STOCKHOLMS UNIVERSITET,
MATEMATISKA INSTITUTIONEN,
Avd. Matematisk statistik

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

Uppgift 1

(A) Antalet observationer är $N = 6$ och antalet parametrar $k = 2$. Designmatrisen blir

$$\mathbf{A} = \begin{pmatrix} 2.1 & 0.7 \\ 1.1 & 0.6 \\ 1.2 & 0.8 \\ 1.3 & 1.1 \\ 3.1 & 1.2 \\ 2.2 & 0.7. \end{pmatrix}$$

.

(B) Låt $\mathbf{Y} = (Y_1, \ldots, Y_6)^T = (1.3, 0.4, -0.1, 0.9, 2.1, 1.6)$ vara observationsvektorn. Allmänna formeln för minsta kvadrat-skattningen ger

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\hat{\beta}_1, \hat{\beta}_2)^T \\ &= (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Y} \\ &= \begin{pmatrix} 23.2 & 9.78 \\ 9.78 & 4.63 \end{pmatrix}^{-1} \begin{pmatrix} 14.25 \\ 5.7 \end{pmatrix} \\ &= \begin{pmatrix} 0.3935 & -0.8311 \\ -0.8311 & 1.9715 \end{pmatrix} \begin{pmatrix} 14.25 \\ 5.7 \end{pmatrix} \\ &= \begin{pmatrix} 0.8694 \\ -0.6055. \end{pmatrix} \end{aligned}$$

(C) Antalet frihetsgrader för att skatta feltermsvariansen är $N - k = 6 - 2 = 4$. Det följer att

$$\hat{\sigma}^2 = \frac{\text{Kvs(Residual)}}{4} = \frac{0.7014}{4} = 0.1753.$$

(D) Konfidensintervallet ges av

$$\hat{\beta}_1 \pm t_{0.025}(4)\sqrt{\hat{\sigma}^2 \mathbf{S}_{11}^{-1}},$$

där $\mathbf{S}_{11}^{-1}$ motsvarar element $(1, 1)$ i matrisen $\mathbf{S}^{-1} = (\mathbf{A}^T\mathbf{A})^{-1}$. Tolkningen är att i det fall intervallet inte innehåller 0 så har SMB en signifikant betydelse. Vi använder de värden vi har och uttrycker intervallet som en funktion av den relevanta t-kvantilen

$$0.8694 \pm t_{0.025}(4) \times \sqrt{0.1753 \times 0.3935}.$$

(Värdet $t_{0.025} = 2.7764$ ger att intervallet inte innehåller 0).

(E) Antagandet innebär att residualerna ska ha samma varians för alla observationer. Detta kan exempelvis verifieras genom att plotta residualerna mot de anpassade värdena. En sådan plott ska inte visa några trender eller mönster för residualerna.

Uppgift 2

Detta motsvarar en tillämpning av den allmänna linjära modellen $\mathbf{Y} = \mathbf{A}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$ med

$$\mathbf{A} = \begin{pmatrix} x_1 & z_1 \\ \vdots & \vdots \\ x_N & z_N. \end{pmatrix}$$

och $\boldsymbol{\theta} = (\beta_1, \beta_2)^T$. Motsvarande skattning ges av

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{Y}.$$

(A) Givet $\widehat{\boldsymbol{\theta}} = (\widehat{\beta}_1, \widehat{\beta}_2)^T$ och ovanstående erhålls med enkla matrisberäkningar att

$$\begin{aligned}\widehat{\beta}_1 &= \frac{\sum_i z_i^2 \sum_i x_i Y_i - \sum_i x_i z_i \sum_i z_i Y_i}{\sum_i x_i^2 \sum_i z_i^2 - (\sum_i x_i z_i)^2},\\ \widehat{\beta}_1 &= \frac{-\sum_i x_i z_i \sum_i x_i Y_i + \sum_i x_i^2 \sum_i z_i Y_i}{\sum_i x_i^2 \sum_i z_i^2 - (\sum_i x_i z_i)^2}.\end{aligned}$$

(B) Eftersom detta är en tillämpning av den allmänna linjära modellen följer att $E(\widehat{\boldsymbol{\theta}}) = \boldsymbol{\theta}$, vilket visas enligt vanligt förfarande (se Sundbergs kompendium kapitel 2.3). Av detta följer att $E(\hat{\beta}_1) = \beta_1$ och $E(\hat{\beta}_2) = \beta_2$.

