

9. Zusammenfassung der Testtheorie

Def. Messen (Gutjahr, 1972, u.a.): "Zuordnen von Zahlen zu Objekten, wobei die Relationen zwischen den Objekten durch die Relationen zwischen den Zahlen reflektiert werden sollen.

Def. Test (Lienert, 1967): Ein Test ist ein wissenschaftliches Routineverfahren, zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale, mit dem Ziel einer möglichst quantitativen Aussage, über den relativen Grad der individuellen Merkmalsausprägung.

Testgütekriterien: 1. Objektivität 2. Reliabilität 3. Validität 4. Nebenkriterien
(Normiertheit, Vergleichbarkeit, Oekonomie, Nützlichkeit)

Itemschwierigkeit:

$$p = \frac{\text{Anzahl richtiger Lösungen}}{\text{Gesamtzahl der Antworten auf dem Item}}$$

Trennschärfe eines Items:

Korrelation Item mit Gesamtscore (Summenwert)

Klassische Testtheorie:

Annahme der KTT: Es sei möglich eine Messung beliebig oft zu wiederholen.

$$X = T + F$$

4 Axiome:

1. $E(F) = 0$
2. $\rho(F, T) = 0$
3. $\rho(F_1, T_2) = 0$
4. $\rho(F_1, F_2) = 0$

Denkmodell: Beobachtete Varianz ist wahre Varianz plus Fehlervarianz.

$$\sigma^2(X) = \sigma^2(T) + \sigma^2(F)$$

Reliabilität: a) Quadrat der Korrelation von Rohwert und wahren Wert.
b) Korrelation Test und Paralleltest

$$r_{tt} = \frac{\sigma^2(T)}{\sigma^2(X)} = \rho^2(X, T) = \rho(X, X')$$

Praktische Bestimmung der Reliabilität:

1. Retest-Reliabilität
2. Parallelttest-Reliabilität
3. Odd-Even- oder Split-Half-Reliabilität
4. Interne Konsistenz

Formeln zur praktische Bestimmung der Reliabilität:

1. SPEARMAN-BROWN:

$$r_{ttn} = \frac{n * r_{tt}}{1 + (n - 1) * r_{tt}} \quad \text{bei 2 Testhälften:} \quad r_{ttn} = \frac{2 * r_{tt}}{1 + r_{tt}}$$

=> Testverlängerung:

$$n = \frac{r_{ttn} * (1 - r_{tt})}{r_{tt} * (1 - r_{ttn})}$$

2. CRONBACH Alpha:

$$\alpha = \frac{m}{m - 1} * \left(1 - \frac{\sum_{j=1}^m s_j^2}{s_x^2} \right)$$

Standardmessfehler: Standardabweichung durch die Wiederholung der Messung

$$s_F = s_X * \sqrt{1 - r_{tt}}$$

1. Konfidenzintervall: $T = X \pm Z_{krit} * s_F$

2. Standardfehler intraindividuelle Differenzen:

$$s_F (\text{Intra}) = s * \sqrt{2 - (r_{11} + r_{22})}$$

$$X_2 - X_1 = Z_{krit} * s_F(\text{Intra})$$

3. Standardfehler interindividueller Differenzen:

$$s_F(\text{Inter}) = s * \sqrt{2 - (1 - r_{tt})}$$

$$X_A - X_B = Z_{\text{krit}} * s_F(\text{Inter})$$

Reliabilität von Differenzen:

$$r_{tt \text{ diff}} = \frac{r_{tt1} + r_{tt2} - 2 * r_{12}}{2 * (1 - r_{12})}$$

gültig für $r_{12} \neq 1$

Validität: Wie weit kann man aus dem Testverhalten auf reales Verhalten schliessen? Korrelation Test und Kriterium.

$$\text{Validität: } \rho(X, Y) = \frac{\sigma(X, Y)}{\sigma(X) * \sigma(Y)}$$

Arten der Validität:

1. predictive validity (Voraussagevalidität)
2. concurrent validity (Uebereinstimmungsvalidität)
3. content validity (Inhaltsvalidität)
4. construct validity (Konstruktvalidität)

Verdünnungsformel (attenuation formulae):

$$\rho(T_X, T_Y) = \frac{r(X, Y)}{\sqrt{r(X, X') * r(Y, Y')}}}$$

Ableitungen daraus für Testverlängerung (1.), Kriteriumsverlängerung (2.) & beides (3.)

	Val. nach	Validität vor	
	Verlänge- rung	Verlänge- rung	
1.	$\rho(X_n, Y)$	$= \rho(X, Y) *$	$\frac{\sqrt{r(X_n, X_n')}}{\sqrt{r(X, X')}}}$ Reliabilität des Tests nach Verlängerung Reliabilität des Tests vor Verlängerung
2.	$\rho(X, Y_n)$	$= \rho(X, Y) *$	$\frac{\sqrt{r(Y_n, Y_n')}}{\sqrt{r(Y, Y')}}}$ Rel. des Kriteriums nach Verlängerung Rel. des Kriteriums vor Verlängerung
3.	$\rho(X_n, Y_n)$	$= \rho(X, Y) *$	$\frac{\sqrt{r(X_n, X_n') * r(Y_n, Y_n')}}{\sqrt{r(X, X') * r(Y, Y')}}}$

Zusammenhang zwischen Reliabilität und Validität (Attenuation Paradoxon):

$$\rho(X, Y) = \frac{\sum_i r(x_i, Y) * s(x_i)}{\sqrt{\sum_i \sum_j s(x_i) * s(x_j) * r(x_i, x_j)}}$$

Kritik der KTT:

1. Unüberprüfbarkeit der Annahmen
2. Widerspruchsfreiheit
3. Korrelationskoeffizienten sind stichprobenabhängig
4. Die Skalenfestlegung ist abhängig von der Itemstichprobe
5. Der Messfehler wird in allen Bereichen der Skala gleich gross angenommen

Probabilistische Testtheorie:

Aussagen über Auftretenswahrscheinlichkeiten von beobachtbarem Verhalten:

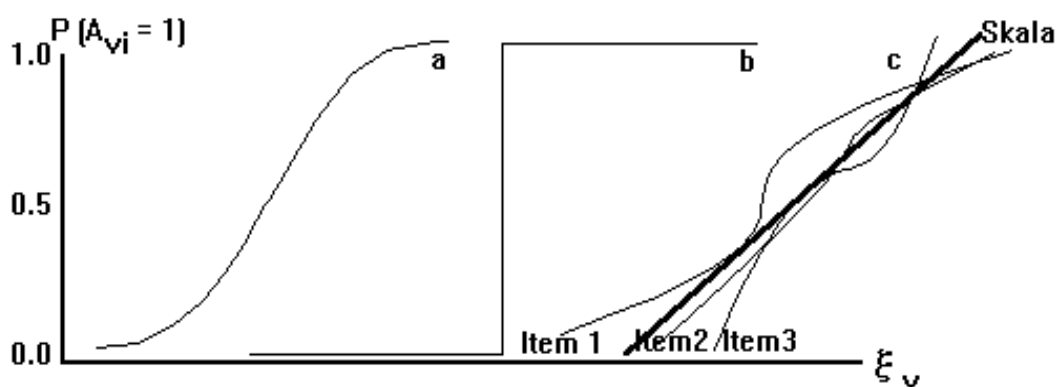
$$P(A_{vi} = 1) = f(\xi_v, \sigma_i)$$

ξ_v ... Fähigkeit der Person

σ_i ... Schwierigkeit des Items i

Die Reaktionswahrscheinlichkeit ist eine Funktion aus Fähigkeit der Person und Itemschwierigkeit.

Itemcharakteristik: Beziehung zwischen Beantwortungswahrscheinlichkeit und Fähigkeit.



a) RASCH-Modell

b) GUTTMAN-Skalierung mit dem Modellgeltungstest:

		Items				
		1	2	3	4	5
Personen	1	1	0	0	0	0
	2	1	1	0	0	0
	3	1	1	1	0	0
	4	1	1	1	1	0

c) KTT: Items werden zu einer Skala zusammengefasst.

