Test der Regressionskoeffizienten, der p-Value

Zur Bestimmung, wie signifikant ein Faktor ist, wird häufig der sogenannte p-Value genannt. Zunächst wird die Hypothese aufgestellt, dass der Koeffizient des betrachteten Faktors aus der multiplen Regression b=0 ist. Der p-Value ist dann die Wahrscheinlichkeit diese Hypothese irrtümlich abzulehnen. Die Wahrscheinlichkeit bestimmt sich allgemein aus dem Verteilungswert der t-Verteilung mit:

$$t = \frac{b}{s_b}$$

b = Koeffizient aus multipler Regression
s_b = Streuung des Koeffizienten

und dem doppelten Wert der Studentverteilung von t mit dem Freiheitsgrad
$f = n - z - 1$ (n = Anzahl Versuche, z = Anzahl Modellterme x_1, x_2, x_3, $x_1 \cdot x_2$, x_1^2 usw.). Der doppelte Wert resultiert aus dem zweiseitigen Test. Mit j als Index für den jeweiligen Faktor schreibt man:

$$t_j = \frac{b_j}{s_{b\,j}}$$

Die Streuung des Regressionskoeffizienten wird bestimmt durch:

$$s_{b\,j} = \sqrt{s^2 X''_{j,j}}$$

wobei s die Standardabweichung des Gesamtmodells ist, die sich aus den Summenquadraten der Abweichungen zwischen gemessenen Zielwerten zu den Modellwerten berechnet:

$$s^2 = \frac{1}{n - z - 1} \sum_{i=1}^{n} \left(Y_i - b_o - \sum_{j=1}^{z} x_{j,i}\, b_j \right)^2$$

mit b_o = konstantes Glied des Modells.

X'' berechnet sich durch:

$$X'' = (X^T X)^{-1} \quad mit \quad X = \begin{bmatrix} 1 & x_{11} & .. & x_{z1} \\ 1 & x_{12} & .. & x_{z2} \\ .. & .. & .. & .. \\ 1 & x_{1n} & .. & x_{zn} \end{bmatrix}$$

Je größer der t-Wert ist, desto kleine wird der p-Value. Unter der zugrundeliegenden Nullhypothese $b_j = 0$ ist ein Signifikanzniveau festzulegen, das in der Regel bei 5% liegt. D.h. liegt der p-Value unter 5%, so ist der betrachtete Faktor signifikant.