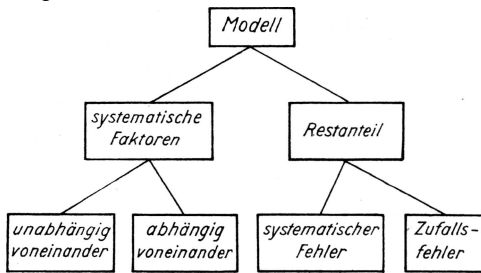


ist es möglich, den Parallelitätsbegriff in unterschiedliche Richtungen auszulegen und damit zu widersprüchlichen Reliabilitätskonzepten zu kommen (GUTTMAN).

Eine Überwindung dieser kritischen Einwände führte zu den Ansätzen der *Verallgemeinerbarkeit von Meßergebnissen*, die von einer Menge nominalparalleler, d. h. inhaltlich gleichartiger Meßvariabler, ausgehen, dem Testkontinuum, und verallgemeinerte Aussagen bezüglich dieses Testkontinuums anstreben (CRONBACH u. a.).

3. Die *Populationsabhängigkeit der Aussagen*. Sie soll mit den Ansätzen der logistischen  $j$  Testmodelle teilweise überwunden werden (BIRNBAUM, RASCH).

Eine andere Klasse von Erweiterungen der klassischen Testtheorie sind die *stochastischen Testmodelle* (LORD, NOVICK, LAZARSELD). Sie haben das Ziel, die statistischen Zusammenhänge zwischen Meßergebnissen und ihren Entstehungsursachen schärfer zu fassen und dadurch tiefere Einblicke in den Phänomenbereich zu erhalten, die dann gleichzeitig zu einer besser gesicherten Interpretation der Ergebnisse führen. Hierzu gehören als erschöpfende Statistiken die logistischen Testmodelle, speziell die latente Strukturanalyse und die Faktorenanalyse. Für alle Ansätze der p. T. gilt, daß ihre Ergebnisse gleichzeitig Ansatzpunkte für die Testkonstruktion liefern. Dabei kommt der  $\hat{I}$  Itemselektion besondere Bedeutung zu. Außerdem lassen sich die Modellansätze danach unterscheiden, welche Wirkanteile und Ursachenkomplexe sie zur theoretischen Erklärung benutzen. Eine grobe Klassifikation verdeutlicht die Abbildung.



**psychologische Testtheorie: Klassifikation zur Unterscheidung der Modellansätze**

Bezogen auf die im Modellansatz verknüpften Faktoren lassen sich die testtheoretischen Ansätze noch danach unterscheiden, welche Parameter sie bezüglich der Meßobjekte und Meßvariablen als Erklärungsansatz enthalten. In dieser Klassifikation ist z. B. die klassische Testtheorie ein Modellansatz mit nur einem systematischen Faktor, der den wahren Testpunktwert angibt und unabhängig vom Restanteil (Fehler) wirkt. Dagegen ist die Faktorenanalyse ein äußerst differenzierter Modellansatz.

Testverfahren, statistische: Verfahren zur Prüfung statist. Hypothesen, d. h. von Annahmen über die Wahrscheinlichkeitsverteilung einer Zufallsgröße. Im *Signifikanztest* wird die Entscheidung anhand einer konkreten Stichprobe getroffen. Die Hypothesenprüfung geht davon aus, daß 1) eine aufzustellende Hypothese  $H_0$  parametrisch oder parameterfrei sein kann. Eine parametrische Hypothese setzt voraus, daß der Verteilungstyp der betrachteten Zufallsgrößen bekannt ist, so daß durch die Hypothese nur noch eine endliche Anzahl von Parametern zu spezifizieren ist, in denen sie sich unterscheiden. Eine parameterfreie Hypothese wird dagegen immer dann aufgestellt, wenn die Voraussetzungen über einen bestimmten Verteilungstyp der Variablen relativ unsicher sind oder wenn ein Prüfverfahren für einen vorliegenden Verteilungstyp nicht vorhanden ist. Unter der Auswahl oder Konstruktion einer Test- oder Prüfgröße versteht man 2) eine Stichprobenfunktion, deren Verteilung unter der Annahme vollständig bekannt sein muß, daß die Hypothese gilt. Die bekanntesten Testgrößen zur Hypothesenprüfung sind der *t-Test* und der *F-Test als parametrische Testverfahren* sowie der  *$\chi^2$ -Test als ein nichtparametrisches Verfahren*. 3) Die Definition eines *kritischen Intervalls*, in das die Testgröße mit einer Wahrscheinlichkeit fällt, die nicht größer als eine Zahl  $a$  mit  $0 < a < 1$  unter der Voraussetzung ist, daß die Hypothese gilt. Die Zahl  $a$  heißt *Signifikanzniveau* oder *Irrtumswahrscheinlichkeit*. Häufig verwendete Werte dafür sind  $a = 0,5; 0,1; 0,01$ . Das kritische Intervall wird auch *Ablehnungsbereich* genannt. Die Begrenzung des kritischen Intervalls bezeichnet man als *Ablehnungsschwelle* oder *Signifikanzgrenze*. Je nach der Wahl des Bereichs unterscheidet man beim Test zwischen einer einseitigen und einer zweiseitigen Fragestellung; eine zweiseitige Fragestellung liegt vor, wenn sich der Ablehnungsbereich auf beiden Seiten der Annahmezone befindet. Die Entscheidung über die Verwendung einer ein- oder zweiseitigen Fragestellung hängt von der Spezifizierung der Alternativhypothese ab. Bei einem Zwei-Stichproben-Problem ist z. B. die Hypothese zu prüfen, ob die beiden Stichproben in bezug auf den Erwartungswert  $m$  aus der gleichen Grundgesamtheit stammen. Die Prüfung der *f Nullhypothese*  $H_0: \mu_1 = \mu_2$  gegen die *Alternativhypothese*  $H_1: m_1 < m_2$ , wenn  $m_1 > m_2$  unmöglich ist oder nicht interessiert, führt zu einem *einseitigen Test*; die Hypothese  $H_0: m_1 = m_2$  gegen  $H_1: m_1 \neq m_2$  führt dagegen zu einem *zweiseitigen Test*. Im nächsten Schritt der Hypothesenprüfung wird 4) eine *Entscheidungsregel* gebildet. Fällt für eine Stichprobe der Wert, den die Testgröße annimmt, in den kritischen Bereich, so wird die Hypothese abgelehnt, anderenfalls ist gegen die Richtigkeit der Hypothese nichts einzuwenden. Bei dieser Entscheidung können zwei Arten von Fehlern auftreten: Im Fehler 1. Art wird die Hypothese  $H_0$