Das RASCH-Modell:

Modellgleichung:

$$P(A_{vi} = 1) = \frac{e^{x_v - s_i}}{1 + e^{x_v - s_i}}$$

Die lokale stochastische Unabhängigkeit:

Die Lösungswahrscheinlichkeit einer Person eines Items 2 hängt mit deren Fähigkeit zusammen und nicht mit der richtigen oder falschen Beantwortung eines vorher bearbeiteten Items 1.

Spezifische Objektivität und Stichprobenunabhängigkeit:

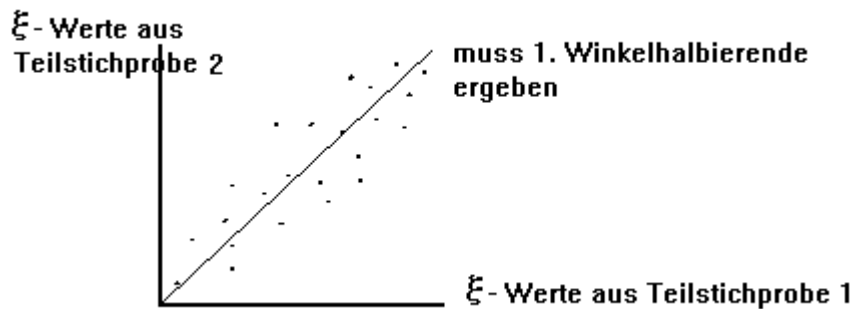
Die Parameterschätzung (Fähigkeit und Schwierigkeit) ist unabhängig von der Itemsstichprobe und der Personenstichprobe (Wechselseitige Stichprobenunabhängigkeit). Die Genauigkeit der Schätzung hängt allerdings von der Stichprobengröße ab.

Der Vergleich zweier Personen ist unabhängig von der Itemauswahl oder der Personenauswahl (spezifisch objektiv).

Die Parameterschätzung und Modellgeltungstest:

Die Parameterschätzung (zwei Unbekannte) erfolgt durch eine bedingte Maximum-Likelihood-Schätzung.

Im Modellgeltungstest werden die geschätzten Parameter aus verschiedenen Stichprobensegmenten im Likelihoodquotiententest verglichen. Oder es erfolgt die graphische Kontrolle:



Items, die dies nicht erfüllen werden ausgesondert.

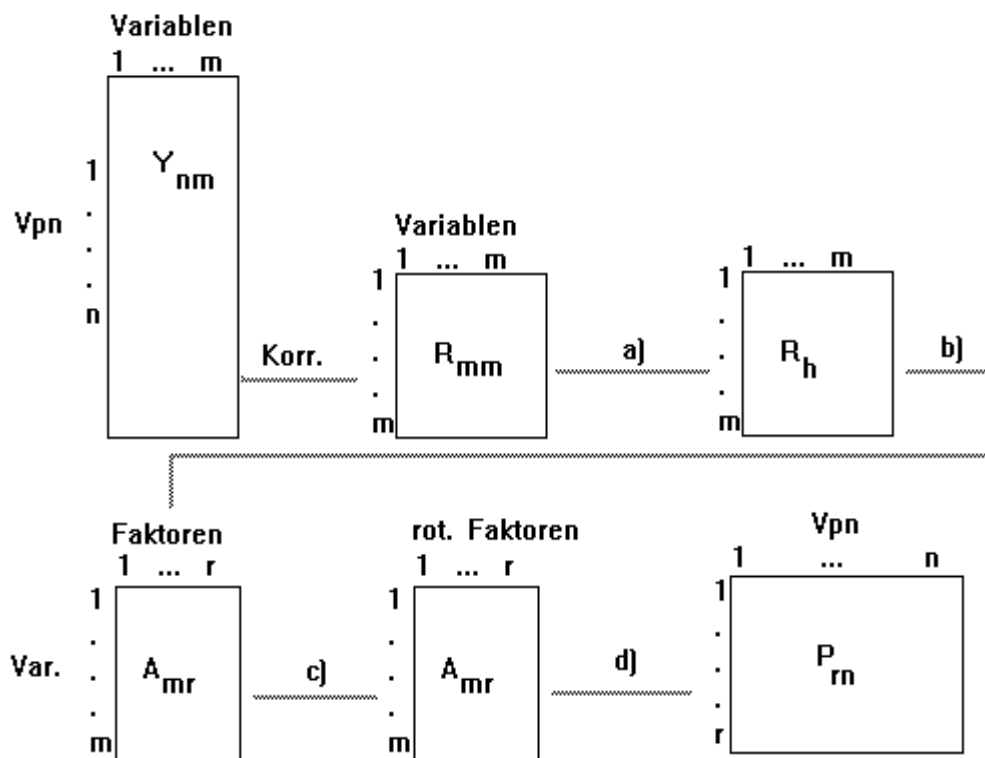
Faktorenanalyse:

Verfahren zur Reduktion von Variablen.

Faktorenanalyse von Tests:

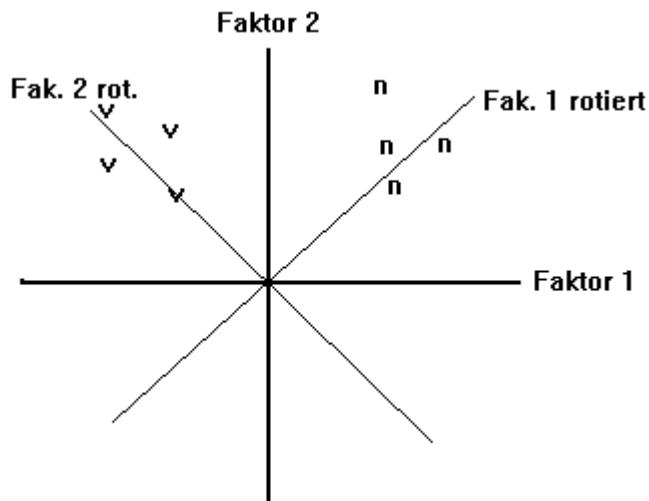
Die Items werden wie Variablen behandelt. Die Reduktion auf Faktoren entspricht der Zuordnung der Items zu Skalen.

Matrizen der FA:



Aus der Datenmatrix ($V_{pn} \times Var$) wird die Interkorrelation der Variablen berechnet ($Var \times Var$). a) Es wird die gemeinsame Varianz der Variablen gesucht (Kommunalitätenproblem). b) Die Variablencluster sollen durch Faktoren ersetzt werden (Faktorenproblem). c) Von den mathematischen wird auf die inhaltlichen Faktoren rotiert (Rotationsproblem).

Bsp: PCA mit VARIMAX-Rotation:



d) Welchen Wert hat eine Person auf einem Faktor (Faktorenwerteproblem).

Die unrotierte Faktorenladungsmatrix A_{mr} und die KTT:

		Faktoren (Skalen)				
		I	II	III	IV	
A_{mr}	1	a_{11}	a_{1II}	a_{1III}	a_{1IV}	a_{11} ... Trennschärfe: Korrelation des Items i mit Skala I (Ladung für Selektion [$<.30$]) $\sum a_{i1}^2 = h_1^2$... Kommunalität $\sum a_{.l}^2 = \lambda_{\max}$... Eigenwert des Faktors (Kennwert für die Varianz des Faktors)
	2	a_{21}	a_{2II}			
	3	a_{31}				
	.					
	.					
Items	.					
	.					
	.					
	m					

Interne Konsistenz:
$$\alpha_{\max} = \frac{m}{m-1} \left(1 - \frac{1}{\lambda_{\max}} \right)$$

Die *Variablenvarianz* der Ausgangsmatrix setzt sich wie folgt zusammen:

stand. Gesamtvarianz					
$s_m^2 = 1 = 100\%$					
Kommunalität			Uniqueness		
h_m^2			u_m^2		
Faktorenladungen			Spezifität		Fehler
a_{m1}^2	a_{m2}^2	...	a_{mr}^2	b_m^2	e_m^2
Reliabilität					
r_{tt}					e_m^2

Voraussetzungen der FA:

$V_{pn} > 100$; Verhältnis (V_{pn} : Variablen) wie (3 : 1); keine dichotomen Variablen.

