

Formelsammlung

Gegeben sei ein multivariates lineares Regressionsmodell $y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + u_i$ für $i = 1, \dots, n$, welches die Annahmen MLR1-MLR6 erfüllt. Es sei $\hat{\beta} = (\hat{\beta}_0, \dots, \hat{\beta}_k)$ der MKQ-Schätzer für $\beta = (\beta_0, \dots, \beta_k)$. Weiterhin sei X die Designmatrix des Modells, SSR die Summe der quadrierten Residuen, R^2 das Bestimmtheitsmaß und s^2 der Schätzer für die Varianz der Residuen.

- Für alle $j = 0, \dots, k$ gilt, dass

$$\frac{\hat{\beta}_j - \beta_j}{s \sqrt{x^{jj}}} \sim t_{n-k-1},$$

wobei x^{jj} der $j + 1$ -te Diagonaleintrag der Matrix $(X^T X)^{-1}$ ist.

- Es sei r ein Vektor mit $k + 1$ Einträgen. Dann gilt, dass

$$\frac{r^T \hat{\beta} - r^T \beta}{s \sqrt{r^T (X^T X)^{-1} r}} \sim t_{n-k-1}.$$

- Sei jetzt zusätzlich das eingeschränkte Modell $y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{k-q} x_{i(k-q)} + u_i$ für $i = 1, \dots, n$ mit $q \leq k$, Bestimmtheitsmaß R_r^2 und Summe der quadrierten Residuen SSR_r gegeben. Dann gilt, dass

$$\frac{(SSR_r - SSR)(n - k - 1)}{q SSR} = \frac{(R^2 - R_r^2)(n - k - 1)}{q(1 - R^2)} \sim F_{q, n-k-1}.$$

Insbesondere ist

$$\frac{R^2(n - k - 1)}{k(1 - R^2)} \sim F_{k, n-k-1}.$$

- Es sei R eine Matrix mit q Zeilen und $k + 1$ Spalten und c ein Vektor mit q Einträgen. Dann gilt, dass

$$\frac{(R\hat{\beta} - c)^T (R(X^T X)^{-1} R^T)^{-1} (R\hat{\beta} - c)}{qs^2} \sim F_{q, n-k-1}.$$

- Nimm an, dass sich die Daten, mit denen das obige lineare Modell geschätzt wird, in 2 Klassen (sagen wir a und b) aufteilen lassen und diese Unterscheidung nicht in Form einer Dummyvariablen im obigen Modell auftaucht. Die Regressionsparameter des linearen Modells können separat nur unter Nutzung der Daten der Klasse a bzw. der Klasse b geschätzt werden, die resultierenden Summen der quadrierten Residuen werden mit SSR_a und SSR_b bezeichnet. Dann gilt, dass

$$\frac{(SSR - SSR_a - SSR_b)(n - 2(k + 1))}{(SSR_a + SSR_b)(k + 1)} \sim F_{k+1, n-2(k+1)}.$$

Alle Verteilungen gelten auch, wenn eine gepoolte MKQ-Schätzung auf Paneldatenmodelle angewandt wird.