Uitwerkingen (kort) van tentamen RVA van 2 juli 2014

Vraag 1

a.

Beschouw de vector van waarnemingen Y en de vector van voorspellingen $\hat{Y}$, dus Y_1 is het i^e element van Y en $\hat{Y}_i$ is het i^e element van $\hat{Y}$.

Het is voldoende aan te tonen: $E\left(Y-\hat{Y}\right)=0$

Aangezien: $E(Y)=X\beta$ en

$E\left(\hat{Y}\right)=E\left(X\hat{\beta}\right)=E(X(X^TX)^{-1}X^TY)=X(X^TX)^{-1}X^T\ E(Y)=\mathrm{X}\left(\mathrm{X^TX}\right)^{-1}\mathrm{X^TX\beta}=\mathrm{X\beta}$

volgt snel $E\left(Y-\hat{Y}\right)=E(Y)-E\left(\hat{Y}\right)=X\beta-X\beta=0$

b.

We kunnen schrijven: $Y-\hat{Y}=(I-X(X^TX)^{-1}X^T)Y=MY$ met $M=\ I-X(X^TX)^{-1}X^T=I-H$.

Er geldt onder andere: var($Y-\hat{Y}$) $=\sigma^2M=\sigma^2(I-H)$.

Dus $var\left(Y_i-\hat{Y}_i\right)=\sigma^2(1-h_{ii})$, hier is h_{ii} een positief getal want $\sigma^2h_{ii}=var\left(\hat{Y}_i\right)$ omdat $var\left(\hat{Y}\right)=\sigma^2H$.

De varianties $var\left(Y_i-\hat{Y}_i\right)$ zijn dus alle (iets) kleiner dan σ^2.

c.

Omdat β_0 deel uitmaakt van het model zijn alle elementen van de eerste kolom van X gelijk aan 1.

De vector van residuen $Y-\hat{Y}$ staat loodrecht op alle kolommen van X: $X^T\left(Y-\hat{Y}\right)=0$

Het loodrecht staan van $Y-\hat{Y}$ op de eerste kolom houdt in:

$\sum_i\left(Y_i-\hat{Y}_i\right)\times 1=\sum_i\left(Y_i-\hat{Y}_i\right)=0$

Dat $X^T\left(Y-\hat{Y}\right)=0$ geldt volgt uit $Y-\hat{Y}=(I-X(X^TX)^{-1}X^T)Y=MY$:

$X^T\left(Y-\hat{Y}\right)=X^TMY=\cdots=(X^T-X^T)Y=0$

Vraag 2

Voor uitwerkingen van deze opgave verwijzen we naar de lecture notes van HC9.

Vraag 3

a.

Laat Y_{ijk} de k^e meting zijn bij niveau i van factor 1 en niveau j van factor 2 ($i=1,2,\ldots,r;j=1,2,\ldots,s;k=1,2,\ldots,m$). Bij variantie-analyse met 2 factoren zonder interactie gaan we uit van onafhankelijke metingen Y_{ijk} die

$N(\mu+\alpha_i+\beta_j,\sigma^2)$-verdeeld zijn. Anders geschreven:

$Y_{ijk}=\ \mu+\alpha_i+\beta_j+\varepsilon_{ijk}$

Met onafhankelijk storingen ε_{ijk} die $N(0,\sigma^2)$-verdeeld zijn.

We moeten nog restricties stellen voor de parameters α_i and β_j .

In het volgende gaan we uit van: $\alpha_1=0$ en $\beta_1=0$.

Voor de representatie $Y=X\beta+\varepsilon$ kunnen we het volgende zeggen:

Vector Y is de vector van metingen Y_{ijk}.

Vector ε is de vector van storingen ε_{ijk} (dezelfde volgorde als bij vector Y).

Elementen van vector β: $\mu,\alpha_2,\alpha_3,\ldots,\alpha_r,\beta_2,\beta_3,\ldots,\beta_s$.

De matrix X:

Laat $a^T = (a_1, a_2, \dots, a_{r+s-1})$ een rij van X zijn, deze rij wordt bepaald door de meting Y_{ijk} die erbij hoort: $E\left(Y_{ijk}\right) = \mu + \alpha_i + \beta_j = a^T\beta$.
Bij $i = j = 1$ geldt $a_1 = 0$ rest elementen van a zijn nul.
Bij $i = 2$ en $j = 3$ geldt: $a_1 = 1, a_2 = 1, a_{r+2} = 1$, rest van de elementen van a zijn nul.
Etc.
b.
De 5 punten van de toets:

(1) Voor model zie onderdeel a

(2) We toetsen H_0: $(\beta_2, \beta_3, \dots, \beta_s) = 0$ tegen H_1: $(\beta_2, \beta_3, \dots, \beta_s) \neq 0$

(3) Toetsingsgrootheid: $F = \frac{(SSE_H - SSE_M)/(s-1)}{SSE_M/(n-r-s+1)}$

Met $SSE_M = \sum_{ijk}\left(Y_{ijk} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j\right)^2$, de kwadraatsom van de residuen van het volle model,

Met $SSE_H = \sum_{ijk}\left(Y_{ijk} - \hat{\mu} - \hat{\alpha}_i\right)^2$, de kwadraatsom van de residuen van het model onder de nulhypothese,

Het aantal vrijheidsgraden $k - r$ van het algemene toetsingsprobleem hebben we vervangen door $s - 1$ (het aantal parameters dat nul gesteld wordt onder de nulhypothese).

