ met $c = (c_1, \dots, c_m)^T$, $x_i = (x_{i,1}, \dots, x_{i,m})^T$ en A een $m \times m$-diagonaalmatrix, waarbij het element A_{ii} gelijk is aan a_i.

a 10 pt) Geef formules voor de regressiecoëfficiënten in de getransformeerde variabelen.

b 3 pt) Geef een voorbeeld waarbij bovenstaande transformatie handig is.

Einde.