Entscheidungstheorie:

Klassifikation von *Diagnostischen Entscheidungen* in ausgesuchten Anwendungsgebieten der Psychologie:

Personen	
<i>Eignungsdiagnostik</i> <i>differentielle Indikation</i> <i>Personalauswahl</i>	<i>Psychotherapie</i> <i>psycholog. Trainingsgruppen</i> <i>(Schule, Klinik, ABU)</i>
Selektion	Modifikation
<i>psychologische Beratung</i> <i>(Schule, Beruf)</i>	<i>Deko-Psychologie (Arbeits-,</i> <i>Umweltgestaltung)</i> <i>Familientherapie</i>
Bedingungen	

Kategorisierung von *Personalentscheidungsproblemen:*

	Zurückweisung möglich	Zurückweisung nicht möglich
Eine Stelle ist offen	Selektion	Annahme
Zwei oder mehr Stellen sind offen	multiple Selektion	Klassifikation

1. Selektion:

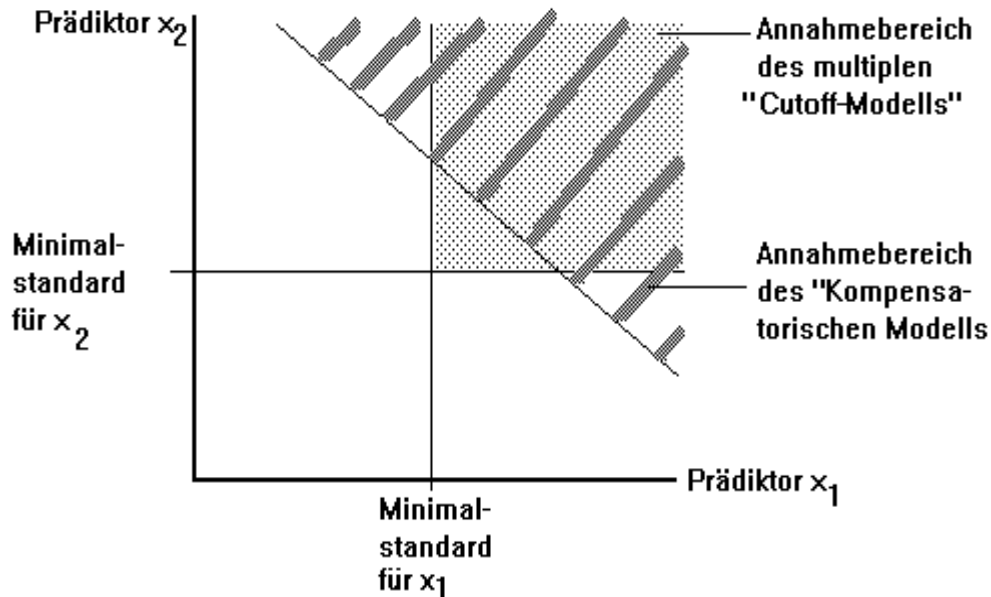
$$y' = b_1 x_1 + b_2 x_2 + \dots + b_n x_n$$

y' ... Vorhergesagtes Kriterium
(z.B. Belastungsfähigkeit)

$x_1 \dots x_n$... Prädiktoren zur Vorhersage des
Kriteriums (z.B. Leistungen im
Subtest 1 bis n)

$b_1 \dots b_n$... Gewichte für die Prädiktoren

Cutoff-Modell (Minimalstandards) vs. Kompensatorisches Modell



2. multiple Selektion:

$$\text{Stelle } a \quad y'_a = a b_1 x_1 + a b_2 x_2 + \dots + a b_n x_n$$

$$\text{Stelle } b \quad y'_b = b b_1 x_1 + b b_2 x_2 + \dots + b b_n x_n$$

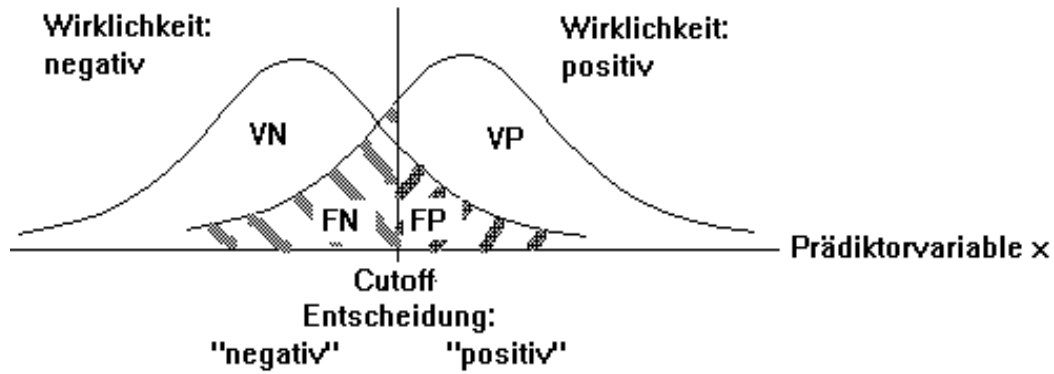
3. Annahme

4. Klassifikation:

Unterschied zwischen den Stellen a und b :

$$(y_a - y_b)' = (a_{b1} - b_{b1}) x_1 + (a_{b2} - b_{b2}) x_2 + \dots + (a_{bn} - b_{bn}) x_n$$

Outcomes of Prediction: Ergebnisse der Vorhersage
 Vergleich der Wirklichkeit und der Entscheidung:



Ergebnisse der Entscheidung:

		Entscheidung:		
		"negativ"	"positiv"	
Wirklichkeit: positiv	falsche Negative $P\{FN\} = FN / (P + N)$	valide Positive $P\{VP\} = VP / (P + N)$	Basisrate: $BR = P / (P + N)$	
	valide Negative $P\{VN\} = VN / (P + N)$	falsche Positive $P\{FP\} = FP / (P + N)$	Negativrate: $1 - BR = N / (P + N)$	
		$(1 - SR) = N' / (P' + N')$	Selektionsrate: $SR = P' / (P' + N')$	

Zusammenhang Wirklichkeit und Entscheidung (Validität: Vorhersage der Wirklichkeit aufgrund des Tests):

$$\phi_{yy'} = \frac{P(VP) - BR * SR}{\sqrt{BR * (1 - BR) * SR * (1 - SR)}}$$

Wahrscheinlichkeit der validen Positiven mit Testanwendung:

$$P(VP) = BR * SR + \phi_{yy'} * \sqrt{BR * (1 - BR) * SR * (1 - SR)}$$

Wahrscheinlichkeit der validen Positiven bei zufälliger Entscheidung:

$$P(VP) = BR * SR \quad (\text{bei Zufallszuweisung})$$

Erwarteter Nutzen der Ergebnisse der Entscheidung (expectet utility):

$$EU = U_t + N_{VP} * U_{VP} + N_{FN} * U_{FN} + N_{VN} * U_{VN} + N_{FP} * U_{FP}$$

EU ... erwarteter Nutzen (expectet utility)

U_t ... Nutzen der Testanwendung (oft negativ, da Test Geld kostet)

N_{VP} ... Anzahl valider Positiver

U_{VP} ... Nutzen valider Positiver

Der inkrementelle Nutzen:

Der Zuwachs an Nutzen durch die Testanwendung ergibt sich aus der Differenz zwischen Nutzen der Anwendung des Tests und Nutzen der zufälligen Entscheidung ohne Test.

Ist die Differenz negativ => keine Anwendung

Ist die Differenz positiv => Anwendung des Tests

10. Antworten zu den Übungsaufgaben

10.1. Einleitung

1. Ein Test ist ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale, mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung.
2. Bei der normorientierten Messung wird ein Messwert im Vergleich zur "Normpopulation" interpretiert. Hat eine Person einen Wert, der mehr als eine Standardabweichung über dem Mittelwert der Vergleichspopulation liegt, so hat sie eine überdurchschnittliche Ausprägung des gemessenen Merkmals.
Bei der kriterienorientierten Messung (Fricke, 1974) wird ein Messwert zur Vorhersage eines Kriteriums herangezogen. Hat zum Beispiel ein Schüler in einem Rechtschreibetest einen bestimmten Punktwert erreicht, wird verglichen, ob er die Leistung erreicht hat, die im Lehrplan festgelegt ist. Es wird ein Schluss von der Testleistung auf das Kriterium "Rechtschreibfähigkeit laut Lehrplan" gemacht.
3. Unter operationalisieren versteht man das "Uebersetzen" eines Begriffes oder Konstruktes in Operationen zur Messung, also die "Messbarmachung" von Variablen.
4. Bei physiologischen Messungen handelt es sich um Verhaltensbeobachtungen, da das Verhalten im Mittelpunkt steht. Biofeedback ist eine Methode zur Selbstbeobachtung (Rückmeldung) physiologischer Parameter, während in psychophysiologischen Untersuchungen eher Fremdbeobachtung vorliegt.
Man kann die physiologischen Messungen aber auch als "objektiven Test" im Sinne Cattells sehen, denn die Personen haben keine Verfälschungsmöglichkeit. Für den Fall der physiologischen Messung ist die Einordnung in Kerlingers Aufstellung der sozialwissenschaftlichen Datenerhebungsmethoden nicht eindeutig.

