STOCKHOLMS UNIVERSITET,
MATEMATISKA INSTITUTIONEN,
Avd. Matematisk statistik

Tentamen: Linjära statistiska modeller (MT5001), 2022-08-11

Kristoffer Lindensjö. E-post: kristoffer.lindensjo@math.su.se.

Tillåtna hjälpmedel: Miniräknare och formelblad (tillhandahålles av institutionen).

Återlämning: information meddelas via kursforum.

Tentamen består av 5 uppgifter. Varje korrekt löst uppgift ger 10 poäng.

- Resonemang ska vara klara, tydliga och kortfattade.
- Svar ska motiveras om inte annat framgår.
- Börja varje uppgift på nytt papper.
- Numrera tydligt varje blad med uppgift och bladordning.
- Skriv ditt kodnummer på varje blad du lämnar in (men inget namn).

Preliminära betygsgränser:

A	B	C	D	E
45	40	35	30	25

Vissa av följande kvantiler kan komma att bli användbara

$$\begin{aligned}
F_{0.05}(2,200) &= 3.04106\\
F_{0.05}(2,199) &= 3.04129\\
F_{0.05}(2,198) &= 3.04152\\
F_{0.05}(2,27) &= 3.35413\\
F_{0.05}(2,28) &= 3,34039\\
F_{0.05}(2,29) &= 3.32765\\
F_{0.05}(2,30) &= 3.31583.
\end{aligned}$$

Lycka till!

Uppgift 1

En finansanalytiker antar att avkastningen för en aktie kan uttryckas som

$$Y_t = \beta_1 SMB_t + \beta_2 HML_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2) \quad \text{iid},$$

där Y_t är avkastningen vid tiden t för aktien, SMB_t står för "small minus big" och representerar skillnaden i avkastning mellan stora och små bolag (för hela marknaden) vid tiden t samt HML_t står för "high minus low" och representerar skillnaden i avkastning mellan så kallade värdebolag och tillväxtbolag (för hela marknaden) vid tiden t. Alla värden uttrycks i procent, d.v.s. 1 innebär 1 %.

Följande data finns tillgänglig.

Tid t	Avkastning Y_t	"Small minus big" SMB_t	"High minus low" HML_t
1	1.3	2.1	0.7
2	0.4	1.1	0.6
3	-0.1	1.2	0.8
4	0.9	1.3	1.1
5	2.1	3.1	1.2
6	1.6	2.2	0.7

Table 1: Avkastning, "small minus big"-faktor och "high minus low"-faktor.

(A) Bestäm designmatrisen **A** för modellen. (2 p)

(B) Beräkna minsta kvadrat-skattningen av parametervektorn $(\hat{\beta}_1, \hat{\beta}_2)$ givet att

$\sum_t SMB_t^2 = 23.2$

$\sum_t HML_t^2 = 4.63$

$\sum_t SMB_t HML_t = 9.78$

$\sum SMB_t Y_t = 14.25$

$\sum_t HML_t Y_t = 5.7.$ (4 p)

(C) Beräkna skattningen av feltermsvariansen σ^2 givet att

$$\text{Kvs(Residual)} = 0.7014.$$

(1 p)

(D) Ange ett 95-procentigt konfidensintervall för β_1 och tolka resultatet. (2 p)

(E) Beskriv vad antagandet om likavarians (homoskedasticitet) innebär och hur det kan granskas när man gör multipel linjär regression. (1 p)

Uppgift 2

Betrakta följande modell för multipel linjär regression

$$Y_i = \beta_1 x_i + \beta_2 z_i + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2) \ iid, i = 1, ..., N$$

(A) Ange minsta-kvadrat-skattningarna för $\hat{\beta}_1$ och $\hat{\beta}_2$ uttryckta med $x_1, ..., x_N$, $z_1, ..., z_N$, $Y_1, ..., Y_N$.

Ledning: kom ihåg att

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

(4 p)

(B) Härled väntevärdena för $\hat{\beta}_1$ och $\hat{\beta}_2$. (4 p)

(C) Ange ett villkor uttryckt med $x_1, ..., x_N, z_1, ..., z_N$ som är sådant att under detta villkor så kommer $\hat{\beta}_1$ och $\hat{\beta}_2$ att vara oberoende (ditt svar behöver inte motiveras). (2 p)

Uppgift 3

I tabellen nedan återfinns data gällande livslängd i månader y_{ij} för oberoende tester av olika typer av lampor (där $i = 1, 2, 3$ motsvarar lampptyp). Alla tester utfördes på samma sätt.

Lamptyp 1	Lamptyp 2	Lamptyp 2
0.687	0.837	0.107
0.891	1.453	1.582
0.186	2.658	0.742
0.694	2.676	0.143
0.905	2.360	1.531
0.770	2.346	0.455
0.737	0.560	1.187
0.851	1.905	1.792
0.115	0.812	1.312
0.167	1.165	1.660

Följande stickprovsvarianser kan beräknas för de olika typerna av lampor

$$s_1^2 = 0.100, s_2^2 = 0.660, s_3^2 = 0.409.$$

För alla livslängder sedda tillsammans erhålls stickprovsvariansen 0.565.

(A) Du bestämmer dig för att använda en variansanalys-modell med ensidig indelning av modelltyp I.

Vilket antagande gällande varianserna för lamptypernas livslängder görs implicit givet detta modellval? (4 p)

(B) Använd modellen nämnd i **(A)** och data för att testa på 5%-nivån om typerna av lampor kan anses vara likvärdiga gällande livslängd.

(6 p)

Uppgift 4

Betrakta vår vanliga allmänna linjära modell

$$\mathbf{Y} = \mathbf{A}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}).$$

Antag att $\dim(\boldsymbol{\theta}) = 2$ och att σ^2 är en känd konstant.

(A) Hitta fördelningen för skattningen $\widehat{\theta}_1$. (3 p)

(B) Hitta fördelningen för $\widehat{\theta}_1 + \widehat{\theta}_2$. (4 p)

(C) Hitta ett konfidensintervall för uttrycket i **(B)** (3 p)

Uppgift 5

Ange definitionen för en allmän MA(1)-tidsserie. Visa huruvida den är kovariansstationär eller inte. (10 p)