

Lösningar till tentamensskrivning för kursen  
Linjära statistiska modeller

30 november 2023 8–13

*Examinator:* Ola Hössjer, tel. 070/672 12 18, ola@math.su.se

---

**Uppgift 1**

a) Antalet frihetsgrader för Regression och Residual är 1 respektive  $20-2=18$ .  
Det följer att

$$F\text{-kvot} = \frac{\text{Mkvs(Regression)}}{\text{Mkvs(Residual)}} = \frac{\text{Kvs(Regression)}}{\text{Kvs(Residual)}/18} = \frac{5.2}{19.3/18} = 4.85.$$

Eftersom detta värde överstiger  $F_{0.05}(1, 18) = 4.4$ , kan vi på signifikansnivån 5% förkasta nollhypotesen att närhet till butik inte har någon inverkan på kundnöjdhet.

b) Skattningen av feltermernas standardavvikelse ges av

$$\hat{\sigma} = \sqrt{\text{Mkvs(Residual)}} = \sqrt{\frac{\text{Kvs(Residual)}}{18}} = \sqrt{\frac{19.3}{18}} = 1.036. \quad (1)$$

c) Ett 95% konfidensintervall för  $\sigma$  kan skrivas som

$$I_\sigma = \left( \hat{\sigma} \cdot \sqrt{\frac{18}{\chi_{0.025}^2(18)}}, \hat{\sigma} \cdot \sqrt{\frac{18}{\chi_{0.975}^2(18)}} \right) = (0.782, 1.531), \quad (2)$$

där  $\chi_p^2(f)$  är  $1-p$ -kvantilen för en  $\chi^2(f)$ -fördelning. Med hjälp av (1)-(2) kan vi bestämma värdena på de två  $\chi^2$ -kvantiler som ingår i konfidensintervallet för  $\sigma$ . Vi får:

$$\begin{aligned} \chi_{0.975}^2(18) &= 18 \cdot (\hat{\sigma}/1.531)^2 = 8.23, \\ \chi_{0.025}^2(18) &= 18 \cdot (\hat{\sigma}/0.782)^2 = 31.53. \end{aligned}$$

## Uppgift 2

a) Parametervektorn för grundmodellen är  $\boldsymbol{\theta} = (\tilde{\alpha}_1, \tilde{\alpha}_2, \beta_1, \beta_2, \beta_3)^T$ . Hypotesmodellen att det inte finns några systematiska skillnader mellan skadeprijsättningen i de två regionerna, svarar mot nollhypotesen  $H_0 : \tilde{\alpha}_1 = \tilde{\alpha}_2 = \tilde{\alpha}$ . Vi kan formulera det som

$$\boldsymbol{\theta} = \begin{pmatrix} \tilde{\alpha} \\ \tilde{\alpha} \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \tilde{\alpha} \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \mathbf{B}\boldsymbol{\lambda}.$$

b) Antalet frihetsgrader för residualerna är  $24 - 5 = 19$ , och skillnaden i antal parametrar mellan grund- och hypotesmodell är  $k - l = 5 - 4 = 1$ . För att testa nollhypotesen bildar vi därför

$$\text{F-kvot} = \frac{\text{Mkvs(Avv från hyp)}}{\text{Mkvs(Residual)}} = \frac{\text{Kvs(Avv från hyp)}}{\text{Kvs(Residual)/19}} = \frac{1.8}{22.3/19} = 1.53,$$

som är mindre än  $F_{0.05}(1, 19) = 4.4$ . Vi kan därför inte förkasta nollhypotesen.

c) Grundmodellens underrum spänns upp av de fem kolumnerna i designmatrisen

$$\mathbf{A} = (\mathbf{1}_1, \mathbf{1}_2, \mathbf{x}_1 - \bar{x}_1, \mathbf{x}_2 - \bar{x}_2, \mathbf{x}_3 - \bar{x}_3) = (\mathbf{1}_1, \mathbf{1}_2, \mathbf{X}),$$

där  $\mathbf{1}_1 = (1, \dots, 1, 0, \dots, 0)^T$ ,  $\mathbf{1}_2 = (0, \dots, 0, 1, \dots, 1)^T$ ,  $\mathbf{x}_j = (x_{j1}, \dots, x_{j24})^T$ ,  $\bar{x}_j = (\bar{x}_j, \dots, \bar{x}_j)^T$  och  $\mathbf{X}$  är den del av designmatrisen som härrör från de förklarande variablerna. Minsta kvadrat-skattningarna av modellparametrarna ges därför av

$$\begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y} = (\mathbf{A}^T \mathbf{A})^{-1} \begin{pmatrix} \sum_{i=1}^{12} Y_i \\ \sum_{i=13}^{24} Y_i \\ \sum_{i=1}^{24} Y_i (x_{1i} - \bar{x}_1) \\ \sum_{i=1}^{24} Y_i (x_{2i} - \bar{x}_2) \\ \sum_{i=1}^{24} Y_i (x_{3i} - \bar{x}_3) \end{pmatrix}. \quad (3)$$