10.2. Historisches

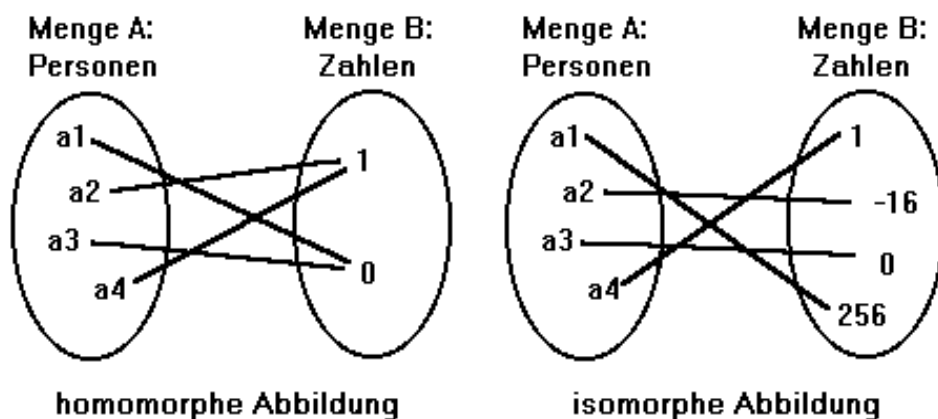
1. Binet und Simon (1904) entwickelten den ersten praktisch angewandten Intelligenztest und führten später den Begriff des "Intelligenzalters" ein. William Stern setzte das Intelligenzalter und das Lebensalter in Beziehung. Dieser Intelligenzquotient kam dann in Termans "Stanford-Binet-Test" zur Anwendung und hatte die Form:

$$IQ = \frac{\text{Intelligenzalter}}{\text{Lebensalter}} * 100$$

2. Beide benutzten unterschiedliche Varianten der Faktorenanalyse.

10.3. Grundlagen der Messung

- 1.



2. Messen ist die homomorphe Abbildung oder Repräsentation eines empirischen Relativs in ein numerisches Relativ.
3. Skalierung ist die Einrichtung einer Skala zur Durchführung einer Messung (Ahrens, 1974).

10.4. Klassische Testtheorie

1.

a)
$$\rho(A,B) = \frac{\mathbf{s}(A,B)}{\mathbf{s}(A) * \mathbf{s}(B)} = \frac{9}{4 * 5} = .45$$

b)
$$\rho(T_A, T_B) = \frac{\mathbf{r}(A,B)}{\sqrt{r_{ttA} * r_{ttB}}} = \frac{.45}{\sqrt{.50 * .85}} = .69$$

2. a) Man will ein sehr heterogenes Kriterium messen z.B.: Angst. Dieses Kriterium umfasst viele Facetten. Ein Test, der alle diese Facetten erfassen soll braucht ebenso sehr heterogene Items. Für eine hohe Reliabilität sind aber homogene Items die Voraussetzung. Man könnte sich eine Subskala z.B. Prüfungsangst vorstellen, die sehr reliabel gemessen werden kann.

b)

$$\rho(X,Y) = \frac{\sum_i \mathbf{r}(x_i, Y) * \mathbf{s}(x_i)}{\sqrt{\sum_i \sum_j \mathbf{s}(x_i) * \mathbf{s}(x_j) * \mathbf{r}(x_i, x_j)}}$$

Die Iteminterkorrelation zwischen x_i und x_j (steht im Nenner) sollte klein sein, um eine hohe Validität zu erzielen. Für die Reliabilität ist aber wichtig, dass die Items zuverlässig dasselbe messen, also hoch interkorrelieren.

3.

a) Retest-Reliabilität: Vorteil: Einfach durchzuführen, ohne Konstruktion eines neuen Tests.

Nachteile: Stabilität und Reliabilität werden vermischt. Zeiteffekte zwischen erster und zweiter Messung. Reaktive Effekte durch das mehrmalige Messen selbst.

b) Paralleltest-Reliabilität &

c) Odd-Even- oder Split-Half-Reliabilität: Vorteil: Einmalige Vorgabe des Tests. Nachteil: Hoher Konstruktionsaufwand. Voraussetzung ist wirkliche Parallelität der Testhälften.

d) Interne Konsistenz: Vorteile: Keine Konstruktion neuer Tests. Einmalige Vorgabe des Tests. Einfach durchzuführen.

4. a) Die Daten müssen Normalverteilt sein. b) Die Korrelationskoeffizienten und damit auch der Reliabilitätskoeffizient sind stichprobenabhängig. c) Die Daten müssen intervallskaliert sein.

5. Das Quadrat des Korrelationskoeffizienten r_{TX} ist die Reliabilität von Test X

6.

$$r_{tt}^2 = \frac{n * r_{tt}}{1 + (n - 1) * r_{tt}} = \frac{1/3 * .73}{1 + (1/3 - 1) * .73} = .47$$

7. Man bräuchte eigentlich nur den einen Prädiktor, da die anderen neun keine inkrementelle (zusätzliche) Validität bringen.

8. Formel: $r_{AB} = \frac{s_T^2}{s_A^2}$

$$s_T^2 = .70 * 25 = 17,5$$

$$s_X^2 = s_T^2 + s_F^2$$

$$s_F^2 = 7,5$$

9. Testverlängerung durch hinzufügen paralleler Items.

10. Die Reliabilität wird durch Korrelation berechnet. Korrelationen sind aber in hohem Masse stichprobenabhängig. Je grösser die Streuung in der Population, desto grösser wird die Korrelation. Die erste Gruppe hat eine grössere Streuung bzgl. des gemessenen Merkmals und somit eine grössere Reliabilität.

11. Uebereinstimmungsvalidität

12. Das was die Skala A an dem Kriterium aufklärt wird schon durch andere Teile der Testbatterie erfasst, so dass die Skala keine zusätzliche Aufklärung erbringt (sie hat keine inkrementelle Validität).

13. a) interne Konsistenz (Cronbach Alpha)
 b) Stabilitätskoeffizient (Retestreliaibilität)
 c) Parallelttestkoeffizient

14. $r_{tt} = 0$

15.

a) Reliabilitätsproblem

Bestimmungsart: Retestreliaibilität

b) Validitätsproblem

Bestimmungsart: Inhaltsvalidität (content validity)

c) Validitätsproblem

Bestimmungsart: Uebereinstimmungsvalidität (concurrent validity)

d) Validitätsproblem

Bestimmungsart: Konstruktvalidität (construct validity)/ Extremgruppenvalidierung

16. Der Korrelationskoeffizient zwischen zwei Halbttests (gemeint ist nicht die Spearman-Brown Formel) ist eine

Unterschätzung

der Reliabilität des Tests.

17. Nenne mindestens zwei Faktoren, die die Validität eines Tests beeinflussen:

a) Reliabilität des Tests

b) Reliabilität der Erfassung des Kriteriums

c) Kriterienauswahl

d) Korrelation zwischen Test und Kriterium

18. Welches wäre von den unten angegebenen Tests T1, T2, T3 die beste Kombination von 2 Tests zur Voraussage des Kriteriums C. Erkläre den Sachverhalt.

Korrelationsmatrix:

	C	T1	T2
T1	.43		
T2	.41	.72	
T3	.32	.04	.12

a) T1 & T3

b) Erklärung: T1 und T3 haben eine geringe Interkorrelation, messen also unterschiedliche Facetten des Kriteriums, während T1 und T2 weitgehend dasselbe messen.

