

6.3.4 Konfidenz und Vorhersage Intervalle

Da \hat{b} und \hat{a} normalverteilt sind, können wir KIs für a und b mit der t -Verteilung erstellen. Für b :

$$\hat{b} \pm t_{(n-2), 1-\frac{\alpha}{2}} \frac{s}{\sqrt{\sum x_i^2 - n\bar{x}^2}}$$

Um unsere Unsicherheit um des Modells auszudrücken, brauchen wir die Varianz von $\hat{a} + \hat{b}x_0$:

$$\begin{aligned} V[\hat{a} + \hat{b}x_0] &= V[\hat{a}] + x_0^2 V[\hat{b}] + 2x_0 \text{Cov}[\hat{a}, \hat{b}] \\ &= s^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x_i^2 - n\bar{x}^2} \right) \end{aligned}$$

$\frac{1}{n}$ kommt von der Stichprobengröße, $\frac{(x_0 - \bar{x})^2}{\sum x_i^2 - n\bar{x}^2}$ hängt von der Entfernung von x_0 von \bar{x} ab.

Ein $(100 - \alpha)\%$ KI für $\hat{a} + \hat{b}x_0 = E[Y|X = x_0]$:

$$(\hat{a} + \hat{b}x_0) \pm t_{(n-2), 1-\frac{\alpha}{2}} s \sqrt{\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x_i^2 - n\bar{x}^2} \right)}$$

Ein $(100 - \alpha)\%$ Vorhersage Intervall für $Y|X = x_0$ ist:

$$(\hat{a} + \hat{b}x_0) \pm t_{(n-2), 1-\frac{\alpha}{2}} s \sqrt{\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x_i^2 - n\bar{x}^2} \right)}$$

6.4 Residuen

$$\begin{aligned}e_i &= y_i - (\hat{a} + \hat{b}x_i) \\ &= y_i - \hat{y}_i\end{aligned}$$

$\{e_i\}$ sind nicht eine Stichprobe aus $N(0, \sigma^2)$, weil wir a und b nur geschätzt haben. Wenn wir a und b genau wußten, wären $e_i = y_i - (a + bx_i)$ eine Stichprobe aus $N(0, \sigma^2)$.

Die Residuen werden benutzt, um die Güte des Modells eingehender zu überprüfen. Die folgenden Graphiken sind hilfreich:

- (1) e_i v. \hat{y}_i Residuen v. Vorhergesagte
- (2) e_i v. i Residuen v. Reihenfolge
- (3) e_i v. w_i Residuen v. eine andere erklärende Variable
- (4) Histogramm/Boxplot/Dotplot von e_i

6.5 Multiple Lineare Regression

Um die Wirkungen von mehreren erklärenden Variablen zu untersuchen:

$$Y = b_0 + \sum_{j=1}^p b_j X_j + \epsilon$$

Das Modell wird als Matrixgleichung geschrieben:

$$Y = Xb + \epsilon$$

n ist die Anzahl Beobachtungen

p ist die Anzahl erklärender Variablen

Y ($n \times 1$) ist die abhängige Variable

$\{X_j\}$ ($n \times (p + 1)$) sind die erklärenden Variablen

b ($(p + 1) \times 1$) ist der Koeffizientenvektor

ϵ ($n \times 1$) \sim u.i.v. $N(0, \sigma^2)$ ist der Fehlerterm

6.5.1 KQ Schätzung von b

$$\min C = \sum \epsilon_i^2 = \min \epsilon' \epsilon = \min (Y - Xb)'(Y - Xb)$$

$$\frac{dC}{db} = 0 \Rightarrow -X'Y + X'Xb = 0$$

$$\Rightarrow \hat{b} = (X'X)^{-1}X'Y$$

$$V[\hat{b}] = V[(X'X)^{-1}X'Y]$$

$$= (X'X)^{-1}X'((X'X)^{-1}X')'V[Y]$$

$$= (X'X)^{-1}\sigma^2$$

weil $V[Y] = V[\epsilon] = \sigma^2 I$

Als Beispiel kann man die einfache lineare Regression in dieser allgemeinen Form herleiten. N.B. für $p = 1$ gilt

$$(X'X)^{-1} = \frac{1}{n(\sum x_i^2 - n\bar{x}^2)} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix}$$

6.5.2 ML Schätzer = KQ Schätzer in Regression

Mit dem Modell

$$Y = Xb + \epsilon$$

wo $\epsilon \sim N(0, \sigma^2 I)$, nehmen wir an, dass

$$Y_i \sim N\left(b_0 + \sum_{j=1}^p b_j X_{ij}, \sigma^2\right) \quad \text{u.i.v.}$$

$$L(y, x; b, \sigma^2) = \prod_{i=1}^n \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2\sigma^2} \left(y_i - b_0 - \sum_{j=1}^p b_j x_{ij}\right)^2}$$

$$l = \log L$$

$$= -n \log \sigma - \frac{n}{2} \log 2\pi - \frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - b_0 - \sum_{j=1}^p b_j x_{ij}\right)^2$$

Für σ bekannt wird das Maximum von l beim Minimum von $\epsilon' \epsilon$ erreicht.

6.5.3 Gütekriterien

$$R^2 = \frac{\text{erklärte (Modell) Variabilität}}{\text{Gesamt-Variabilität}} = \frac{MSS}{TSS}$$
$$= \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

Je mehr Variablen einbezogen werden, desto größer wird R^2 . Ein Vorschlag ist das adjustierte R^2 :

$$R^2 = 1 - \frac{RSS}{TSS}$$

wo RSS = Rest-Variabilität

$$R_{adj}^2 = 1 - \frac{RSS/(n - p - 1)}{TSS/(n - 1)}$$

$RSS/(n - p - 1) = s^2$ schätzt σ^2

$TSS/(n - 1)$ schätzt σ_Y^2

$$R_{adj}^2 = 1 - \frac{(n - 1)}{(n - p - 1)}(1 - R^2)$$

Beim kleinen R^2 kann R_{adj}^2 sogar < 0 sein, aber man sollte Modelle mit solchen Werten sowieso nicht in Betracht ziehen. Bei größeren Datensätzen ist $R_{adj}^2 \approx R^2$.