

Glossar – Abkürzungen und Begriffe

- **BIJU-Studie**

Studie des Max-Planck-Instituts für Bildungsforschung (Berlin) zu den Bildungsverläufen und psychosozialen Entwicklungen im Jugend- und Erwachsenenalter.

Weitere Informationen unter: www.biju.mpg.de

- **BETA(-GEWICHT)**

⇒ Regressionsanalyse

- **d , EFFEKTSTÄRKE d , STANDARDISIERTE EFFEKTSTÄRKE d**

ist ein standardisiertes Maß für Merkmalsunterschiede zwischen zwei Gruppen. Sie wird berechnet, indem die Differenz der \Rightarrow Mittelwerte der Gruppen durch die gemeinsame \Rightarrow Standardabweichung dividiert wird.

- **EFFEKTSTÄRKE**

⇒ d

- **Eta², Bestimmtheitsmaß Eta²**

ist die Maßzahl für die erklärte Varianz bei der \Rightarrow Kovarianzanalyse und der \Rightarrow Varianzanalyse. Eta² bezeichnet den Varianzanteil, den man erhält, wenn man die Einzelwerte durch den jeweiligen Gruppenmittelwert ersetzt, daraus die Varianz berechnet und durch die ursprüngliche Varianz teilt; es variiert zwischen 0 und 1. Z. B. gibt das mit der Schulform verbundene Eta² für ein Leistungsmerkmal Auskunft darüber, welchen Anteil Schulformunterschiede (Mittelwertdifferenzen) an den Leistungsunterschieden zwischen den Schülerinnen und Schülern überhaupt haben.

- **IALS**

„International Adult Literacy Study“, von der IEA durchgeführt.

- **IEA**

„International Association for the Evaluation of Educational Achievement“, Amsterdam

- **IGLU**

„Internationale Grundschul-Lese-Untersuchung“

IGLU wird international unter der Bezeichnung „Progress in International Reading Literacy Study“ (PIRLS) von der IEA durchgeführt.

- **ITEM**

Der kleinste Baustein in einem empirischen Erhebungsinstrument.

Beispiele: Frage in einem Fragebogen, Aufgabe in einem Mathematiktest.

- **ITEM-RESPONSE-THEORIE (IRT)**

Mit Tests sollen Leistungen der Schüler/innen festgestellt werden. Die Schüler/innen, so der Ausgangspunkt, verfügen über bestimmte Fähigkeiten, die nicht direkt beobachtbar sind (latent traits). Was sich beobachten (messen) lässt, ist, wie die Schüler/innen auf die vorgegebenen Items reagieren, wie sie also z.B. die Aufgaben in einem Mathematiktest lösen.

Ob ein Item bewältigt wird oder nicht, hängt von zwei Größen ab: Der Kompetenz (den Fähigkeiten) der Person und der Schwierigkeit (den Anforderungen) des Items. Lassen sich beide Aspekte (Personenkompetenz, Itemschwierigkeit) auf derselben Dimension verorten, gilt im Prinzip, dass eine Person dann ein Item löst, wenn ihr Kompetenzniveau über dem Schwierigkeitsniveau liegt.

Die klassische Annahme der Testtheorie lautet: Es gibt einen deterministischen Zusammenhang dergestalt, dass eine Person alle Items löst, die unterhalb ihres Fähigkeitsniveaus liegen und keines oberhalb. In probabilistischen Modellen (z.B. bei der → RASCH-SKALIERUNG) wird von Lösungswahrscheinlichkeiten ausgegangen. Mit diesen Modellen verträglich ist der beobachtbare Umstand, dass zuweilen sehr fähige Personen auch manches einfache Item nicht bewältigen und manch „blindes Huhn auch ein Korn findet“. Der strikte Zusammenhang zwischen Fähigkeit und Schwierigkeit wird zugunsten eines tendenziellen aufgegeben: Je fähiger eine Person ist, desto größer ist die Wahrscheinlichkeit, Items verschiedener Schwierigkeitsstufen zu lösen. Diese Wahrscheinlichkeit nimmt mit zunehmender Schwierigkeit ab.

