Normwerte der Testdiagnostik

– ihre Berechnungen und Umrechnungen ineinander –

Aufsatz und Excel-Projekt zum Gebrauch neben der Lehrveranstaltung
"Konventionelle und computergestützte Testdiagnostik"
sowie für die allgemeine Information zu diesem Thema

Autor:
Dr. Helmut Stauche
Institut für Erziehungswissenschaft
Universität Jena

Februar 2008

Die hier betrachteten Normwerte dienen dem Einsatz pädagogisch-psychologischer Tests, die nach den Prinzipien der klassischen Testtheorie konstruiert wurden. Unter klassischer Testtheorie soll dabei die Sparte diagnostischer Tests verstanden werden, die am Anfang des 20. Jahrhunderts ihren Ausgang nahm, in der Mitte desselben Jahrhunderts im wesentlichen von Gulliksen systematisiert wurde und mit der präzisierenden Weiterentwicklung durch Lord und Novick am Ende der sechziger Jahre einen weiteren Aufschwung erlebte (vgl. Michel und Conrad 1992, S. 16). Der Löwenanteil der auf dem Markt befindlichen pädagogisch-psychologischen Tests ist klassisch konstruiert.

A. Normalverteilungsbasierte Normwerte

Eine der wesentlichen Grundannahmen der klassischen Testtheorie ist, dass das betrachtete Testmerkmal normal verteilt ist und Intervallskalenniveau besitzt. Nur dieser Ausgang lässt die dort notwendigen Berechnungen von arithmetischen Mittelwerten und Streuungswerten zu. Der Output eines jeden klassischen Tests ist die Positionierung der getesteten Person in der Verteilung der ihr zugeordneten Normstichprobe, das heißt, die Angabe, wie weit nach unten oder oben entfernt das Ergebnis der Testperson vom Mittelwert der Eichstichprobe liegt.
Sinnstiftend ist dabei die Einteilung des gesamten Spektrums der normal verteilten Werte in Bereiche, die ihre Grenzen aus dem Wert der Standardabweichung der Testvariablen beziehen. Man spricht von einfacher, doppelter und dreifacher Streuungsbreite. Eine Erweiterung auf ein noch breiteres Intervall macht keinen Sinn, weil innerhalb der dreifachen Streuungsbreite bereits ca. 99,73 % aller erzielten Werte liegen. Dennoch müssen Testergebnisse auch für Ausreißer nach unten oder nach oben angebbar sein.

Abbildung 1 zeigt diesen Sachverhalt in der oberen Zeile der Abszissenbeschriftung.

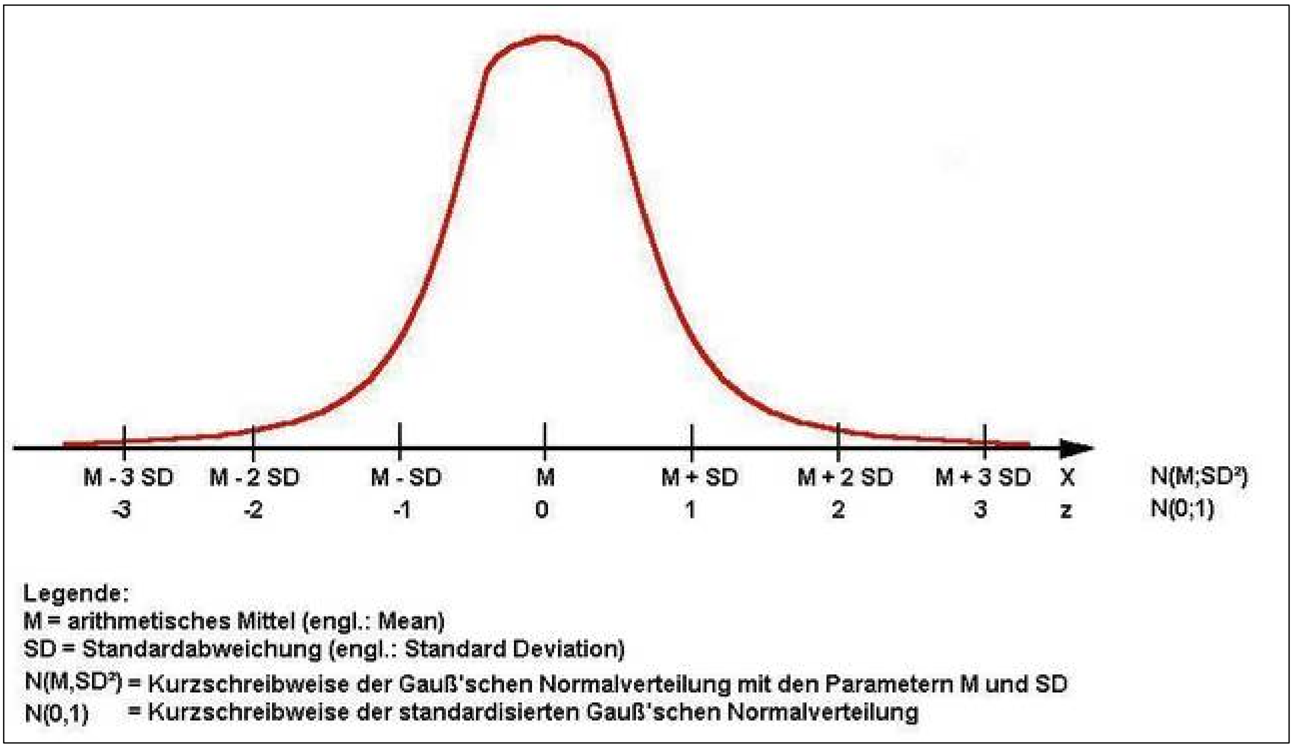

Abbildung 1[1]