19.

a) Formel:

$$n = \frac{r_{tt_n} * (1 - r_{tt})}{r_{tt} * (1 - r_{tt_n})} = \frac{.90 * (1 - .75)}{.75 * (1 - .90)} = 3$$

Der Test muss 3 mal verlängert werden, d.h. er hätte 45 Items. Es müssen also 30 Items zugefügt werden.

b) Konstruktionsaufwand/ Oekonomie/ Ermüdung der Probanden.

20. Formel:

$$sF(\text{Inter}) = s * \sqrt{2 * (1 - r_{tt})} = 15 * \sqrt{2 * (1 - .84)} = 8.5$$

$$sF(\text{Inter}) = 8.5$$

$$X_A - X_B = Z_{\text{krit}} * sF(\text{Inter}) = 1.96 * 8.5 = 16.6$$

Hans und Dieter unterscheiden sich um 15 IQ-Punkte. Um mit 95 % Sicherheit (Irrtumswahrscheinlichkeit 5%) von einem Unterschied sprechen zu können, müssten sich beide aber um mindestens 16.6 IQ-Punkte unterscheiden.

21. Wie gross ist die Reliabilität r_{tt} eines Tests, dessen Standardabweichung $s(x) = 15$ und dessen Standardmessfehler $s(F) = 15$.

$$r_{tt} = 0$$

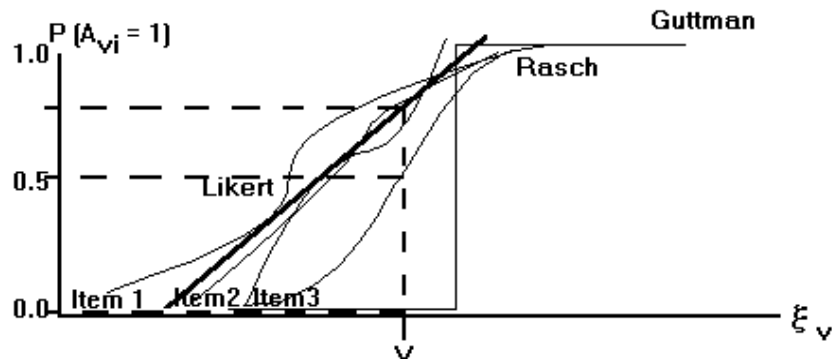
10.5. Probabilistische und klassische Testtheorie

1. Zwischen dem beobachtbaren Verhalten in der Testsituation und der latenten, nicht beobachtbaren Eigenschaft besteht ein korrelativer Zusammenhang. Eine Veränderung der latenten Eigenschaft hat eine Veränderung im Verhalten zur Folge.
2. Die klassische Testtheorie nimmt an, dass sich der beobachtete Wert aus dem "wahren" Wert und einem Fehleranteil zusammensetzt. Die probabilistische Testtheorie geht davon aus, dass die Wahrscheinlichkeit einer Reaktion einer Vp eine Funktion ist, aus der Fähigkeit der Person und der Schwierigkeit der Aufgabe.
3.
 - a) auf der Fähigkeitsebene: *beide gleich*
 - b) auf der Testscoreebene: *Hans*
 - c) auf der Fähigkeitsebene: *beide gleich*
 - d) auf der Testscoreebene: *Dieter*
- e) Der Vergleich ist nicht spezifisch objektiv, da er nicht unabhängig von der Itemauswahl ist. Die *spezifische Objektivität* wäre nur bei parallelen Itemkennlinien gegeben.
4. Falls man für das Gastarbeiterkind die Normen des Tests für die deutsche Population verwendet, kommt man zu folgendem Resultat:

Ueberschätzung der Schulleistung

(Bei gleicher Intelligenz des deutschen und des ausländischen Kindes würde man auf der Grundlage der deutschen Normen das ausländische Kind überfordern, und damit benachteiligen)

5. a)



b) Bsp.: In diesem Beispiel hat die eingetragene Person V eine Beantwortungswahrscheinlichkeit von 0 (0%) für das Guttman-Item, eine Wahrscheinlichkeit von .50 (50%) für das Rasch-Item und eine Wahrscheinlichkeit von .75 (75%) für die Likert-Skala.

10.6. Faktorenanalyse

1. Formeln: $s_m^2 = 1 = h_m^2 + b_m^2 + e_m^2$; $r_{tt} = h_m^2 + b_m^2$

$$\sum a_m^2 = h^2$$

1:	$h_m^2 = .73$	$b_m^2 = .07$	$e_m^2 = .20$
2:	$h_m^2 = .45$	$b_m^2 = .05$	$e_m^2 = .50$
3:	$h_m^2 = .81$	$b_m^2 = .04$	$e_m^2 = .15$
4:	$h_m^2 = .75$	$b_m^2 = .05$	$e_m^2 = .20$

a) Spezifität: b_m^2 Fehleranteil: e_m^2

b) Wahre Varianz: Varianz ohne Fehler: $h_m^2 + b_m^2$

2. Die klassische Testtheorie und die Faktorenanalyse enthalten folgende Konzepte, die man aufgrund ihres Bedeutungsgehaltes als nahezu identisch zuordnen kann:

Faktorenladung und Trennschärfe

3. Die folgende Aussage trifft für die Methode der Faktorenanalyse nicht zu:

Die Aufgabe besteht in der Bestimmung der Korrelation von Variablenpaaren.

10.7. Testkonstruktion

1.
 - a) Theorieteil mit Itembegründung
 - b) Testdurchführung/ -auswertung/ -interpretation
 - c) Stichprobenbeschreibung (Eichstichprobe)
 - d) Reliabilität
 - e) Validität
 - f) Literaturverzeichnis
 - g) Normentabellen

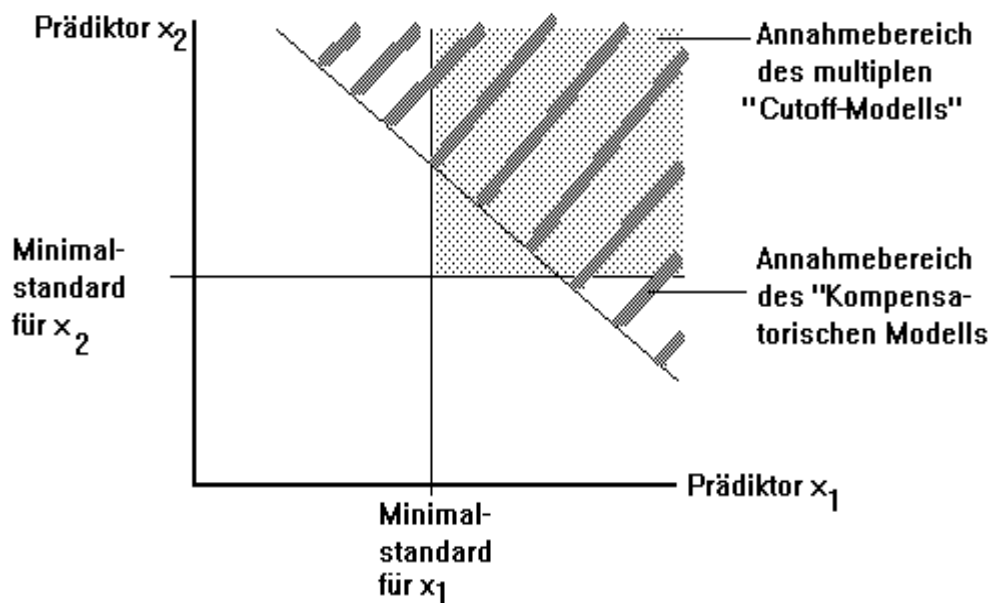
2. Items können anhand der Trennschärfe, der Itemschwierigkeit und der Homogenität ausgewählt werden. Dies sind formale Gesichtspunkte zur Selektion. Die Inhaltsvalidität (wie repräsentativ ist einen Frage für die Fragestellung) ist ein inhaltlicher Aspekte, der nicht berechnet aber beurteilt werden kann.

10.8. Entscheidungstheorie

1.