- **KESS**

„Kompetenzen und Einstellungen von Schülerinnen und Schülern“, Hamburg

- **KORRELATION**

Korrelation meint den statistischen Zusammenhang zwischen zwei Merkmalen. Die Enge des Zusammenhanges wird mittels des Korrelationskoeffizienten ausgedrückt. I.d.R. (aber nicht zwangsläufig) misst der Korrelationskoeffizient lineare Zusammenhänge, also solche der Proportionalität (je - desto; Beispiel: Körpergröße und Körpergewicht).

Korrelationskoeffizienten sind Zusammenhangsmaße, die zwischen -1 und +1 variieren können. Das Vorzeichen zeigt an, in welcher Richtung ein Zusammenhang besteht (gleichsinnig/proportional oder ungleichsinnig/umgekehrt proportional), der Betrag des Koeffizienten quantifiziert das Ausmaß des (linearen) Zusammenhanges. Der Koeffizient darf aber nicht als Prozentanteil des maximal möglichen Zusammenhanges interpretiert werden. Sein Quadrat macht eine Aussage über den Anteil gemeinsamer Varianz, d.h. in welchem Ausmaße die beiden miteinander korrelierte Merkmale sich (gleichsinnig oder ungleichsinnig) gemeinsam verändern. (Und das Quadrat z.B. eines Koeffizienten von 0,5 beträgt nur 0,25.)

Korrelative Zusammenhänge dürfen nicht automatisch als kausale interpretiert werden. Wenn zwei Merkmale miteinander korrelieren, kann dies vielfältige Ursachen haben. Erforderlich ist in jedem Falle eine eingehende inhaltliche Überprüfung.

- **LAU**

„Aspekte der Lernausgangslage und der Lernentwicklung“, Hamburg

- **LÖSUNGSWAHRSCHEINLICHKEIT**

Die Lösungswahrscheinlichkeit einer Aufgabe gibt an, wie groß die Wahrscheinlichkeit ist, dass irgendein Schüler bzw. irgendeine Schülerin diese Aufgabe löst. Die Lösungswahrscheinlichkeit wird mit dem Wert p (vom englischen „probability“) angegeben und liegt zwischen 0 und 1. Eine Lösungswahrscheinlichkeit von $p = 0,47$ beispielsweise besagt, dass 47 Prozent der Schülerinnen und Schüler einer definierten Gruppe diese Aufgabe lösen.

- **MITTELWERT**

Kurzbezeichnung für den arithmetischen Mittelwert. Der Mittelwert darf berechnet werden, wenn mindestens \Rightarrow intervallskalierte Daten vorliegen. Er ist die Summe der Einzelwerte aller Fälle dividiert durch die Fallzahl.

- **MULTIPLE REGRESSIONSANALYSE**

\Rightarrow Regressionsanalyse

- **MULTIPLES R, MULTIPLES R²**

\Rightarrow Regressionsanalyse

- **N**

bezeichnet die Anzahl der Fälle, z. B. die Anzahl der Personen, die mit einem bestimmten Test erfasst wurden bzw. eine bestimmte Aufgabe gelöst haben.

- **ODDS**

Odds (engl.; deutsch: Wettquotient“) entspricht der Häufigkeit eines Ereignisses dividiert durch die Häufigkeit des dazu komplementären Ereignisses. Werden z. B. im Durchschnitt pro 100 Geburten 51 Jungen geboren, betragen die odds für ein zufällig ausgewähltes Baby, ein Junge zu sein: $51/49 = 1,04$. Die Bezeichnung „Wettquotient“ rührt daher, dass bei einer fairen Wette diejenigen, die auf das komplementäre Ereignis „Mädchen“ setzen, zusätzlich zur Rückzahlung des eigenen Einsatzes das 1,04 fache als Gewinn erhalten.

Wenn die odds eines Ereignisses größer als 1 sind, so ist das Eintreten dieses Ereignisses wahrscheinlicher als sein Nicht-Eintreten (die Odds eines sicheren Ereignisses sind unendlich); wenn die Odds eines Ereignisses kleiner als 1 sind, so ist das Nicht-Eintreten des Ereignisses wahrscheinlicher als sein Eintreten (die odds eines unmöglichen Ereignisses betragen 0).