Även om  $\mathbf{1}_1$  och  $\mathbf{1}_2$  är ortogonala mot varandra, är de i allmänhet *inte* ortogonala mot de tre sista kolumnerna  $\mathbf{x}_1 - \bar{x}_1$ ,  $\mathbf{x}_2 - \bar{x}_2$ ,  $\mathbf{x}_3 - \bar{x}_3$  i designmatrisen. Därför gäller i allmänhet att

$$(\mathbf{A}^T \mathbf{A})^{-1} \neq \begin{pmatrix} 1/12 & 0 & \mathbf{0}^T \\ 0 & 1/12 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{0} & (\mathbf{X}^T \mathbf{X})^{-1} \end{pmatrix}, \quad (4)$$

där  $\mathbf{0} = (0, 0, 0)^T$ . Av detta och ekvation (3) följer att de två givna formlerna för  $\hat{\alpha}_1$  och  $\hat{\alpha}_2$  i allmänhet är fel. Observera att om  $\mathbf{1}_1$  och  $\mathbf{1}_2$  hade varit

ortogonala mot kolumnerna i  $\mathbf{X}$ , så hade högerledet i (4) varit lika med  $(\mathbf{A}^T \mathbf{A})^{-1}$ , och då hade de föreslagna formlerna för  $\hat{\alpha}_1$  och  $\hat{\alpha}_2$  varit korrekta. Hypotesmodellen underrum spänns däremot upp av de fyra kolumnerna i designmatrisen

$$\mathbf{A}_0 = \mathbf{A}\mathbf{B} = (\mathbf{1}, \mathbf{x}_1 - \bar{x}_1, \mathbf{x}_2 - \bar{x}_2, \mathbf{x}_3 - \bar{x}_3),$$

och på grund av definitionen av  $\bar{x}_1$ ,  $\bar{x}_2$  och  $\bar{x}_3$  i problemtexten är  $\mathbf{1} = (1, \dots, 1)^T$  ortogonal mot de tre sista kolumnerna i  $\mathbf{A}_0$ , det vill säga

$$(\mathbf{A}_0^T \mathbf{A}_0)^{-1} = \begin{pmatrix} 1/24 & \mathbf{0}^T \\ \mathbf{0} & (\mathbf{X}^T \mathbf{X})^{-1} \end{pmatrix}.$$

Eftersom  $\hat{\alpha}$  är första komponenten i minsta kvadrat-skattningen  $\hat{\boldsymbol{\lambda}} = (\mathbf{A}_0^T \mathbf{A}_0)^{-1} \mathbf{A}_0^T \mathbf{Y}$  av regressionsparametrarna  $\boldsymbol{\lambda}$  för hypotesmodellen, följer att

$$\hat{\alpha} = [(\mathbf{A}_0^T \mathbf{A}_0)^{-1} \mathbf{A}_0^T \mathbf{Y}]_1 = \frac{1}{24} (\mathbf{A}_0^T \mathbf{Y})_1 = \frac{1}{24} \sum_{i=1}^{24} Y_i.$$

### Uppgift 3

a) Vi börjar med att beräkna

$$\begin{aligned} \gamma_0 &= \text{Var}(X_t) \\ &= \text{Var}(\varepsilon_t - \theta\varepsilon_{t-1}) \\ &= \text{Var}(\varepsilon_t) + \theta^2 \text{Var}(\varepsilon_{t-1}) - 2\theta \text{Cov}(\varepsilon_t, \varepsilon_{t-1}) \\ &= \sigma_\varepsilon^2 + \theta^2 \sigma_\varepsilon^2 - 2\theta \cdot 0 \\ &= (1 + \theta^2) \sigma_\varepsilon^2 \end{aligned}$$

och

$$\begin{aligned} \gamma_1 &= \text{Cov}(X_t, X_{t+1}) \\ &= \text{Cov}(\varepsilon_t - \theta\varepsilon_{t-1}, \varepsilon_{t+1} - \theta\varepsilon_t) \\ &= \text{Cov}(\varepsilon_t, \varepsilon_{t+1}) - \theta \text{Var}(\varepsilon_t) - \theta \text{Cov}(\varepsilon_{t-1}, \varepsilon_{t+1}) + \theta^2 \text{Cov}(\varepsilon_{t-1}, \varepsilon_t) \\ &= 0 - \theta \sigma_\varepsilon^2 - \theta \cdot 0 + \theta^2 \cdot 0 \\ &= -\theta \sigma_\varepsilon^2. \end{aligned}$$