Die ebenfalls in dieser Abbildung gezeigte standardisierte Normalverteilung N(0;1) – siehe untere Zeile der Abszissenbeschriftung – macht die Angabe von Positionen in der Verteilung erst praktikabel. Durch die Transformation der Werte der Testvariablen X in sogenannte z-Werte gemäß der Transformationsvorschrift **z = (X – M) / SD** entsteht eine Normalverteilung mit dem lokalen Maximum bei Null und standardisierter Flankensteilheit, bedingt durch die einheitliche Varianz von 1.

Diese z-standardisierte Normalverteilung erleichtert vor allem das Vergleichen von Ergebnissen unterschiedlicher Tests, unterschiedlicher Subtests eines komplexen Testmerkmals oder zu unterschiedlichen Normstichproben.

Dennoch sind z-Werte nicht anschaulich: Sie rangieren – wie der Abbildung entnommen werden kann – im praktischen Bereich (alle minus Menge der Ausreißer) zwischen -3 und +3 – und sind fast durchweg Dezimalzahlen. Dies veranlasste die Theoretiker der klassischen Testtheorie, verschiedene Lineartransformationen der z-Skala zu entwerfen.

Alle so konstruierten Normwerte spreizen den Wertevorrat der z-Skala um einen Faktor und rücken den Mittelwert der Verteilung auf eine positive Ganzzahl:

Normwert = Spreizungsfaktor • z + Abszisse des Mittelwerts

[1] In dieser Abbildung fehlt bewusst die Ordinate, um sowohl die nicht standardisierte als auch die standardisierte Normalverteilung darstellen zu können.

Die so entstandenen gebräuchlichsten Normwerte sind (ausgehend von z):

Normwert	Spreizungs-faktor	Lage des Mittelwerts	Formel	Mittel-wert	Bereich ohne Interventionsbe-darf	praktischer Bereich
T	10	50	T=10z+50	50	40...60	20...80
IQ	15	100	IQ=5z+100	100	85...115	55...145
Standard-werte	10	100	SW=10z+100	100	90...110	70...130
Standard Nine	2	5	SN=2z+5	5	4...6	1...9

Tabelle 1

Zwei Anmerkungen hierzu:

1. Der IQ (Intelligenzquotient) dient häufig zur Angabe der Testergebnisse von Intelligenz- und anderen kognitiven Tests. Der Name IQ hat sich gehalten, obwohl in der Formel des IQ kein Quotient vorkommt (siehe Tabelle 1). Er geht auf eine historische Definition zurück, nach der das Intelligenzalter und das Lebensalter nach einer so genannten Äquivalenznorm in Beziehung gesetzt wurden:
   IQ = Intelligenzalter / Lebensalter •100
   Eine solche Definition macht jedoch nur Sinn für einen relativ kurzen Lebensabschnitt, in dem die kognitive Entwicklung des Menschen hinreichend linear verläuft. Auf Äquivalenznormen wird in diesem Aufsatz nicht näher eingegangen.

2. Die Standard Nine haben einen begrenzten Wertevorrat von 1 bis 9, was ihren Namen ausmacht. Damit wird aber die dreifache Streuungsbreite nicht bedient. Bezieht man diese ein, gelangt man zum seltsam anmutenden Wertevorrat von -1 bis 11. Auch diese Skala existiert, man bezeichnet sie als Centile.

Abschließend zur Vorstellung einiger Normwertskalen unterstütze ich vollinhaltlich eine Aussage von Dieterich aus dem Jahre 1973:
„Es gibt keinen vernünftigen Grund, weshalb so viele Normskalen im Gebrauch sind. Man könnte sich eigentlich einmal auf ein paar vernünftige Skalen einigen. Dabei könnte man einen oder zwei Mittelwerte festlegen und einige verschiedene Streuungen, sodaß man fein differenzierende und grob differenzierende Tests konstruieren kann.“

B. Der Prozentrang (PR)

Ein weiterer stark eingeführter und weithin bekannter Normwert bei pädagogisch-psychologischen Testverfahren ist der Prozentrang. Dieser ist eine besondere Häufigkeitsangabe, die nicht notwendigerweise an Normalverteiltheit der Messwerte in der Normstichprobe gebunden ist, was durchaus als Vorteil dieses Wertes anzusehen ist.
Der Prozentrang ist – wie auch die prozentuale relative Häufigkeit – eine Prozentangabe. Zur Unterscheidung von letzterer, von der er sich in seiner Aussage und damit zahlenmäßig unterscheidet, wird er ohne das Prozentzeichen und darüber hinaus immer als Ganzzahl angegeben.
Erreichte eine Testperson den Prozentrang der Höhe a, dann besagt dieser, dass a % der Testpersonen der Vergleichsstichprobe eine niedrigere (oder gleiche) Ausprägung des Testmerkmals aufweisen. Dabei ist es egal, ob es sich um ein Persönlichkeitsmerkmal oder ein Leistungsmerkmal handelt.