Beispiel: Wenn von 100 deutschen Jugendlichen 72,2 einen Ausbildungsvertrag bekommen, dann betragen deren odds $72,2/27,8 = 2,60$. Erhalten von 100 ausländischen Jugendliche 31,6 einen Ausbildungsvertrag, so sind deren odds $31,6/68,4 = 0,46$. (Wer also bei einem deutschen Jugendlichen darauf setzt, dass er einen Ausbildungsvertrag erhält, gewinnt lediglich $(27,8/72,2)$ das 0,39-fache zu seinem Einsatz; die Wette auf den Ausbildungsvertrag eines ausländischen Jugendlichen erbringt wegen der Seltenheit des Erfolgs hingegen das 2,16-fache $(68,4/31,6)$.)

- **ODDS RATIO**

Ein odds ratio erhält man bei der Division der odds der Zielgruppe durch die odds der Referenzgruppe.

Beispiel: In Fortführung des Beispiels mit den Ausbildungsverträgen ergibt sich der odds ratio $2,60/0,46 = (72,2/27,8) / (31,6/68,4) = 5,62$.

Dieser Wert darf nicht als Wahrscheinlichkeit interpretiert werden. Daher ist es missverständlich zu sagen, dass die „Chance“ eines deutschen Jugendlichen, einen Ausbildungsvertrag zu bekommen, 5,6 mal so hoch sei wie die eines ausländischen Jugendlichen, weil „Chance“ als Wahrscheinlichkeit interpretiert wird. Tatsächlich ist diese Wahrscheinlichkeit nur $72,2/31,6 = 2,3$ mal so hoch.

- **OVERSAMPLING**

gezielte und damit kontrollierbare Übervertretung bestimmter Gruppen in einer Stichprobenziehung

- **PIRLS**

siehe IGLU

- **r**

Abkürzung für Pearsons r , \Rightarrow Korrelation.

- **RASCH-SKALIERUNG**

Eine Skalierungsmethode, die auf der probabilistischen \rightarrow Item-Response-Theorie beruht.

Ihr Vorteil beruht darauf, dass sich die Fähigkeiten von Personen und die Schwierigkeiten von Items auf derselben Dimension abbilden lassen, wie dies z.B. bei PISA und IGLU vorgeführt wird.

Durch die Zuordnung von Items zu bestimmten RASCH-Werten (vgl. Abbildung 3.1), die i.W. im Bereich von 60 bis 140 liegen, können die unterschiedlichen Niveaus inhaltlich interpretiert werden. Hierzu sind die jeweiligen Items daraufhin zu untersuchen, welche Anforderungen sie beinhalten (Beherrschen der Grundrechenarten, Anwenden einfacher Kalküle etc.; vgl. KLIEME, NEUBRAND & LÜDTKE 2001). Können auf diese Art und Weise die unterschiedlichen RASCH-Niveaus im Hinblick auf die Items inhaltlich bestimmt werden, so lässt sich dies auf die Kompetenzen der Personen (Schüler/innen) übertragen: Die herausgearbeiteten Anforderungen, die in den Items stecken, sind die Kompetenzen, die die Schüler/innen aufweisen müssen, um die Items (mit einer gewissen Wahrscheinlichkeit) zu bewältigen.

Zu beachten ist, dass die konkrete Größe von Mittelwert und Streuung der RASCH-Werte (100 und 15 im vorliegenden Fall) sich nicht zwingend aus dem mathematischen Teil der RASCH-Skalierung ergeben. Dort ergeben sich durch das Rechenverfahren bedingt i.d.R. "krumme" Werte, die erst durch anschließende Normierung auf übliche und anschauliche Größenordnungen gebracht werden. Das stellt keine Manipulation dar und verfälscht nicht die Verhältnisse, denn die Anordnung der Personen und der Items auf der kombinierten Fähigkeits-/Schwierigkeitsdimension ändert sich hierdurch nicht.