För  $k \geq 2$  gäller att  $X_t = \varepsilon_t - \theta_{t-1} \varepsilon_{t-1}$  och  $X_{t+k} = \varepsilon_{t+k} - \theta \varepsilon_{t+k-1}$  är funktioner av olika feltermerna. Därför är  $X_t$  oberoende av  $X_{t+k}$  då  $k \geq 2$ , vilket medför att  $\gamma_k = 0$  för  $k \geq 2$ .

b) Det följer av a) och stationariteten hos  $\{X_t\}$  att

$$\rho_k = \frac{\text{Cov}(X_t, X_{t+k})}{\sqrt{\text{Var}(X_t)} \sqrt{\text{Var}(X_{t+k})}} = \frac{\gamma_k}{\sqrt{\gamma_0} \sqrt{\gamma_0}} = \frac{\gamma_k}{\gamma_0} = \begin{cases} 1, & k = 0, \\ -\theta/(1 + \theta^2), & k = 1, \\ 0, & k \geq 2. \end{cases}$$

c) Vi börjar med fallet  $k = 1$ . Eftersom  $X_t$  och  $X_{t+k}$  är oberoende då  $k \geq 2$  följer att

$$\hat{X}_{T+1} = E(X_{T+1}|\mathbf{X}_T) = E(X_{T+1}|X_T) = \rho_1 X_T = -\frac{\theta}{1+\theta^2} X_T,$$

där vi i tredje ledet använde oss av ledningen med  $Y = X_T$ ,  $Z = X_{T+1}$ ,  $\text{Var}(Y) = \text{Var}(Z) = \gamma_0 = (1 + \theta^2)\sigma_\varepsilon^2$  och  $\text{Corr}(Y, Z) = \rho_1$ . För  $k \geq 2$  kan vi utnyttja att  $X_{T+k}$  är oberoende av  $\mathbf{X}_T$ , varav följer att

$$\hat{X}_{T+k} = E(X_{T+k}|\mathbf{X}_T) = E(X_{T+k}) = 0.$$

#### Uppgift 4

a) Antalet observationer är  $N = 5$ , och antalet parametrar  $k = 2$ . Designmatrisen

$$\mathbf{A} = \begin{pmatrix} 1.3 & 2.6 \\ 0.7 & 2.0 \\ 1.5 & 2.0 \\ 0.9 & 1.4 \\ 2.1 & 2.9 \end{pmatrix}$$

har dimension  $N \times k$ . Dess två kolumner innehåller värdena på respektive förklarande variabel från alla mätningar. Observera att eftersom modellen saknar intercept har  $\mathbf{A}$  ingen kolumn med bara ettor.

b) Låt  $\mathbf{Y} = (Y_1, \dots, Y_5)^T = (1.6, 1.2, 1.5, 1.2, 1.9)^T$  vara observationsvektorn. Vi använder allmänna formeln för minsta kvadrat-skattningen och ledningen för att räkna ut estimatet

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\hat{\beta}_1, \hat{\beta}_2)^T \\ &= (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y} \\ &= \begin{pmatrix} 9.65 & 15.13 \\ 15.13 & 25.13 \end{pmatrix}^{-1} \begin{pmatrix} 10.24 \\ 16.75 \end{pmatrix} \\ &= \frac{1}{9.65 \cdot 25.13 - 15.13^2} \begin{pmatrix} 25.13 & -15.13 \\ -15.13 & 9.65 \end{pmatrix} \begin{pmatrix} 10.24 \\ 16.75 \end{pmatrix} \\ &= \begin{pmatrix} 0.2873 \\ 0.4936 \end{pmatrix} \end{aligned} \tag{5}$$

av  $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$ .

c) Eftersom antalet frihetsgrader för att skatta feltermens variansen är  $N - k = 5 - 2 = 3$ , följer att

$$\hat{\sigma}^2 = \text{Mkvs}(\text{Residual}) = \frac{\text{Kvs}(\text{Residual})}{3} = \frac{0.0909}{3} = 0.0303.$$

d) Vi ska utnyttja dualiteten mellan konfidensellipsoider och tester, det vill säga att konfidensellipsoiden för  $\boldsymbol{\beta}$  består av mängden av alla  $\boldsymbol{\beta}_0 =$

