### 5.2.2 Tests bezüglich der Modellparameter

Nachdem die Parameter geschätzt wurden, ist von Interesse, ob die Komponenten des Parametervektors $\vec{\beta}$ signifikant von Null verschieden sind. Das heißt, wir testen die

Nullhypothese: $\beta_i = 0$

gegen die

Alternativhypothese: $\beta_i \neq 0$.

Kann bei der i-ten Komponente ($\beta_i$) die Nullhypothese verworfen werden, so hat die i-te unabhängige Variable einen signifikanten Einfluss auf die abhängige Variable (hier sind nur die Steigungsparameter von Interesse, d.h. i > 0). Allgemein ist $\beta_0$ der Achsenabschnitt, die Komponente $\beta_1$ ist die Steigung bezüglich der ersten unabhängigen Variablen im Modell (im einfachen linearen Fall ist diese die einzige), $\beta_2$ ist die Steigung bezüglich der zweiten unabhängigen Variablen, u.s.w.. Die Variablen, die keinen signifikanten Einfluss haben, können unter Umständen aus der Modellgleichung gestrichen werden (bei einer multiplen Regression). Für weitere Untersuchungen und zur Bestätigung müsste dann eine weitere Regressionsanalyse ohne diese Variablen und mit neuen Daten durchgeführt werden.

Für diesen Test verwendet man die Quadratsummen der Varianzzerlegung. Es sei im Folgenden sei r = k + 1 (r ist hier nicht der Korrelationskoeffizient (!)). Es gilt:

SSE/(n - r) ist ein Schätzer für die Varianz der Fehler $\sigma^2$, also

$$\hat{\sigma}^2 = \mathrm{SSE}/(n - r)\ .$$

Im Beispiel ist $\hat{\sigma}^2 = 35{,}12639\ldots$ .

Sei $\vec{B}$ der Zufallsvektor, dessen Realisierung $\hat{\vec{\beta}}$ ist, dann gilt

$$E(\vec{B}) = \vec{\beta} \text{ und } Var(\vec{B}) = \sigma^2 \left(X^T X\right)^{-1} .$$

Somit kann man Prüfgrößen für den obigen Test bezüglich der Modellparameter definieren:

$$t_i = \frac{\hat{\beta}_i - 0}{\sqrt{a_{ii}}} ,$$

wobei $a_{ii}$ das i-te (i = 0, 1, …, k) Hauptdiagonalelement der Matrix

$$A = \hat{\sigma}^2 \left(X^T X\right)^{-1}$$

ist, einer Schätzmatrix für $Var(\vec{B})$. Die Indizierung beginnt also bei 0, analog der Indizierung der Komponenten von $\vec{B}$ und $\vec{\beta}$.

Dabei sind (wie immer unter $H_0$) die $t_i$ Realisierungen von t-verteilten Zufallsvariablen mit n - r Freiheitsgraden. Also kann $H_0$ auf einem Signifikanzniveau von $\alpha$ verworfen werden, wenn gilt:

$$F_{t_{n-r}}(|\, t_i \,|) \geq 1 - \alpha / 2 \Leftrightarrow \alpha \geq 2(1 - F_{t_{n-r}}(|\, t_i \,|)) = p - Wert_i$$

Im Beispiel ist demnach der Parameter $\beta_1$ (d.h. die Steigung der Regressionsgeraden) signifikant von Null verschieden (0,0199 ≤ 0,05), während der Parameter $\beta_0$ (der Achsenabschnitt der Regressionsgeraden) nicht signifikant von Null verschieden ist (0,1004 > 0,05). Da die Steigung signifikant ungleich Null ist, konnte ein Einfluss der Körpergröße (v2) auf das Körpergewicht (v1) nachgewiesen werden.