

vorliegt: Ein hoher Schwierigkeitsindex p entspricht einer geringen Schwierigkeit der Meßvariablen und umgekehrt. Ist die Testskala umgekehrt gerichtet, dann entsprechen hohe p -Werte auch einer hohen Schwierigkeit der Meßvariablen. Für eine Meßvariable x_i wird der Schwierigkeitsindex p_i aus den Meßergebnissen einer Stichprobe geschätzt:

$$p_i = \frac{1}{N} \sum_{j=1}^N x_{ij}$$

wenn N den Stichprobenumfang bezeichnet. Für dichotome Meßvariable mit den Antwortalternativen 0 und 1 ist es üblich, diesen Schwierigkeitsindex p_i in Form eines Prozentwertes p_i ($\% = p_i \cdot 100\%$) anzugeben, der dann die Prozentzahl derjenigen Meßobjekte kennzeichnet, die das Meßergebnis (die Antwortalternative) 1 erzeugt haben.

Der Kennwert p_i ist im Rahmen der klassischen Testtheorie eine wichtige Kenngröße für die f Itemselektion.

Itemselektion: Begriff der psychologischen Testtheorie (t Testentwicklung), der das Problem bezeichnet, aus einer größeren Gesamtmenge von n Meßvariablen eine Teilmenge von n_i Variablen auszuwählen, die die latente Dimension bzw. Eigenschaft oder Fähigkeit ebenfalls hinreichend erfaßt. Damit geht es inhaltlich darum, theoretische Verfahren und Überlegungen anzugeben, die es gestatten, aus einer Variablenmenge eine Teilmenge so auszuwählen, daß, bezogen auf die experimentelle Fragestellung, der *Phänomenbereich adäquat* und *aussagekräftig* erfaßt wird. Damit wird die I. ein wesentliches Teilproblem bei der Konstruktion von Tests (t Testentwicklung). Im testtheoretischen Sinne ergeben diese Anforderungen für die I., daß, bezogen auf ein Kriterium der latenten Dimension, die ausgewählten n_i Meßvariablen eine maximal hohe *multiple Korrelation* mit einem geeigneten Kriterium ergeben, z. B. mit dem Gesamtwert oder mit einem Außenkriterium. Dabei erfolgt die Auswahl sowohl auf der Basis der wahren Korrelationen als auch unter Berücksichtigung des Stichprobenfehlers.

Da dieser Problemkomplex statistisch nicht exakt gelöst ist, wird ein empirischer Zugang zur Selektion notwendig. Dafür wird häufig folgende Vorgehensweise gewählt: die Selektion von Meßvariablen auf der Grundlage einer Stichprobe und die Abschätzung der multiplen Korrelation der ausgewählten Meßvariablen mit Hilfe einer Eichstichprobe. Zur Auswahl der Meßvariablen sind Ausschöpfungsprozeduren üblich, die sich als Vorwärts-Selektion und Rückwärts-Elimination unterscheiden lassen: 1) Die *Vorwärts-Selektionsprozeduren* (WHERRY, DWYER, SUMMERFIELD und LUBIN) realisieren eine sequentielle Selektion der Meßvariablen in der Weise, daß in jedem Schritt

diejenige Meßvariable ausgewählt wird, die den größten Zuwachs zur Validität mit den bisherigen Meßvariablen ergibt. Dies entspricht jeweils der maximalen partiellen Korrelation. 2) Die *Rückwärts-Eliminationsprozeduren* (HORST, MAC- EWAN) beginnen mit der Gesamtmenge der n Meßvariablen und eliminieren dann sukzessive Meßvariable, so daß die Verringerung in der multiplen Korrelation in jedem Schritt minimiert wird.

Für praktische Fragestellungen erfolgt die I. häufig über die Itemparameter \hat{I} Trennschärfe und J Itemschwierigkeit. Dabei wird versucht, eine hohe Trennschärfe zu erreichen und gleichzeitig eine der Fragestellung angepaßte Verteilung der Schwierigkeitsindizes zu sichern.

Itemvalidität: Begriff der psychologischen \hat{I} Testtheorie, der den allgemeinen Güteparameter f Validität eines Testes bezüglich eines externen Kriteriums Y auf die einzelnen Meßvariablen oder Items eines Tests spezifiziert. Besteht ein Gesamtest X aus der Menge x_1, \dots, x_m einzelner Meßvariablen, dann gibt die Validität von X bezüglich Y über ein korrelatives Zusammenhangsmaß die inhaltliche Übereinstimmung des Testes X mit dem Außenkriterium an. Analog wird die I. eines einzelnen Items x_k bezüglich des Validitätskriteriums Y als korrelativer Zusammenhang $p(x_k, Y)$ definiert und kennzeichnet die inhaltliche Übereinstimmung der einzelnen Meßvariablen mit dem Außenkriterium. Dann gilt folgender Zusammenhang der I. en mit der Validität des Gesamttestes:

$$p(X, Y) = \frac{1}{n} \sum_{k=1}^m r_{kY}$$

$$\text{wobei } r_{kY} = \frac{2}{m} \sum_{k=1}^m r_{kY} Q(x_k, Y)$$

und Q ein Produkt-Moment-Korrelationskoeffizient ist. Dabei bezeichnet $g(x_k, Y)$ die I. des Items x_k bezüglich Y und gibt die inhaltliche Übereinstimmung mit Y an. Da sich die Validität des Gesamttestes X bezüglich Y aus den einzelnen I. en der Meßvariablen zusammensetzt, folgt für die \hat{I} Itemselektion, daß man durch die Auswahl hochvalider Items einen hochvaliden Test zusammenstellen kann (immer bezogen auf das gleiche Validitätskriterium Y). In praktischen Untersuchungen ist es häufig möglich und auf Grund der Itemselektion sinnvoll, die I. durch die Methode der Kreuzvalidierung zu schätzen. Dabei wird i. allg. nach der Itemselektion und der Bestimmung der Regressionsgewichte, die anhand der Daten einer Stichprobe gewonnen wurden, durch die Verwendung einer neuen, zweiten Stichprobe eine Validität dadurch geschätzt, daß die Aussagen und j Prognosen bzw. \setminus Prädiktionen aus der ersten Stichprobe auf die Gültigkeit bezüglich der zweiten Stichprobe überprüft werden.