Het aantal vrijheidsgraden $n - k - 1$ hebben we vervangen door $n - r - s + 1$, in het laatste is $n = r \times s \times m$ het totaal aantal metingen en $r + s - 1$ is de dimensie van β in ons probleem.

(4) Volgens de theorie heeft F onder de nulhypothese de F-verdeling met $r + s - 1$ en $n - r - s + 1$ vrijheidgraden.

(5) We verwerpen de nulhypothese als $F \geq c$.

Vraag 4
Voor uitwerkingen van deze opgave verwijzen we naar de lecture notes, het bewijs van Stelling 9 (HC5).

Vraag 5
a.
Het model is ook hier weer in de vorm $Y = X\beta + \varepsilon$, uitgaande van de gesuggereerde volgorde voor de elementen van Y hebben we de volgende matrix X:

$$X = \begin{pmatrix} 1 & -3 & 5 \\ 1 & -2 & 0 \\ 1 & -1 & -3 \\ 1 & 0 & -4 \\ 1 & 1 & -3 \\ 1 & 2 & 0 \\ 1 & 3 & 5 \end{pmatrix}$$

De schatters worden bepaald door:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (X^T X)^{-1} X^T Y$$

Verdere uitwerking:

$$X^TX = \begin{pmatrix} 9 & 0 & 0 \\ 0 & 28 & 0 \\ 0 & 0 & 84 \end{pmatrix}$$

$\hat{\beta}_0 = \sum_i Y_i/9$, $\hat{\beta}_1 = (-3 \times Y_1 - 2 \times Y_2 - Y_3 + Y_5 + 2 \times Y_6 + 3 \times Y_7)/28$,
$\hat{\beta}_2 = (5 \times Y_1 - 3 \times Y_3 - 4 \times Y_4 - 3 \times Y_5 + 5 \times Y_7)/84$.

$b.$

Er geldt: $var(\hat{\beta}) = \sigma^2 (X^TX)^{-1}$

Dus $var(\hat{\beta}_1) = \sigma^2 d$ met d het (2,2)-element van $(X^TX)^{-1}$, dus $d = \frac{1}{28}$.

Daarom $se(\hat{\beta}_1) = S/\sqrt{28}$

$c.$

Volgens de theorie geldt: $\frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \sim t_{n-k-1} = t_{7-2-1} = t_4$

Laat c het getal zijn dat volgens de t_4-verdeling zo gekozen is dat
$P\left(-c < \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} < c\right) = 0.95$

Dan heeft het 95%-BI voor β_1 als grenzen: $\hat{\beta}_1 \pm c \times se(\hat{\beta}_1)$,

Immers $-c < \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} < c$ en $\hat{\beta}_1 - c \times se(\hat{\beta}_1) < \beta_1 < \hat{\beta}_1 + c \times se(\hat{\beta}_1)$ zijn gelijkwaardig.

Vraag 6

$a.$

Model van de variantie-analyse met 2 factoren met interactie: $Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$

Definieer $Y = \begin{pmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \end{pmatrix}$ en $\beta = \begin{pmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \gamma_{22} \end{pmatrix}$ dan is het model van de vorm $Y = X\beta + \varepsilon$

Met de volgende matrix X:

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$b.$

Schatter van σ^2: $S^2 = \sum_{ijk} \left(Y_{ijk} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_{ij}\right)^2 / (n - k - 1)$

Het algemene aantal vrijheidgraden $n - k - 1$ moeten we nog vertalen naar deze context.

$k + 1$ is dimensie van β, is dus 4. $n = 8$ (aantal waarnemingen).

Dus: $S^2 = \sum_{ijk} \left(Y_{ijk} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_{ij}\right)^2 / 4$

In het algemeen heeft $(n - k - 1)S^2/\sigma^2$ een χ^2_{n-k-1}-verdeling.

In deze vraag: $4S^2/\sigma^2$ heeft een χ^2_4-verdeling

c.

De 5 punten zijn als volgt:

(1) $Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$ met onafhankelijke storingen ε_{ijk} die $N(0, \sigma^2)$-verdeeld zijn (is ook in vorm $Y = X\beta + \varepsilon$ te schrijven)

(2) We toetsen $H_0: \beta_2 = \gamma_{22} = 0$ tegen $H_1: \beta_2 \neq 0 \vee \gamma_{22} \neq 0$

(3) Toetsingsgrootheid: $F = \frac{(SSE_H - SSE_M)/2}{SSE_M/4}$

Met $SSE_M = \sum_{ijk}\left(Y_{ijk} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_{ij}\right)^2$, de kwadraatsom van de residuen van het volle model,

Met $SSE_H = \sum_{ijk}\left(Y_{ijk} - \hat{\mu} - \hat{\alpha}_i\right)^2$, de kwadraatsom van de residuen van het model onder de nulhypothese,

Het aantal vrijheidsgraden $k - r$ van het algemene toetsingsprobleem is hier gelijk aan 2 (het aantal parameters dat nul gesteld wordt onder de nulhypothese).

Het algemene aantal vrijheidsgraden $n - k - 1$ is gelijk aan $8 - 3 - 1 = 4$.

(4) Onder de nulhypothese heeft F de F-verdeling met 2 en 4 vrijheidsgraden.

(5) We verwerpen de nulhypothese als $F \geq c$.