- **REGRESSIONSANALYSE**

Die (multiple) Regressionsanalyse ist ein Analyseverfahren, das den Zusammenhang zwischen einer \Rightarrow intervallskalierten abhängigen (zu erklärenden) Variable (dem sogenannten Kriterium) und mehreren, ebenfalls intervallskalierten unabhängigen (erklärenden) Variablen (den sogenannten Prädiktoren) aufdeckt. Bei der Berechnung der Regressionsgleichung werden die \Rightarrow Korrelationen der Prädiktoren untereinander berücksichtigt. Die Maßzahl für den Zusammenhang zwischen allen Prädiktoren einerseits und dem Kriterium andererseits ist das „multiple R“. Es kann wie \Rightarrow Pearsons r interpretiert werden, kann allerdings keine negativen Werte annehmen. Das quadrierte multiple R entspricht der erklärten Varianz. Der eigenständige Beitrag jedes einzelnen Prädiktors (bei Konstanthaltung der anderen Prädiktoren) zur Aufklärung der Unterschiede im Kriterium wird mit den Beta-Gewichten (Beta-Koeffizienten) angegeben.

- **RELIABILITÄT**

Reliabilität ist ein Gütekriterium für sozialwissenschaftliche Messungen. Sie entspricht der Zuverlässigkeit einer Messung. Reliabel ist ein Test oder eine Skala, wenn nur geringe Messfehler auftreten. Es gibt verschiedene Methoden, die Reliabilität zu überprüfen, z. B. Cronbachs \Rightarrow Alpha.

- **SIGNIFIKANZ**

Bedeutsamkeit, im vorliegenden Zusammenhang **statistische Bedeutsamkeit** im Sinne von Überzufälligkeit.

Empirische Untersuchungen beruhen i.a. auf Stichproben. Gewollt sind aber Aussagen über die zugrundeliegende Grundgesamtheit. Die (sog. Inferenz-)Statistik hilft dabei, die Verallgemeinerung von der Stichprobe auf die Grundgesamtheit abzusichern.

Beispiel: Zeigen Jungen oder Mädchen bessere Leistungen in Mathematik? Je einer Gruppe von Jungen und Mädchen wird derselbe Mathematiktest vorgegeben, der zu zwei Mittelwerten führt, einen für die Jungen, einen für die Mädchen, etwa $MW(J)=11,4$ und $MW(M)=10,5$. Ist diese Differenz ein zufälliges Ergebnis, das nur für die Stichprobe gilt, oder gilt es für alle Jungen und Mädchen?

Die Statistik stellt Verfahren bereit, anhand derer diese Frage (innerhalb gewisser Grenzen) beantwortet werden kann. Als signifikant gilt das Ergebnis dann, wenn der statistische Test zum Ergebnis führt, dass die gefundene Differenz überzufällig, also unabhängig von der konkreten Stichprobe ist. Verschiedene Größen sind beim statistischen Testen zu berücksichtigen, u.a. die Größe der Stichproben; es gilt: Je größer die Stichproben, desto eher ergeben sich Signifikanzen.

Statistische Signifikanz und inhaltliche Bedeutsamkeit gehen nicht zwangsläufig miteinander einher. Bei großen Stichproben sind selbst kleinste Differenzen statistisch signifikant, aber inhaltlich bedeutungslos. Statistik ersetzt also nicht die inhaltliche Auseinandersetzung.

- **SKALA**

1. Kurzbezeichnung für die Ausprägungen einer Einschätzskala (Ratingskala). Bei der Vorgabe der Antwortmöglichkeiten von 1 = „trifft völlig zu“ bis 4 = „trifft überhaupt nicht zu“ im Schülerfragebogen spricht man z. B. von einer vierstufigen Skala.

2. Inhaltlich zusammenfassende Einzelitems können, z. B. durch Aufsummieren oder Mittelwertbildung, zu einer Skala zusammengefasst werden. Ein Beispiel ist die Skala „(positive) Erwartungen an die Berufliche Schule“, die auf dem Wege der individuellen Mittelwertbildung über 5 Fragen aus dem Fragebogen für Schülerinnen und Schüler ermittelt wurde.

- **SKALIERUNG**

Eine Skala bezeichnet das einem Messvorgang zugrunde liegende Bezugssystem, das, worauf z.B. eine zu testende Person verortet werden soll, etwa eine Skala zur Lesekompetenz oder zur Rechenfertigkeit.