		Entscheidung:		
		"negativ"	"positiv"	
Wirklichkeit: positiv	falsche Negative $P(\text{FN}) = \frac{\text{FN}}{\text{P} + \text{N}}$	valide Positive $P(\text{VP}) = \frac{\text{VP}}{\text{P} + \text{N}}$	Basisrate: $\text{BR} = \frac{\text{P}}{\text{P} + \text{N}}$	
Wirklichkeit: negativ	valide Negative $P(\text{VN}) = \frac{\text{VN}}{\text{P} + \text{N}}$	falsche Positive $P(\text{FP}) = \frac{\text{FP}}{\text{P} + \text{N}}$	Negativrate: $1 - \text{BR} = \frac{\text{N}}{\text{P} + \text{N}}$	
		$(1 - \text{SR}) = \frac{\text{N}'}{\text{P}' + \text{N}'}$	Selektionsrate: $\text{SR} = \frac{\text{P}'}{\text{P}' + \text{N}'}$	

2. Das Cutoff-Modell ermöglicht eine bequeme Entscheidung, da der Bewerber einen Mindestwert erfüllen muss. Beim Kompensatorischen Modell können Schwächen durch Stärken ausgeglichen werden. Es kommt somit auf die genaue Fragestellung an, welches Modell besser geeignet ist. Bsp.: Zur Eignung als Pilot muss eine Mindestsehschärfe vorhanden sein (Cutoff-Modell). Zur Ausbildung als Chemielaborant kann man Schwächen im schulischen Wissen durch Stärken in der Konzentrationsleistung ausgleichen.



3. Je restriktiver der Cutoff-Punkt gewählt wird, umso mehr Personen, die in Wirklichkeit geeignet wären werden abgelehnt. Somit steigt die Anzahl falscher Negativer.
4. a) Klassifikationsproblem
- b) Die Testskalen X_1 und X_2 haben differentielle Validität, d.h. bringen etwas zur unterschiedlichen Zuweisung zu den Richtungen.
 - c) Die Testskala X_3 hat keine differentielle Validität, d.h. bringen nichts zur unterschiedlichen Zuweisung zu den Richtungen.

5. Die Anwendung einer Testbatterie bei einer Stichprobe ergab folgende Daten:

FN:	VP:	BR:
.20	.60	.80
VN:	FP:	
.10	.10	
	SR:	
	.70	

a) $BR = .80$; $SR = .70$.

b) Prozent valider Entscheidungen: $(VP + VN) \cdot 100 = 70\%$

Literaturverzeichnis

- Ahrens, H. J. (1974). *Multidimensionale Skalierung*. Weinheim: Belz Verlag.
- Ajzen, I. & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs: Prentice-Hall.
- Allport, G. W. (1935). Attitudes. In C. Murchinson (Ed.), *A handbook of social psychology*. Worcester, Mass.: Clark University Press, 798-844.
- Amelang, M. & Bartussek, D. (1981). *Differentielle Psychologie und Persönlichkeitsforschung*. Stuttgart, Berlin, Köln, Mainz: Kohlhammer.
- American Psychological Association (1954). Technical recommendations for psychological tests and diagnostic techniques. *Suppl. Psych. Bull.* 51.
- Amthauer, R. (1953). *Intelligenz-Struktur-Test*. Göttingen.
- Binet, A. & Simon, T. A. (1905). Méthodes nouvelles pour la diagnostic du niveau intellectuel des anormaux. *L'Année Psych.* 11, 191-336.
- Brickenkamp, R. (1975). *Handbuch psychologischer und pädagogischer Tests*. Göttingen: Hogrefe.
- Buros, O. K. (1938, 1941, 1949, 1953, 1959, 1965). *The Mental Measurement Yearbook*. Highland Park.
- Burt, C. & Williams, E. L. (1962). The influence of motivation on the results of intelligence tests. *Brit. J. statist. Psych.* 15, 129-135.
- Campbell, D. T. (1963). Social attitudes and other acquired behavioral dispositions. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 6). New York: McGraw-Hill, 94-172.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.

- Campbell, N. R. (1920). *Physics. The elements*. Cambridge: Univ. Press.
- Cattell, J. McKeen (1890). Mental tests and measurements. *Mind* 15, 373-381. American Psychological Association 1893
- Conrad, W. (1976). Erstellung von RASCH-Skalen für die Angstfragebogen FS 5-10 und KAT. *Diagnostika* 22, 110-125.
- Darwin, C. (1859). *On the origin of the species by means of natural selection*. London.
- Diehl, J. M. & Kohr, H. U. (1989). *Deskriptive Statistik (8. Aufl.)*. Eschborn bei Frankfurt am Main: Klotz.
- Drenth, P. J. S. (1969). *Der psychologische Test. Eine Einführung in seine Theorie und Anwendung*. München: Barth.
- Ebbinghaus, H. (1897). Über eine neue Methode zur Prüfung geistiger Fähigkeiten und ihre Anwendung bei Schulkindern. *Z. Psych.* 13, 401-459.
- Ellwein, T., Lippert, E. & Zoll, R. (1975). *Politische Betätigung in der Bundesrepublik Deutschland*. Göttingen.
- Esquirol, J. E. D. (1838). *Des maladies mentales considérées sous les rapports médical, hygiénique et médico-légal*. Paris.
- Fahrenberg, J. (1964). Objektive Tests. In R. Heiss(Hrsg.), *Handbuch der Psychologie in 12 Bänden, Band 6, Psychologische Diagnostik* (S.483-487). Göttingen: Hogrefe.
- Fahrenberg, J., Ewert, U. & Maier, N. (1987). *Reanalyse des Freiburger Persönlichkeitsinventars FPI*. Psychologisches Institut der Universität Freiburg i. Br.
- Fischer, G. H. (1968). Neue Entwicklungen in der Psychologischen Testtheorie. In: G. H. Fischer (Hrsg.), *Psychologische Testtheorie*. Bern: Huber.

- Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests. Grundlagen und Anwendungen*. Bern: Huber.
- Fricke, R. (1974). *Kriteriumsorientierte Leistungsmessung*. Stuttgart: Kohlhammer.
- Guilford, J. P. (1954). *Psychometric methods* (2nd ed). New York.
- Gutjahr, W. (1972). *Die Messung psychischer Eigenschaften*. Berlin: VEB Deutscher Verlag der Wissenschaft.
- Guttman, L. A. (1944). A basis for scaling qualitative data. *American Sociological Review*, 9, 139-150.
- Guttman, L. (1950). The basis of scalogram analysis. In: S. A. Stouffer (ed.), *Measurement and prediction*. Princeton, N. J.
- Hathaway, S. R. & McKinley, J. C. (1940, 1943). *Minnesota Multiphasic Personality Inventory*. Minneapolis.
- Hilke, R. (1980). *Grundlagen normorientierter und kriteriumorientierter Tests*. Bern: Huber.
- Hiltmann, H. (1966). *Kompodium der psychodiagnostischen Tests*. Bern.
- Jäger, A. O. (1986). Validität von Intelligenztests. *Diagnostica*, 32, Heft 4, 272-289.
- Jäger, R. S. (Hrsg.).(1988). *Psychologische Diagnostik: ein Lehrbuch*. München, Weinheim: Psychologie Verlags Union.
- Kerlinger, F. N. (1979). *Grundlagen der Sozialwissenschaften*. Weinheim: Belz Verlag. (Original erschienen 1964: *Foundations of Behavioral Research*).
- Klauer, K. J. (1987). *Kriteriumsorientierte Tests*. Göttingen, Toronto, Zürich: Hogrefe.
- Krantz, H. D., Luce, R. D., Suppes, P. & Tversky, A. (1971). *Foundations of measurement*. Vol I. New York, London: Academic Press.

