# Tentamen Regressie en Variantie-Analyse (191530440)
# Woensdag 2 juli 2014 van 13:45 tot 16:45 uur

*Dit tentamen bestaat uit 6 vragen. Een rekenmachine is NIET toegestaan.*

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 het model van multiple regressie met $k$ verklarende variabelen en een constante $\beta_0$, waarbij we uitgaan van onafhankelijke en normaal verdeelde storingen met verwachting 0 en gelijke variantie $\sigma^2$. We bestuderen de residuen $Y_i - \hat{Y}_i$ .
$a$. Toon aan: $E(Y_i - \hat{Y}_i) = 0$
$b$. De residuen zijn voor grote waarden van de steekproefgrootte $n$ bij benadering normaal verdeeld met verwachting 0 en variantie $\sigma^2$. De variantie $var(Y_i - \hat{Y}_i)$ is niet exact gelijk aan $\sigma^2$ . Motiveer met formules welke van de volgende twee beweringen waar is:
(1) de varianties $var(Y_i - \hat{Y}_i)$ zijn alle (iets) groter dan $\sigma^2$,
(2) de varianties $var(Y_i - \hat{Y}_i)$ zijn alle (iets) kleiner dan $\sigma^2$.
$c$. Toon aan dat de som van de residuen $Y_i - \hat{Y}_i$ altijd exact gelijk is aan 0 .

**Vraag 2**
Beschouw het model van multiple regressie met $k$ verklarende variabelen en een constante $\beta_0$. We beschouwen nu het geval dat de storingen $\varepsilon_i$ verdeeld zijn volgens een AR(1)-model.
$a$. Geef aan wat het AR(1)-model inhoudt (geef alle modelveronderstellingen).
$b$. Laat zien dat $var(\varepsilon) = \sigma^2 \Omega$ geldt met $\Omega$ een matrix die alleen van een correlatiecoëfficiënt $\rho$ afhangt.
$c$. Geef de formule voor de ‘estimated generalized least squares estimator’ van de vector $\beta$.

**Vraag 3**
We beschouwen variantie-analyse met 2 factoren zonder interactie.
$a$. Geef aan hoe we het model van de variantie-analyse met 2 factoren zonder interactie in de vorm “$Y = X\beta + \varepsilon$” kunnen schrijven.
$b$. Pas algemene toetsingstheorie van regressiemodellen toe om te toetsen of de tweede factor eigenlijk wel in het model thuishoort. Volg het schema van 5 punten vermeld aan het begin van dit tentamen.
(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 $k+1$ $(k+1 < n)$. Toon aan dat de gebruikelijke schatter $S^2$ van $\sigma^2$ als volgt verdeeld is:

$\frac{(n-k-1)S^2}{\sigma^2}$ heeft de $\chi^2_{n-k-1}$-verdeling.

**Vraag 5**

De concentratie van een stof *A* wordt op tijdstippen $t_1 = -3$, $t_2 = -2$, $t_3 = -1$, $t_4 = 0, t_5 = 1$, $t_6 = 2$ en $t_7 = 3$ gemeten. Men wil bestuderen hoe de verwachting van de meting verandert als functie van de tijd. Voor de metingen $Y_1, Y_2, \ldots, Y_7$ geldt de volgende modelvergelijking:

$$Y_i = \beta_0 + \beta_1 t_i + \beta_2 (t_i^2 - 4) + \varepsilon_i$$

met onafhankelijke storingen $\varepsilon_i$ die verdeeld zijn volgens $N(0, \sigma^2)$.

$a.$ Leid formules af voor de (kleinste kwadraten) schatters $\hat{\beta}_0$, $\hat{\beta}_1$ en $\hat{\beta}_2$, vereenvoudig de formules als functie van de waarnemingen $Y_i$ zo veel mogelijk.

$b.$ Bepaal $se(\hat{\beta}_1)$, de geschatte standaardafwijking van $\hat{\beta}_1$, als functie van $S$. (De gebruikelijke schatter $S$ van $\sigma$ hoef je niet te herschrijven in termen van deze opgave.)

$c.$ Geef aan hoe we een 95%-betrouwbaarheidsinterval voor $\beta_1$ kunnen construeren.

**Vraag 6**

In een fabriek wordt geëxperimenteerd om de (verwachte) productie zo veel als mogelijk te verhogen. In een deelexperiment worden twee factoren $A$ en $B$ onderscheiden, met elk twee niveaus. Metingen van de productie $Y_{ijk}$ worden gepland volgens volgend schema

| Factor $A$ | Factor $B$ | |
|---|---|---|
| | 1 | 2 |
| 1 | $Y_{111}$ en $Y_{112}$ | $Y_{121}$ en $Y_{122}$ |
| 2 | $Y_{211}$ en $Y_{212}$ | $Y_{221}$ en $Y_{222}$ |

We gaan uit van variantie-analyse met 2 factoren met interactie, waarbij we uitgaan van de volgende restricties op de parameters: $\alpha_1 = \beta_1 = \gamma_{11} = \gamma_{12} = \gamma_{21} = 0$ .

$a.$ Geef aan hoe we in dit geval het model in de vorm "$Y = X\beta + \varepsilon$" kunnen schrijven.

$b.$ Geef een algemene formule voor de schatter van $\sigma^2$ (variantie van de storingen) en geef aan wat de verdeling is in de context van deze opgave.

$c.$ Geef (in algemene termen) aan hoe we $H_0: \beta_2 = \gamma_{22} = 0$ tegen $H_1: \beta_2 \neq 0 \vee \gamma_{22} \neq 0$ kunnen toetsen. Volg het schema van 5 punten vermeld aan het begin van dit tentamen.

**Normering:**

De vragen 1 t/m 6 worden beoordeeld met een "rapportcijfer" (cijfer tussen 1 en 10). Tentamencijfer is hiervan het gemiddelde.