Skalierung ist der Vorgang, eine Skala zu erstellen. Ausgangspunkt sind in der Regel eine größere oder kleinere Menge von Items, von denen angenommen wird, sie trügen etwas zur zu messenden Eigenschaft bei. Gesammelt werden könnten beispielsweise verschiedene Rechenaufgaben, die einer Stichprobe vorgelegt werden. Das Problem besteht nun (wie in einer Klassenarbeit) darin, die Einzellösungen so zusammenzufassen, dass jeder Person der Stichprobe ein möglichst aussagekräftiges Gesamtergebnis auf der Dimension Rechenfähigkeit zugeordnet werden kann.

Ein häufig gewähltes Verfahren besteht darin, jedem Item eine bestimmte maximal mögliche Punktzahl zuzuordnen. Summenbildung ergibt dann die einzelnen Ausprägungen der Skala. Dies Vorgehen geht u.a. davon aus, dass alle Items von ihren Anforderungen her einen Beitrag genau zu der zu messenden Eigenschaft leisten, dessen Ausmaß angemessen durch die Punkte repräsentiert wird.

Die Verfahren der Testkonstruktion systematisieren den Vorgang der Skalierung. Hierbei gibt es verschiedene Ansätze; → ITEM-RESPONSE-THEORIE.

- **STREUUNG (STANDARDABWEICHUNG, standard deviation (s, sd oder SD))**

Die Streuung ist ein Maß für die Heterogenität. Die Streuung ist kein absolutes Maß, d.h. einer Streuung von z.B. 4,4 lässt sich nicht ansehen, ob dahinter eine große oder kleine Heterogenität steht. Streuungen lassen sich i.d.R. erst dann sinnvoll interpretieren, wenn zwei Streuungen miteinander verglichen werden können.

Beispiel: Erhoben werden die Mathematikleistungen in zwei Parallelklassen. Mittelwert und Streuung betragen in der ersten Klasse $MW=22,3$ - $s=2,9$ und in der zweiten $MW=22,5$ - $s=4,4$. Daraus lässt sich ablesen, dass bei demselben mittleren Leistungsniveau (nahezu identische Mittelwerte) die Schüler/innen der ersten Klasse in ihren Mathematikleistungen homogener als jene der zweiten Klasse sind.

In die Berechnung der Streuung geht der Mittelwert ein. Dessen Kenntnis erleichtert neben dem exemplarischen Vergleich zweier Streuungen das Einschätzen, ob eine Streuung groß oder klein ist. Hierzu muss die Höhe der Streuung auf die Größe des Mittelwertes bezogen werden.

Beispiel: Eine Streuung von $s=4,4$ ist eher klein, beträgt der Mittelwert $MW=35,7$, und groß bei einem Mittelwert von $MW=5,1$.

Ist die Art der Verteilung bekannt, aus der die Daten stammen, für die Mittelwert und Streuung berechnet wurden, dann lassen sich weitergehende Aussagen treffen. Für die Normalverteilung gilt, dass innerhalb des Bereiches $MW-s$ und $MW+s$ rund zwei Drittel aller Werte liegen.

Beispiel: Die für PISA 2000 entwickelten Skalen aus den nationalen Ergänzungstests besitzen einen Mittelwert von 100 und eine Streuung von 30. Damit liegen die Leistungen von zwei Drittel aller in Deutschland untersuchten Schüler/innen im Bereich von $100-30=70$ und $100+30=130$.

Das Quadrat der Streuung heißt Varianz.

- **VARIANZ**

Die Varianz ist das Quadrat der \Rightarrow Standardabweichung. Mathematisch ist die Varianz der Durchschnitt aus den quadrierten Abweichungen aller Einzelwerte vom Gesamtmittelwert.

- **VARIANZANALYSE**

Dieses Analyseverfahren ist mit der \Rightarrow Regressionsanalyse verwandt. Auch hier geht es um die Aufklärung individueller Unterschiede in einer abhängigen Variable durch eine oder mehrere unabhängige Variablen. Der Hauptunterschied zur Regressionsanalyse besteht darin, dass die unabhängigen Variablen nicht \Rightarrow intervallskaliert zu sein brauchen. Die Maßzahl für die erklärte Varianz heißt \Rightarrow η^2 .

- **Z-Transformation**

Durch eine Z-Transformation werden die Abweichungen der Werte der einzelnen Fälle vom Mittelwert durch die Standardabweichung dividiert. Als Ergebnis erhält man eine neue Verteilung mit Mittelwert 0 und Standardabweichung 1.