Zu unterscheiden ist die Berechnung der Prozentrangwerte für einen eigenen Test (das muss nicht unbedingt ein nach den strengen Regeln der Testdiagnostik hergestellter sein – auch für eine beliebige bepunktbare Leistungsanforderung lässt sich der Prozentrang berechnen) und der auf den Prozentrang hinauslaufende Output eines markteingeführten Papier-Bleistift- oder Computertests.
Im ersten Fall bildet die eigene Stichprobe (Klasse, Seminargruppe etc.) die Vergleichsgrundlage, im zweiten Fall die – fast immer repräsentativ aufgebaute – Normstichprobe des Testverfahrens.

C. Vergleich

Stellt man normalverteilungsbasierte Normwerte und Prozentränge gegenüber, dann ist vor allem ein Ergebnis evident:
Während normalverteilungsbasierte Normwerte mit den erreichten Testpunkten mit r=1 – also maximal – korrelieren, ist dies für Prozentränge nicht der Fall. Konstante normalverteilungsbasierte Normwertdifferenzen können zu ganz unterschiedlichen Prozentrangdifferenzen gehören. Der Prozentrang darf bei einem Leistungstest von daher nicht proportio-

nal zur Leistung selbst aufgefasst werden. Dies wird deutlich im Rechenbeispiel des Excel-Projekts im Blatt 5.

D. Besseres Verständnis der Normwerte durch ein Excel-Projekt

Das diesem Aufsatz beigefügte Excel-Projekt **normwerte.xls**, in dessen Hintergrund zahlreiche VBA-Programme laufen, will das oben Gesagte illustrieren.
Alle Berechnungen und Demonstrationen beruhen auf einem Test mit einem Maximalscore von 25 Testpunkten.

– Im **Blatt 1** wird demonstriert, wie man ausgehend von bekannten absoluten Häufigkeiten der Testpunkte zum Prozentrang gelangt. Um nicht mit allzu großen Zahlen zu operieren, wurde eine (Norm-)Stichprobe von N = 348 zugrunde gelegt.
  Der Weg führt von der absoluten Häufigkeit über die kumulative Häufigkeit und die prozentuale kumulative Häufigkeit zum Prozentrang.

– Analog dazu arbeiten die Buttons im **Blatt 2**. Allerdings kann der Benutzer hier im Bereich von 0 bis maximal 25 Testpunkten eigene absolute Häufigkeiten eintragen. Dieses Blatt ist damit auch für eine <u>Echtberechnung des Prozentranges</u> bei eigenen Tests geeignet.

– Das **Blatt 3** zeigt – wiederum ausgehend von den Testpunkten und deren absoluter Verteilung wie im Blatt 1 – die Genese der z-Werte und aller oben beschriebenen z-basierten Normwertskalen. Bei den Standard-Nine-Werten wurde die dreifache Streuungsbreite definitionsgemäß ausgeblendet.

– Im **Blatt 4** wird gezeigt, dass man Prozentrangwerte bei Bekanntheit des Umfangs der Stichprobe in einer rückwärts gerichteten Rechnung in die absoluten Häufigkeiten umrechnen kann und von da aus natürlich weiter in z-Werte und in die z-basierten Skalen. Wichtig ist, dass dies nur ausgeführt werden kann, wenn von Normalverteilung des betreffenden Testmerkmals ausgegangen werden darf.
  Ein derartiges Vorgehen ist dann sinnvoll, wenn Ergebnisse verschiedener Tests zu einer Testperson verglichen werden sollen, und wenn bei einem der Tests nur der

Prozentrangwert und beim anderen nur ein normalverteilungsbasierter Wert ausgegeben wird.

- Schließlich befasst sich das **Blatt 5** mit der oben bereits angeführten Problematik der Interpretation von Prozentrangdifferenzen (vgl. Abschnitt C.). Als Beispiel wird der Vergleich eines symmetrischen Intervalls um T = 50 herangezogen und es wird gezeigt, dass diese T-Differenzen i.a. zu nicht symmetrischen Prozentrangdifferenzen führen.
- Eine Übersicht über die Funktionen des Projekts findet sich im **Blatt 0**.

Literatur:

Dieterich, R. (1973). Psychodiagnostik. München: Reinhardt (Uni-Taschenbücher: UTB 273).

Michel, L. & Conrad, W. (1982). Theoretische Grundlagen psychometrischer Tests. In K. J. Groffmann & L. Michel (Hrsg.), Enzyklopädie der Psychologie, Themenbereich B: Methodologie und Methoden, Serie II: Psychologische Diagnostik (Band I: Grundlagen psychologischer Diagnostik, S. 1-129). Göttingen: Hogrefe.