Tentamen Regressie en Variantie-Analyse (153044)
Vrijdag 9 april van 8:45 tot 11:45 uur

Dit tentamen betaat 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
Wat is het statistisch model als we het hebben over multiple regressie met 2 verklarende variabelen? (Beschrijving van data, kansverdelingen en verdere aannames.)

Vraag 2
Ga uit van het algemene regressiemodel (k verklarende variabelen, model met constante β_0) met de Gauss-Markov-voorwaarden. Toon de volgende gelijkheid aan:

$$\sum_i (y_i - \bar{y})^2 = \sum_i (\hat{y}_i - \bar{y})^2 + \sum_i e_i^2 \ .$$

Geef de betekenis/interpretatie aan van de verschillende kwadraatsommen.

Vraag 3
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 $k+1$ $(k+1 < n)$. Geef aan hoe we de nulhypothese $H : C\beta - \gamma = 0$ tegen de alternatieve hypothese $A : C\beta - \gamma \neq 0$ kunnen toetsen. (Bij het beantwoorden van deze vraag worden voor kansverdelingen en formules van toetsingsgrootheden **geen bewijzen** verwacht.)

Vraag 4
Beschrijf de constructie van een voorspellingsinterval (confidence interval for a future observation y_0) in geval van enkelvoudige lineaire regressie ('simple linear regression'). Geef daarbij ook de (geschatte) standaardafwijking die de lengte van het interval bepaalt.

Vraag 5
We gaan uit van het algemene regressiemodel met k verklarende variabelen. Toon aan dat de vector van residuen e en de vector b (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 6
Ga uit van het algemene regressiemodel (k verklarende variabelen, model met constante β_0) met de Gauss-Markov-voorwaarden. Toon aan dat

$s^2 = \sum_{i=1}^{n} e_i^2 /(n-k-1)$ een zuivere schatter is van σ^2.

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 $H: C\beta - \gamma = 0$ tegen de alternatieve hypothese $A: C\beta - \gamma \neq 0$. De $m \times (k+1)$ matrix C heeft rang m, met $m < k+1$. Bewijs dat de relevante toetsingsgrootheid 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.