- LaPiere, R. T. (1934). Attitudes vs. action. *Social Forces*, 13, 230-237.
- Lienert, G. A. (1967). *Testaufbau und Testanalyse*. Weinheim.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, No. 140.
- Lippert, E., Schneider, P. & Wakenhut, R. (1977). Zur Stabilität probabilistischer Skalierungsverfahren. *Psychologische Beiträge*, 19, 588-599.
- Murray, H. A. (1935, 1943). *Thematic Apperception Test*. Cambridge.
- Osgood, C. E., Suici, G. J. & Tannenbaum, P. H. (1953). *The measurement of meaning*. Urbano.
- Pawlik, K. (Hrsg.). (1976). *Diagnose der Diagnostik*. Stuttgart: Klett.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Kopenhagen: The Danish Institut of Educational Research.
- Raven, J. C. (1938). *Progressive Matrices*. London.
- Reicherts, M. (1985). Kriteriumorientierte Messung in der Klinischen Psychologie: Die Entwicklung eines Tests zur Belastungsbewältigung. *Zeitschrift für Klinische Psychologie, Psychopathologie und Psychotherapie*, 33, Heft 4, 313-336.
- Rorschach, H. (1921). *Psychodiagnostik*. Bern.
- Rosenthal, R. (1976). *Experimenter effects in behavioral research*. New York: Irvington.
- Rosenzweig, S. (1957). *Der Rosenzweig Picture Frustration-Test*. Göttingen: Hogrefe.
- Rost, J. & Spada, H. (1978). Probabilistische Testtheorie. In: K. J. Klauer (Hrsg.), *Handbuch der pädagogischen Diagnostik*, Band 1. S. 59-83.
- Sarason, S. B., Davidson, K., Lighthall, F. & Waite, R. (1958). A test anxiety scale for children. *Child Development* 29, 105-113.

- Spada, H. (unveröff.) *Einführung in die Testtheorie und Testkonstruktion*. Auszüge aus einem Lehrveranstaltungsskriptum von H. Spada & J. Rost unter Rückgriff auf Unterlagen von Scheiblechner.
- Spearman, C. (1910). Correlation calculated from faculty data. *Brit. J. Psych.* 3, 271-295.
- SPSS-X (1988). *User's Guide (3rd ed.)*. Chicago: SPSS Inc.
- Stern, W. (1900). *Über Psychologie der individuellen Differenzen*. Berlin.
- Stevens, S. S. (1951). Mathematics, measurement and psychophysics. In: S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York.
- Suppes, P. & Zinnes, J. L. (1963). Basic measurement theory. In R. D. Luce, R. R. Bush & E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol 1. New York: Wiley.
- Terman, L. M. & Merrill, M. A. (1937). *Measuring intelligence*. Boston.
- Terman, L. M. (1981). *The Stanford-revision and extension of the Binet-Simon scale of measuring intelligence*. Baltimore.
- Thomas, W. I., & Znaniecki, F. (1918). *The Polish peasant in Europe and America*. Vol. 1. Boston: Badger.
- Thorndike, R. L. (1910). *Personnel Selection*. New York.
- Thurstone, L. L. (1931). The measurement of attitudes. *Journal of Abnormal and Social Psychology*, 26, 249-269.
- Thurston, L. L. (1938). Primary mental abilities. *Psychometr. Monogr.* 1.
- Überla, K. (1968). *Faktorenanalyse. Eine systematische Einführung für Psychologen, Mediziner, Wirtschafts- und Sozialwissenschaftler*. Berlin.
- Upmeyer, A. (1985). *Soziale Urteilsbildung*. Stuttgart: Kohlhammer

Wechsler, D. (1939, 1946). *Wechsler-Bellevue intelligence-scale I and II*. New York.

Wiggins, J. S. (1973). *Personality and Prediction. Principals of personality assessment*.
Reading, MA: Addison-Wesley.

Wilkinson, L. (1988). *SYSTAT. The System of Statistics*. Evaston, IL: SYSTAT, Inc.

Wittmann, W. W. (1985). *Evaluationsforschung*. Berlin: Springer.

Wottawa, H. (1980). *Grundriss der Testtheorie*. München Juventa Verlag.

Anhang

Verzeichnis der Abbildungen

Abbildung 1: Zuordnung von Zahlen zu Objekten	25
Abbildung 3: Zuordnung von Zahlen nach ihrem Rang (Ordinalskalierung)	30
Abbildung 4: Itemschwierigkeit bei dichotomen und kontinuierlichen Items	37
Abbildung 5: Zusammenhang zwischen Gesamtsummenwert und Beantwortung von dichotomen und kontinuierlichen Items	38
Abbildung 6: Regression der Messwerte X auf die wahren Werte T.....	41
Abbildung 7: Praktische Reliabilitätsbestimmung	45
Abbildung 8: Variabilität bei der Messung psychologischer und physikalischer Eigenschaften.....	47
Abbildung 9: Zusammenhang der beobachtbaren Messungen und den nicht beobachtbaren latenten Dimensionen.....	56
Abbildung 10: Venn-Diagramm zur Erläuterung des "Attenuation Paradoxon".....	58
Abbildung 11: Itemcharakteristiken für verschiedene Testmodelle	67
Abbildung 12: Graphischer Modellgeltungstest des Rasch-Modells	71
Abbildung 13: Variablencluster im Raum (Diehl & Kohr, 1989, Abbildung 54)	77
Abbildung 14: Varimax-Rotation.....	78
Abbildung 15: Die Matrizen der Faktorenanalyse	79
Abbildung 16: Unrotierte Faktorenladungsmatrix und die klassische Testtheorie	81
Abbildung 17: Die Variablenvarianz in der Faktorenanalyse	82
Abbildung 18: Organisation der Daten für Computereingabe.....	97
Abbildung 18: Diagnostische Entscheidungen in der Psychologie	102
Abbildung 19: Vier Fälle bei der Personalentscheidung	103
Abbildung 20: Cutoff-Modell und Kompensatorisches Modell im Vergleich.....	104
Abbildung 21: Häufigkeitsverteilungen der negativen (N) und positiven (P) Personen, mit den möglichen validen (V) und falschen (F) Entscheidungen.	106

Verzeichnis der Tabellen

Tabelle 1: Sozialwissenschaftliche Datenerhebungsmethoden.....	7
Tabelle 2: Vier-Felder-Tafel.....	29
Tabelle 3: Kerlinger (1979, S. 671)	32
Tabelle 4: Skalen-Transformation nach Kokott (unveröffentlicht)	34
Tabelle 5: Stichprobenabhängigkeit von Korrelationskoeffizienten.....	59
Tabelle 6: Modellgeltungstest bei Guttman-Skalierung.....	68
Tabelle 7: (Diehl & Kohr, 1989, Tabelle 48).....	76
Tabelle 8: (Diehl & Kohr, 1989, Tabelle 51).....	80
Tabelle 9: Beispiel zum Diätverhalten.....	89
Tabelle 10: Vier-Felder-Tafel der möglichen Ergebnisse von Entscheidungen.....	107
Tabelle 11: Vier-Felder-Tafel der Wahrscheinlichkeiten für die möglichen Ausgänge von Entscheidungen.....	107

Index

A

Abbildungsregel, 27
absoluter Nullpunkt, 31
Absolutskala, 28;31
Absolutwerte, 51
Alkoholgefährdung, 14
Allgemeine Psychologie, 6
Allgemeiner Leistungstest, 11
American Psychological Association, 19
Annahme, 105
Anonymität, 84
Antworttendenzen, 84
apparatives Testverfahren, 15
Army General Classification Test, 22
Army-Alpha-Test, 21
Army-Beta-Test, 21
Assessment, 106
attenuation formulae, 56
Aufgabengütekriterien, 37
Axiom, 40

B

Basisrate, 107
Bedeutsamkeitsproblem, 27
bedingte Maximum-Likelihood-Schätzung, 70
Behandlungserfolg, 15
behavioral category, 88
Berufserfolg, 15
Beschreibung, 14
Bestimmung der Reliabilität, 45
Bewertung, 15
Biofeedback, 126
Biographie, 22

C

Crash-Test, 10
criterion of ambiguity, 86
Cronbach Alpha, 46
Cutoff-Modell, 104

D

Datenerhebungsmethode, 6
Datenkontrolle, 99
Definition eines Tests, 10
Determinanten, 66

Determinationskoeffizient, 42
deterministisches Modell, 67
deviation, 19
Diagnose, 102
Diagnostik, 22
dichotome Items, 47
Differentialdiagnose, 18
Differentielle Psychologie, 6
diskriminante Validität, 55
Disposition, 26