$(\beta_{10}, \beta_{20})^T$  sådana att nollhypotesen

$$H_0 : \beta = \beta_0$$

inte förkastas på signifikansnivån 5%. Under nollhypotesen gäller att  $\mu = E(\mathbf{Y}) = \mathbf{A}\beta_0$ , som vi vill jämföra med  $\hat{\mu} = \mathbf{A}\hat{\beta}$ . Vi ställer därför upp teststorheten

$$\text{F-kvot} = \frac{\|\hat{\mu} - \mu\|^2/k}{\hat{\sigma}^2} = \frac{\|\mathbf{A}(\hat{\beta} - \beta_0)\|^2}{2 \cdot \hat{\sigma}^2} = (\hat{\beta} - \beta_0)^T \mathbf{C} (\hat{\beta} - \beta_0), \quad (6)$$

för att testa  $H_0$ , med

$$\mathbf{C} = \frac{1}{2 \cdot \hat{\sigma}^2} \mathbf{A}^T \mathbf{A} = \frac{1}{2 \cdot 0.0303} \begin{pmatrix} 9.65 & 15.13 \\ 15.13 & 25.13 \end{pmatrix} = \begin{pmatrix} 159.2 & 249.7 \\ 249.7 & 414.7 \end{pmatrix},$$

och där nollhypotesen förkastas då (6) överstiger  $F_{0.05}(k, N-k) = F_{0.05}(2, 3) = 9.552$ . Genom att kombinera denna ekvation med (5), ser vi att den sökta konfidensellipsoiden ges av

$$E = \{(\beta_{10}, \beta_{20})^T; 159.2(0.287 - \beta_{10})^2 + 2 \cdot 249.7(0.287 - \beta_{10})(0.494 - \beta_{20}) + 414.7 \cdot (0.494 - \beta_{20})^2 < 9.552\}.$$

## Uppgift 5

a) Det totala stickprovsmedelvärdet är

$$\bar{Y}_{..} = \frac{1}{5}(\bar{Y}_{1.} + \dots + \bar{Y}_{5.}) = \frac{1}{5}(3.1 + 5.2 + 4.2 + 4.6 + 3.9) = 4.2.$$

Eftersom antal frihetsgrader för variationskällan mellan brädor är  $5 - 1 = 4$ , ser vi att

$$\begin{aligned} \text{Mkvs(Mellan brädor)} &= \frac{1}{4} \sum_{i=1}^5 \sum_{j=1}^4 (\bar{Y}_{ij} - \bar{Y}_{i.})^2 = \sum_{i=1}^5 (\bar{Y}_{i.} - \bar{Y}_{..})^2 \\ &= (3.1 - 4.2)^2 + (5.2 - 4.2)^2 + (4.2 - 4.2)^2 + (4.6 - 4.2)^2 + (3.9 - 4.2)^2 = 2.46. \end{aligned}$$

b) Vi noterar först att stickprovsvariansen för bräda  $i$  ges av

$$s_i^2 = \frac{1}{3} \sum_{j=1}^4 (Y_{ij} - \bar{Y}_{i.})^2,$$

och att variationskällan inom brädor har  $5(4 - 1) = 15$  frihetsgrader. Det ger

$$\text{Mkvs(Inom brädor)} = \frac{1}{15} \sum_{i=1}^5 \sum_{j=1}^4 (Y_{ij} - \bar{Y}_{i.})^2 = \frac{1}{5} \sum_{i=1}^5 s_i^2 = 0.13.$$

c) Vi ska utnyttja att de två medelkvadratsummorna i a) och b) är observationer av oberoende stokastiska variabler med fördelningar

$$\begin{aligned}\text{Mkvs}(\text{Mellan brädor}) &\sim (4\sigma_\delta^2 + \sigma^2)\chi^2(4)/4, \\ \text{Mkvs}(\text{Inom brädor}) &\sim \sigma^2\chi^2(15)/15,\end{aligned}$$

där faktorerna  $4\sigma_\delta^2 + \sigma^2$  och  $\sigma^2$  är väntevärdena för respektive kvadratsumma (se formelsamlingen). Av detta följer att

$$F\text{-kvot} = \frac{\text{Mkvs}(\text{Mellan brädor})}{\text{Mkvs}(\text{Inom brädor})} \sim \left( \frac{4}{\sigma^2/\sigma_\delta^2} + 1 \right) F(4, 15). \quad (7)$$

Det sökta konfidensintervallet  $I_{\sigma^2/\sigma_\delta^2}$  består av de varianskvoter  $x$  för vilka ett hypotestest baserat på  $F$ -kvoten i (7) inte förkastar nollhypotesen  $H_0 : \sigma^2/\sigma_\delta^2 = x$  gentemot den alternativa hypotesen  $H_1 : \sigma^2/\sigma_\delta^2 < x$  på signifikansnivån 5%. Ett sådant test förkastar *inte* nollhypotesen då  $F$ -kvoten är mindre än  $(4/x+1)F_{0.05}(4, 15) = (4/x+1)3.056$ . Eftersom det observerade värdet på  $F$ -kvoten är  $2.46/0.13 = 18.92$  enligt a) och b), följer att

$$I_{\sigma^2/\sigma_\delta^2} = \left\{ x; 18.92 < \left( \frac{4}{x} + 1 \right) 3.056 \right\} = \left( 0, \frac{4}{\frac{18.92}{3.056} - 1} \right) = (0, 0.77).$$