ASMNW - Übung 4

Peter von Rohr

2016-05-09

Kontrollfrage 1

Eine Voraussetzung für die Verwendung von Least Squares zur Schätzung der Parameter ist, dass die Matrix $\mathbf{X}$ vollen Kolonnenrang hat. Welche Beziehung resultiert aus dieser Voraussetzung für die Beziehung zwischen den Anzahl Parametern p und die Anzahl Beobachtungen n?

Kontrollfrage 2

Abgesehen von Kolonnenrang, wie lauten die fünf weiteren Bedingungen für das Verwenden von multiplen linearen Regressionen

Kontrollfrage 3

Falls die Bedingung der konstanten Varianz nicht erfüllt ist, hatten wir gezeigt, dass **Generalised Least Squares** verwendet werden kann. Bei generalised least square nehmen wir an, dass

$$var(\epsilon) = \Sigma$$

Diese Co-Varianzmatrix Σ wird mit der Cholesky-Zerlegung und das Produkt $\Sigma = \mathbf{C}\mathbf{C}^T$ zerlegt. Die Zielgrössen $\mathbf{y}$ und die erklärenden Variablen in $\mathbf{X}$ werden mit der Matrix C^{-1} transformiert. Daraus resultieren dann

$$\begin{aligned} \tilde{\mathbf{y}} &= \mathbf{C}^{-1}\mathbf{y} \\ \tilde{\mathbf{X}} &= \mathbf{C}^{-1}\mathbf{X} \end{aligned} \tag{1}$$

Gegeben unser ursprüngliches lineares Modell

$$\mathbf{y} = \mathbf{X}\beta + \epsilon \tag{2}$$

Wie sieht die Beziehung zwischen den Grössen $\tilde{\mathbf{y}}$ und $\tilde{\mathbf{X}}$ aus, insbesondere handelt es sich bei dieser Beziehung wieder um ein lineares Modell und wenn ja, wie sieht dieses aus?

Aufgabe 1

Für unseren Datensatz aus der Vorlesung mit den Zunahmen und dem Geburtsgewicht (siehe http://charlotte-ngs.github.io/GELASM/w12/gain_data.csv) nehmen wir an, dass die Reste ϵ nicht mehr konstant und unkorreliert sind, sondern dass folgende Covarianzmatrix gefunden wurde.

$$\Sigma = \left[\begin{array}{ccccc} 1.44 & 0.12 & 0.08 & 0.09 & 0.12 \\ 0.12 & 2.26 & 0.12 & 0.10 & 0.09 \\ 0.08 & 0.12 & 3.25 & 0.18 & 0.16 \\ 0.09 & 0.10 & 0.18 & 2.91 & 0.14 \\ 0.12 & 0.09 & 0.16 & 0.14 & 2.27 \end{array}\right]$$

Die Matrix ist verfügbar als CSV-Datei unter: http://charlotte-ngs.github.io/GELASM/w12/covar_sigma.csv verfügbar.

Ihre Aufgabe

Schätzen sie die Regressionskoeffizienten unter Berücksichtigung der Covarianzmatrix Σ.

Hinweise

Die Funktion `chol()` macht die Cholesky-Faktorisierung in R. Das Resultat von `chol()` ist eine obere rechte Dreiecksmatrix. Dies entspricht der Matrix $\mathbf{C}^T$ in der Vorlesung. Die Inverse einer Matrix lässt sich mit der Funktion `solve()` berechnen.