E

Eigenschaft, 26
Eignungstest, 11
Eindeutigkeitsproblem, 27
Einstellung, 85;89
Einstellungsmessung, 88
Einstellungsobjekt, 88
Einstellungstest, 11
empirisches Relativ, 27
Entscheidung, 14
Entscheidungstheorie, 102
Entwicklungstest, 11
equal-appearing intervall scale, 86
Ergebnis von Verhalten, 88
erschöpfende Statistik, 69
Erwartungswert, 39
Ethik, 12
expectet utility, 110

F

Fähigkeit der Person, 66
Fahreignung, 15
Faktorenanalyse, 55;76;99
Faktorenproblem, 79
Faktorenwerteproblem, 79
Fehlerkomponente, 40
Fehlervarianz, 40
Flächentransformation, 33
forced choice, 91
Fragebogen, 8
Fragebogenerhebung, 7
Fremdbeobachtung, 8
Funktion, 25
Funktionstest, 11

G

Gesamtvarianz, 40
Gewichte, 104
Graphologie, 22
Grundlagenforschung, 6;15
Gruppeneinteilung, 15
Gruppentest, 21

H

Habituation, 26;39
Handgeschicklichkeit, 15;54
Händigkeit, 11
Handlungsbereitschaft, 26
Häufigkeitsskala, 31
Häufigkeitsverteilung, 106
Hauptkomponentenmethode, 77
HAWIE, 23
HAWIK, 23
Homogenität, 11;38;67;85
Homomorphie, 27
Hypothese, 9;77
Hypothesentesten, 9

I

individualisiertes Testen, 71
individuelles Testen, 21
inhaltliche Validität, 36
Inhaltsanalyse, 7;8
Inhaltsvalidität, 53
inkrementeller Nutzen, 111
Instruktion, 11;20;84
Intelligenz, 11
Intelligenzalter, 20;127
Intelligenzforschung, 6
Intelligenzquotient, 23;127
Intelligenzstruktur, 23
Intelligenztest, 11
Interessentest, 11
interindividuelle Differenz, 6;48
Interkorrelation, 77
Interne Konsistenz, 36;45
Intervallskala, 28;30
Interview, 7;8;22
intraindividuelle Differenz, 48
Intransparenz, 8
Isomorphie, 27
IST, 23

Item, 11;13;37

Itemanalyse, 85

Itempool, 86

Itemsammlung, 86

Itemschwierigkeit, 37;69;85;101

Itemselektion, 85

K

Klassifikation, 11;102;105

Klassische Testtheorie, 36;39

Klinische Psychologie, 21

Klinischer Test, 11

Kommunalität, 82

Kommunalitätenproblem, 79

Kompensatorisches Modell, 104

Konfidenzintervall, 48

Konstrukt, 15;26

Konstrukt-Validität, 36

Konstruktvalidierung, 83

Konstruktvalidität, 53

Kontingenzkoeffizient, 28

konvergente Validität, 55

Korrelationsmatrix, 76

Kovarianz, 41

kriterienorientiert, 16

kriterienorientierte Messung, 16

kriterienorientierten Messung, 126

Kriteriumsvalidität, 36

Kuder-Richardson-20, 47

L

latente Eigenschaft, 10;36

Lebensalter, 127

Leistungstest, 11

Likelihoodquotiententest, 70

Likert Skala, 91

lineare Transformation, 30

logistische Funktion, 67

lokale stochastische Unabhängigkeit, 69

M

Matrizen der FA, 79

Median, 30

Messfehler, 40

Messinstrument, 13;26

Messmodell, 67

Messobjekt, 26

Messung, 10;24

- Messvorgang, 26
Minimalstandards, 104
Missing Data, 84
Mittelwert, 31
MMPI, 22;54
Modellgeltungstests, 70
monotone Transformation, 29
multiple Faktorentheorie, 20
multiple Regression, 103
multiple Selektion, 105
Multitrait-Multimethod Matrix, 55;83
N
natürliche Masseinheit, 31
Nominalskala, 28
Norm, 20
Normalverteilung, 33
Normiertheit, 36
normorientiert, 16
normorientierten Messung, 126
Normskala, 33
Normstichprobe, 60
Normwerte, 33
numerisches Relativ, 27
Nutzen, 102;110
Nützlichkeit, 37
O
objektive Tests, 7;8
Objektivität, 22;36;89
Odd-Even-Reliabilität, 45
Oekonomie, 37
offenen Fragen, 8
operationalisieren, 126
Operationalisierung, 15;26
Ordinalskala, 28
orthogonal, 77
Outcomes of Prediction, 106
P
paper-and-pencil test, 21
Parallellform, 21
Paralleltest, 42
Paralleltest-Reliabilität, 36;45
Parameter, 66
Parameterschätzung, 70
Personenfähigkeit, 69
Persönlichkeitsforschung, 6
Persönlichkeitstest, 11;22
Phänomenologie, 21
physiologische Messung, 17;126
Picture Frustration-Test, 8
Polung der Items, 84
Positivismus, 21
Prädiktor, 103
predictive validity, 53
probabilistische Testtheorie, 66
Produkt-Moment-Korrelation, 31
Progressive Matrices, 21
Projektive Verfahren, 7;8;22
Prüfungsangstskala, 54
Psychologische Diagnostik, 9
Psychologische Tests, 9
Psychometrie, 23
Psychomotorik, 11
psychophysiologische Untersuchung, 126
punkttrachorische Korrelation, 28
Q
Q-Methodologie, 7;8
R
Randsummenwerte, 107
Rangkorrelation, 30
Rangreihe, 29
Reaktivität der Messung, 39
Regression, 41;42
Reliabilität, 36;42;85
Reliabilität von Differenzwerten, 51
Repräsentationsproblem, 27
repräsentative Stichproben, 107
Retest-Reliabilität, 36;45
Rohwerte, 33;51
Rorschachtest, 8;22
Rotationsproblem, 79
S
Schulerfolg, 20
Schulfähigkeit, 11
Schulleistung, 11
Schulreife, 13
Schultest, 11
Schwierigkeit eines Items, 66
Selbstbeobachtung, 8
Selbstbeschreibung, 89
Selektion, 13;102;103

- Selektionsrate, 107
self report, 89
Semantisches Differential, 7;8;91
single action, 88
Skala, 11;24;27
Skalen, 7
Skalenniveau, 32
Skalentyp, 28
Skalenwert, 86
Skalierung, 28
soziale Erwünschtheit, 26
Sozialpsychologie, 85
Soziogramm, 91
Soziometrie, 7;8;91
Spearman-Brown Formel, 46
Spezifische Objektivität, 70
Spezifität, 82
Split-Half-Reliabilität, 45
Stabilitätskoeffizient, 45
Standardmessfehler, 47
Stanford-Binet-Test, 20
statistische Tests, 9
Stichprobenunabhängigkeit, 70
T
Test, 9;19
Testart, 11
Testdurchführung, 85
Testendform, 85
Testentwicklung, 84
Testentwurf, 84
Testhalbierungsmethode, 45
Testkonstruktion, 12
Testplanung, 84
Testscore, 51
Testtheorie, 8;12
Testverlängerung, 57
Testvorform, 85
Testwert, 39
Thematic Apperception Test, 22
Thematischer Apperzeptionstest, 8
Theorie of reasoned action, 88
Trait, 26
Transformation, 27;28;29;30;31
Transitivitätspostulat, 29
Trennschärfe, 38;85;101
U
Uebereinstimmungsvalidität, 53
Uniqueness, 82
Urteilsbildung, 14
V
valide Entscheidungen, 110
Validität, 36;53
Variablencluster, 77
Varianz, 31
Varianz einer Summe, 41
Varianzanalyse, 42
VARIMAX-Rotation, 77
veränderungssensitive Messinstrumente, 52
Verdünnungsformel, 56
Vergleichbarkeit, 36
Verhalten, 26;85;88
Verhaltensbeobachtung, 7;89
Verhaltenseigenschaften, 26
Verhältnisskala, 28;31
Voraussagevalidität, 53
Vorurteile, 12
W
wahre Varianz, 42
wahrer Wert, 39
Wechselwirkung, 26
wissenschaftliche Gütekriterien, 36
Woodworth Personal Data Sheet, 22
Worksheet, 97
Z
Zahl, 25
Zahlen, 12
Ziffer, 25
Zufallszuweisung, 108
Zuverlässigkeit, 22