Unter den erwähnten Bedingungen kommen gemäss *Andress u.a.* als multivariate Analysemethoden explizit nur die latente Klassenanalyse, log-lineare Modelle, der GSK-Ansatz und die logistische Regression in Frage.[392] Die logistische Regression hat dabei den Vorteil, dass sie sowohl für Individualdaten, d.h. unabhängige Variablen mit vielen Ausprägungen, als auch für tabellierte Daten geeignet ist, während die anderen Modelle auf kategoriale Daten beschränkt sind.[393] Die multivariaten Analysen führen wir daher mit logistischen Regressionen durch.

Erklärungskraft der multivariaten Modelle

Mit der logistischen Regression wird der Einfluss mehrerer unabhängiger Variablen auf eine dichotomisierte, abhängige Variable geschätzt. Der partielle Regressions-Koeffizient R gibt Aufschluss über die relative Bedeutung der einzelnen unabhängigen Variablen im Modell. Je grösser R, desto grösser ist der Erklärungsanteil der entsprechenden unabhängigen Variablen im kausalen Modell. Für das Modell insgesamt kann ein Bestimmtheitsmass angewendet werden, das in Anlehnung an das R^2 im Modell der linearen Regression als Pseudo-R^2 bezeichnet wird.[394] Der Wert von Pseudo-R^2 bewegt sich zwischen 0 und 1. Je höher der Wert von Pseudo-R^2, desto grösser ist die prädiktive Wirkung aller unabhängigen Variablen im Modell. Beträgt der Wert von Pseudo-$R^2 = 0$, dann erklären die unabhängigen Variablen die abhängige Variable überhaupt nicht. Ein Wert von $< 0{,}05$ (5 Prozent) weist auf einen geringen kausalen

---

392 *Andress* u.a. 1997. Der Name GSK geht zurück auf die Entwickler dieses Ansatzes (Grizzle, Starmer und Koch). Der Ansatz ist auch bekannt unter dem Namen Minimum-Chi-Quadrat-Methode oder gewichtete Regression. Der GSK-Ansatz basiert auf einer Anwendung des allgemeinen linearen Modells auf kategoriale Daten (*Andress* u.a. 1997: 55).

393 *Andress* u.a. 1997: 19 ff.

394 Pseudo-R^2 wird in der Literatur auch als Likelihood-Ratio-Index oder relative Devianzreduktion bezeichnet. Es gibt verschiedene Vorschläge, wie dieser Wert berechnet werden kann. In dieser Arbeit gelangt die folgende Formel zur Anwendung: «1 – (Log-Likelihood-Funktion des aktuellen Modells/Log-Likelihood-Funktion des konstanten Modells)». Im «aktuellen» Modell werden die Informationen der unabhängigen Variablen verwertet, während im «konstanten» Modell dieser Informationsgehalt ausgeblendet wird. Wenn die beiden Log-Likelihood-Funktionen den gleichen Wert aufweisen, bedeutet dies, dass der Vorhersagewert trotz Berücksichtigung der unabhängigen Variablen nicht verbessert werden konnte. Die Formel ergibt dann «1–1 = 0», d.h. die unabhängigen Variablen haben in diesem Fall keinen Erklärungswert. Vgl. ausführlicher *Andress* u.a. 1997: 261 ff.