(C) Från formelbladet följer att kovariansmatrisen för $\widehat{\boldsymbol{\theta}} = (\widehat{\beta}_1, \widehat{\beta}_2)^T$ ges av $\sigma^2(\boldsymbol{A}^T\boldsymbol{A})^{-1}$. Under antagande om att $\sum_i x_i z_i = 0$ (vilket alltså är det sökta villkoret) ger enkla matrisberäkningar i vårt fall att

$$\begin{aligned}Cov(\widehat{\beta}_1, \widehat{\beta}_2) &= \sigma^2(\boldsymbol{A}^T\boldsymbol{A})^{-1}_{12}\\ &= 0.\end{aligned}$$

Uppgift 3

(A) De antas implicit vara lika; jmfr. Sundbergs kompendium kapitel 4.2.
(B) Vi beräknar

$$KVS_{totalt} = 29 \cdot 0.565 = 16.373$$

$$KVS_{inom} = 9(s_1^2 + s_2^2 + s_3^2) = 10.524$$

$$KVS_{mellan} = KVS_{totalt} - KVS_{inom} = 5.848.$$

Motsvarande F-kvot ges av

$$\frac{5.848/(3-1)}{10.5241/(30-3)} = 7.502 > F_{0.05}(2, 27) = 3.35413.$$

Således förkastas hypotesen motsvarande att lamporna kan anses likvärdiga gällande livslängd.

Uppgift 4

Låt $\boldsymbol{c}$ vara en godtycklig tvådimensionell vektor. Med hjälp av lämpliga resultat ifrån formelbladet erhålls enkelt

$$\boldsymbol{c}^T\widehat{\boldsymbol{\theta}} \sim N(\boldsymbol{c}^T\boldsymbol{\theta}, \sigma^2\boldsymbol{c}^T(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{c}).$$

(A) Med $\boldsymbol{c} = (1,0)^T$ och första resultatet ovan får vi

$$\widehat{\theta}_1 \sim N(\theta_1, \sigma^2(\boldsymbol{A}^T\boldsymbol{A})^{-1}_{11}).$$

(B) och **(C)**: Vi ansätter $\boldsymbol{c} = (1,1)^T$ och erhåller med hjälp av resultaten ovan fördelningen

$$\widehat{\theta}_1 + \widehat{\theta}_2 \sim N(\theta_1 + \theta_2, \sigma^2(1,1)(\boldsymbol{A})^{-1}(1,1)^T).$$

(där uttycket $(1,1)(\boldsymbol{A})^{-1}(1,1)^T$ enkelt kan förenklas). Med grundläggande sannolikhetsteoretiska argument erhålls konfidensintervallet

$$\widehat{\theta}_1 + \widehat{\theta}_2 \pm z_{p/2}\sigma\sqrt{(1,1)(\boldsymbol{A}^T\boldsymbol{A})^{-1}(1,1)^T}.$$

Uppgift 5

En allmän MA(1)-tidsserie kan definieras enligt

$$Y_t = \epsilon_t + \theta\epsilon_{t-1},$$

med $\epsilon_t \sim N(0, \sigma^2)$ iid.

Detta är en kovarians-stationär tidsserie (jmfr. Sundbergs kompendium kapitel 6) oavsett värde på θ, vilket inses med hjälp av följande observationer

$$E(Y_t) = E(\epsilon_t + \theta\epsilon_{t-1}) = 0,$$
$$Var(Y_t) = \sigma^2 + \theta^2\sigma^2$$

och

$$Cov(Y_t, Y_{t+k}) = \begin{cases} \theta\sigma^2, & \text{givet } k = 1 \\ 0, & \text{givet } k > 1. \end{cases}$$