Kapitel 7

Multivariat linjär regression

7.1 ..

7.2 7.2 Klassisk linjär regression

Respons y, r oberoende variabler $Z_1, \dots, Z_r$

Observation nr i: $(Y_i, Z_{1i}, \dots Z_{ri}), i = 1, 2, \dots, n$. Modell

$$Y_i = \beta_0 + \beta_1 Z_{1i} + \cdots + \beta_r Z_{ri} + \varepsilon_i$$

$i = 1, 2, \dots, n$.

Antaganden

1) $E(\varepsilon_i) = 0$
2) $V(\varepsilon_i) = \sigma^2$ oberoende av i
3) $C(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$.

Inför $\boldsymbol{Z}_i' = (1, Z_{1i}, \dots, Z_{ri})$ och $\boldsymbol{\beta}' = (\beta_0, \beta_1, \dots, \beta_r)$ så att

$$Y_i = \boldsymbol{Z}_i'\boldsymbol{\beta} + \varepsilon_i$$

$i = 1, 2, \dots, n$.

Inför nu $\boldsymbol{Y}' = (Y_1, Y_2, \dots, Y_n)$, $\boldsymbol{\varepsilon}' = (\varepsilon_1, \dots, \varepsilon_n)$

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{Z}_1' \\ \vdots \\ \boldsymbol{Z}_n' \end{pmatrix}$$

så att modellen kan skrivas

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

där

1) $E(\boldsymbol{\varepsilon}) = \boldsymbol{0}$
2 & 3) $C(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\boldsymbol{I}$. ...

7.3 7.3 MK skattning

$\boldsymbol{b}$ skattning av $\boldsymbol{\beta}$. Residualkvadratsumman

$$S(\boldsymbol{b}) = \sum_{j=1}^{n}(y_j - b_0 - b_1 z_{j1} - \cdots - b_r z_{jr})^2 = (\boldsymbol{y} - Z\boldsymbol{b})'(\boldsymbol{y} - Z\boldsymbol{b})$$

7.4 ...

7.5 7.5 Inferens från skattad regressionsfunktion

a) Skattning av punkt på regressionsplanet. Y_0 är värdet på Y då $\boldsymbol{z} = \boldsymbol{z}_0$.

$$E(y_0) = \boldsymbol{z}_0'\boldsymbol{\beta}$$

så en punktskattning ges av $\boldsymbol{Z}_0'\hat{\boldsymbol{\beta}}$. Denna är väntevärdesriktig eftersom $\hat{\boldsymbol{\beta}}$ är väntevärdesriktig.

$$V(\boldsymbol{Z}_0'\hat{\boldsymbol{\beta}}) = \boldsymbol{Z}_0'C(\hat{\boldsymbol{\beta}})\boldsymbol{Z}_0 = \sigma^2\boldsymbol{Z}_0'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}_0$$

vilken kan skattas med

$$s^2\boldsymbol{Z}_0'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}_0$$

Ett konfidensintervall för Y_0 ges av

$$\hat{Y}_0 \pm t_{n-r-1}(\alpha/2)s\sqrt{\boldsymbol{Z}_0'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}_0}$$

b) Prognos av ny observation för $\boldsymbol{z} = \boldsymbol{z}_0$. Prognos av Y_0 ges av $\hat{Y}_0 = \boldsymbol{Z}_0'\hat{\boldsymbol{\beta}}$. Prognosfelet ges av $e_0 = Y_0 - \hat{Y}_0$.

$$V(e_0) = V(Y_0 - \hat{Y}_0) = V(\varepsilon_0 - \hat{Y}_0) = V(\varepsilon_0) + V(\hat{Y}_0) = \sigma^2(1 + \boldsymbol{Z}_0'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}_0)$$

σ^2 skattas med s^2. Ett prognosintervall för Y_0 ges av

$$\hat{Y}_0 \pm t_{n-r-1}(\alpha/2)s\sqrt{1 + \boldsymbol{Z}_0'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}_0}$$

7.6 Residualanalys överhoppas

7.7 7.7 Multivariat multipel regression

Nu flera Y-variabler $Y_1, \ldots, Y_m$ som alla beror av $Z_1, \ldots, Z_r$:

$$Y_{ij} = \beta_{0j} + \beta_{1j}Z_{1j} + \cdots + \beta_{rj}Z_{ri} + \varepsilon_{ij}$$

$i = 1, 2, \ldots, n, j = 1, 2, \ldots, m$.

Inför

$$\boldsymbol{Y} = (Y_{ij}) = \begin{pmatrix} Y_{11} & \ldots & Y_{1m} \\ \vdots & & \vdots \\ Y_{n1} & \ldots & Y_{nm} \end{pmatrix}$$

$$\boldsymbol{\beta} = (Y_{ij}) = \begin{pmatrix} \beta_{01} & \ldots & \beta_{0m} \\ \beta_{11} & \ldots & \beta_{1m} \\ \vdots & & \vdots \\ \beta_{r1} & \ldots & \beta_{rm} \end{pmatrix}$$

.....

$$\boldsymbol{Y} = \boldsymbol{Z\beta} + \boldsymbol{\varepsilon}$$

....

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}$$

7.8 7.8 Begreppet linjär regression

7.8.1 Univariat

$(Y, Z_1, \ldots, Z_r)'$ antas multivariat normalfördelad med väntevärdesvektor $\boldsymbol{\mu}$ och kovariansmatris Σ. Dessa kan partitioneras i

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_Y \\ \boldsymbol{\mu}_{\boldsymbol{Z}} \end{pmatrix}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{YY} & \boldsymbol{\sigma}'_{ZY} \\ \boldsymbol{\sigma}_{ZY} & \Sigma_{ZZ} \end{pmatrix}$$

Result 7.12: Den bästa linjära prediktorn av Y ges av $\hat{Y} = \beta_0 + \boldsymbol{\beta}'\boldsymbol{Z}$ där $\boldsymbol{\beta} = \Sigma_{ZZ}^{-1}\boldsymbol{\sigma}_{ZY}$ och $\beta_0 = \mu_Y - \boldsymbol{\beta}'\boldsymbol{\mu}_Z$.

7.8.2 Multivariat

Kan generaliseras till multivariat regression. $(\boldsymbol{Y}, \boldsymbol{Z})'$ normalfördelad. $\boldsymbol{\beta} = \Sigma_{YZ}\Sigma_{ZZ}^{-1}$

$\hat{Y} = \boldsymbol{\mu}_Y + \boldsymbol{\beta}(\boldsymbol{Z} - \boldsymbol{\mu}_Z)$.

Fel-vektorn i prediktionen $\boldsymbol{Y} - \hat{\boldsymbol{Y}}$ har kovariansmatris

$$\Sigma_{YY.Z} = \Sigma_{YY} - \Sigma_{YZ}\Sigma_{ZZ}^{-1}\Sigma'_{YZ}.$$

....

Partiell korrelation

Korrelationskoefficienten ...

Det gäller att

$$\rho_{Y_1Y_2.Z} = \frac{\rho_{Y_1Y_2} - \rho_{Y_1Z}\rho_{Y_2Z}}{\sqrt{(1-\rho_{Y_1Z}^2)(1-\rho_{Y_2Z}^2)}}$$

7.9 7.9 Läses kursivt

7.10 7.10 Pathanalys (Stiganalys)