Tentamen Regressie en Variantie-Analyse (191530440)
Vrijdag 10 april 2015 van 8:45 tot 11:45 uur

Dit tentamen bestaat uit 6 vragen.

Dit tentamen is "gesloten boek", rekenmachine is niet toegstaan.

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
Beschouw enkelvoudige lineaire regressie zonder constante β_0:

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

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 2
We gaan uit van het algemene regressiemodel met k verklarende variabelen en storingen ε_i die onafhankelijk zijn en normaal verdeeld met verwachting nul en gemeenschappelijke variantie σ^2.

a. Toon aan dat de vector van residuen E en de vector van kleinste-kwadraten schatters $\hat{\beta}$ onafhankelijk zijn.

b. We kijken naar enkelvoudige lineaire regressie, het geval $k = 1$.
De aangepaste rechte lijn heeft als vergelijking:
$y = \hat{\beta}_0 + \hat{\beta}_1 x = \bar{Y} + \hat{\beta}_1(x - \bar{x})$,
met $\bar{Y}$ het steekproefgemiddelde van de afhankelijke variabele.
Toon aan dat ook $\bar{Y}$ en $\hat{\beta}_1$ onafhankelijk zijn.

Vraag 3
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 drie kwadraatsommen.

Vraag 4

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 $n \to \infty$).

Vraag 5

Van 5 monsters rubber wordt de hardheid twee keer gemeten, 1 keer volgens methode 1 en 1 keer volgens methode 2. Laat Y_{ij} de meting van monster i volgens methode j zijn ($i = 1,2,\dots,5; j = 1$ of $j = 2$). We gaan uit van variantie-analyse met 2 factoren (zonder interactie):

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} ,$$

met onafhankelijke storingen ε_{ij} die normaal verdeeld zijn met verwachting 0 en variantie σ^2. In deze opgave stellen we de volgende restricties:

$$\mu = 0 \quad \text{en} \quad \beta_1 = -\beta_2 \ .$$

a. Bepaal formules voor de kleinste kwadraten schatters $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4, \hat{\alpha}_5$ en $\hat{\beta}_2$, vereenvoudig deze formules zo veel als mogelijk.

b. Geef aan hoe we $H_0: \beta_2 = 0$ tegen $H_1: \beta_2 \neq 0$ moeten/kunnen toetsen. (Zie de vijf algemene punten van een toets aan het begin van dit tentamen.)

c. Data van dit soort worden wel met "gepaarde data" aangeduid. Voor het toetsingsprobleem van onderdeel b wordt wel de "gepaarde t-toets" gebruikt: men gaat over op verschillen $D_i = Y_{i2} - Y_{i1}$ $(i = 1,2,\dots,n,$ hier $n = 5)$ en men gebruikt als toetsingsgrootheid $T = \frac{\bar{D}}{S_D/\sqrt{n}} = \frac{\bar{D}}{S_D/\sqrt{5}}$,
met $\bar{D}$ het steekproefgemiddelde van de verschillen D_i en S_D de steekproefstandaardafwijking van de verschillen D_i.
Deze toetsingsgrootheid en uw toetsingsgrootheid van onderdeel b moeten gelijkwaardige toetsen opleveren, toon aan dat dit inderdaad het geval is.

Vraag 6

We gaan uit van het algemene regressiemodel met k verklarende variabelen en storingen ε_i die onafhankelijk zijn en normaal verdeeld met verwachting nul en gemeenschappelijke variantie σ^2. Geef aan wat de verdeling is van

$$\sigma^{-2}(\hat{\beta} - \beta)^T (X^T X)(\hat{\beta} - \beta) ,$$

en bewijs deze verdeling.

Normering:

De vragen 1 t/m 6 worden beoordeeld met een "rapportcijfer" (cijfer tussen 1 en 10). Tentamencijfer is hiervan het gemiddelde, eventueel verhoogd met een bonuspunt verdiend met de huiswerkopgaven.