Tentamen Regressie en Variantie-Analyse (191530440)
Vrijdag 13 april 2012 van 8:45 tot 11:45 uur

Dit tentamen bestaat uit 6 vragen en 1 bonuspuntvraag.

Als er gevraagd wordt naar een toets, geef dan aan (voor zover nog niet gegeven) :
(1) De modelveronderstellingen (het "statistisch model"),
(2) De formulering van de nulhypothese en alternatieve hypothese,
(3) De formule voor de toetsingsgrootheid, indien er meerdere formules mogelijk zijn kies dan voor de formule die rekentechnisch het eenvoudigst is.
(4) De verdeling van de toetsingsgrootheid onder de nulhypothese,
(5) Geef aan voor welke waarden van de toetsingsgrootheid de nulhypothese verworpen moet worden. (Als T de toetsingsgrootheid is, verwerp je dan de nulhypothese voor $T \geq c$? Of voor $T \leq -c$? Of voor zowel $T \geq c$ als $T \leq -c$?)

Vraag 1
Beschrijf hoe residuenplots beoordeeld moeten worden als je de modelveronderstellingen (het "statistisch model") controleert.

Vraag 2
We gaan uit van het algemene regressiemodel met k verklarende variabelen. Toon aan dat de vector van residuen E en de vector $\hat{\beta}$ (schatter van β) ongecorreleerd zijn als de Gauss-Markov-voorwaarden vervuld zijn, en onderling onafhankelijk zijn als $Y \sim N(X\beta, \sigma^2 I)$.

Vraag 3
Geef aan hoe we het model van de variantie-analyse met 1 factor in de vorm "$Y = X\beta + \varepsilon$" kunnen schrijven, en pas algemene toetsingstheorie van regressiemodellen toe om de nulhypothese $H_0 : (\alpha_1, ..., \alpha_k) = 0$ te toetsen tegen de alternatieve hypothese $H_1 : (\alpha_1, ..., \alpha_k) \neq 0$. (Bij het beantwoorden van deze vraag worden voor kansverdelingen en formules van toetsingsgrootheden **geen bewijzen** verwacht.)

Vraag 4
Beschouw enkelvoudige lineaire regressie zonder constante β_0 :

$$Y_i = \beta_1 x_i + \varepsilon_i \qquad (i = 1, 2, ..., n),$$

waarbij we aannemen dat de storingen ε_i onderling onafhankelijk en $N(0, \sigma^2)$ -verdeeld zijn. Beschrijf de constructie van betrouwbaarheidsintervallen voor β_1. Geef daarbij ook aan wat de geschatte standaardafwijking ('standard error') van de betrokken schatter is.

Vraag 5

Beschouw het regressiemodel met modelvergelijking

$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \varepsilon_i$

waar de storingen ε_i voldoen aan een AR(1)-model. Laat zien dat $\mathrm{cov}(\varepsilon) = \sigma^2 \Omega$ geldt met Ω een matrix die alleen van een correlatiecoëfficiënt ρ afhangt. Geef de formule voor de 'estimated generalized least squares estimator' van de vector β.

Vraag 6

Beschouw het regressiemodel $Y = X\beta + \varepsilon$ met de Gauss-Markov-voorwaarden. Toon aan dat $\hat{\beta}$ een consistente schatter van β is als $tr(X^T X)^{-1} \to 0$ (als $n \to \infty$).

BONUSPUNTVRAAG

Vraag A

Beschouw het algemene regressiemodel met normale verdelingen:

$Y \sim N(X\beta, \sigma^2 I_n)$,

waar X een $n \times (k+1)$ matrix is van rang $k+1$ $(k+1 < n)$. We toetsen de nulhypothese $H_0 : C\beta - \gamma = 0$ tegen de alternatieve hypothese $H_1 : C\beta - \gamma \neq 0$. De $m \times (k+1)$ matrix C heeft rang m, met $m < k+1$. Bewijs dat de (algemene) toetsingsgrootheid F onder de nulhypothese een $F_{m,n-k-1}$-verdeling heeft.

Normering:

De vragen 1 t/m 6 tellen even zwaar mee, de score van deze zes vragen wordt geschaald naar een cijfer tussen 1 en 10.

Met vraag A kun je een bonuspunt scoren.