Tentamen Regressie en Variantie-Analyse (191530440)
Woensdag 4 juli 2012 van 13:45 tot 16:45 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
Beschouw het regressiemodel $Y = X\beta + \varepsilon$. Toon aan dat de vector van residuen E loodrecht op de kolommen van X staat en dat $S = \| Y - X\beta \|^2 = (Y - X\beta)^T (Y - X\beta)$ als functie van β geminimaliseerd wordt voor $\beta = (X^T X)^{-1} X^T Y$.

Vraag 3
Geef aan hoe we het model van de variantie-analyse met 2 factoren **zonder** interactie in de vorm "$Y = X\beta + \varepsilon$" kunnen schrijven, en pas algemene toetsingstheorie van regressiemodellen toe om te toetsen of de tweede factor eigenlijk wel in het model thuishoort. (Bij het beantwoorden van deze vraag worden voor kansverdelingen en formules van toetsingsgrootheden **geen bewijzen** verwacht.)

Vraag 4
Beschouw het algemene regressiemodel met normale verdelingen: $Y \sim N(X\beta, \sigma^2 I_n)$, met X een $n \times (k+1)$ matrix van rang $k+1$ ($k+1 < n$). Beschrijf de constructie van betrouwbaarheidsintervallen voor $E(Y_0) = x_0^T \beta$ (de verwachting van Y_0 als de vector x_0 de bijbehorende waarden van de verklarende variabelen bevat). Geef daarbij ook aan wat de geschatte standaardafwijking ('standard error') van de betrokken schatter is.

Vraag 5
Toon aan dat $a^T\hat{\beta}$ de BLUE is van $a^T\beta$ als de Gauss-Markov-voorwaarden van het regressiemodel vervuld zijn.

Vraag 6
Beschouw het algemene regressiemodel met normale verdelingen: $Y \sim N(X\beta, \sigma^2 I_n)$, met X een $n \times (k+1)$ matrix van rang $k+1$ ($k+1<n$). Toon aan dat $\dfrac{(n-k-1)S^2}{\sigma^2} = \dfrac{\sum_i E_i^2}{\sigma^2}$ een χ^2_{n-k-1}-verdeling heeft.

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.