

Fragebogen 2

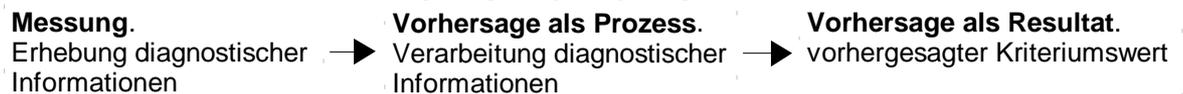
Diagnostischer Prozess

Ausarbeitung: Johannes Geffers [auf Grundlage einiger Vorarbeiten...]

1. Erläutern Sie die Bedeutung der Unterscheidung zwischen Messung und Vorhersage im Rahmen der diagnostischen Urteilsbildung (1.1)! Welches sind die Charakteristika der urteilenden Messung, welches die der mechanischen Messung (1.2), worin unterscheiden sich klinische von statistischen Vorhersagen (1.3)?

1. Bedeutung der **Unterscheidung zwischen Messung und Vorhersage** in der diagnostischen Urteilsbildung.

- Klinische und statistische Vorhersagen in der psychologischen Diagnostik haben einen gemeinsamen Ausgangspunkt, der sich anhand folgenden Schemas veranschaulichen lässt:



- Die Unterscheidung dieser beiden Phasen ist u.a. dann von Bedeutung, wenn beispielsweise die Validitäten der beiden Herangehensweisen in den jeweiligen Phasen und im Gesamt untersucht werden sollen.

2. **Messung** bzw. Informationserhebung. Gliederung der Informationserhebung in ›urteilend‹ (, korrespondiert mit klinischen Methoden, ›judgmental‹) und ›mechanisch‹ (korrespondiert mit statistischen Methoden, ›mechanical‹) (vgl. Sawyer).

a) Die Charakteristika der **urteilenden Messung**.

- Erhebungsmethoden ›frei‹ von regelhaften Beschränkungen; ganz ins Belieben des Diagnostikers gestellt (z.B. unstandardisiertes Interview)
- korrespondiert mit der klinischen Informationserhebung

b) Die Charakteristika der **mechanischen Messung**.

- Erhebungsmethoden folgen festgelegten Regeln, die empirisch überprüfbar sind (z.B. objektive Tests).
- korrespondiert mit der statistischen Informationsverarbeitung

3. **Vorhersage**. Unterscheidung zwischen statistischer und klinischer Prognose.

- Die Unterscheidung bezieht sich *nur* auf die Art und Weise der Informationsverarbeitung bzw. Datenkombination, nicht auf Unterschiede in der Datenerhebung (s.o.).

a) **klinische Vorhersage**.

- Es gehen nicht überprüfte, evtl. nicht einmal explizit formulierte Regeln in die Vorhersage ein.
- Sie werden immer dann getroffen, wenn Kriteriumsinformationen fehlen und die empirischen Zusammenhänge zwischen diagnostischen und kriterialen Informationen unbekannt sind. (Der klinisch vorgehende Diagnostiker verweist dann gewöhnlich auf seine Berufserfahrung oder Intuition.)

b) **statistische Vorhersage**.

- Allein die diagnostischen Informationen (Prädiktoren) gehen unter Bezugnahme auf empirisch gewonnene bzw. überprüfte Regeln und Gesetzmäßigkeiten in die Vorhersage des Kriteriums ein.
 - ◆ Zwei Arten statistischer Vorhersagen dominieren in der psychologischen Diagnostik:

- i. Vorhersagen mit Hilfe von **Erwartungstabellen** (aktuarische Vorhersagen).
- Prognose von diskreten Kriteriumsinformationen
 - Auf Grundlage eines Satzes diagnostischer Informationen werden in einer Clusteranalyse taxonomische Klassen gebildet, die
 - in sich homogen sind
 - sich gegenseitig ausschließen
 - eine zuverlässige Klassifikation von Individuen erlauben
 - erst dann werden empirisch die Wahrscheinlichkeiten bestimmt, mit denen für jede einzelne taxonomische Klasse bestimmte Kriteriumsdaten zu erwarten sind
- b) Vorhersagen mit Hilfe von **multiplen Regressionsgleichungen**.
- V.a. Vorgehen bei quantitativen Kriteriumsinformationen.
 - Im einfachsten Fall gilt folgende Prognosegleichung:

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

x_1, \dots, x_n Stehen als *Variablen* für die Werte von n Prädiktoren (diagnostischen Informationen).

y'_c Steht für den mit Hilfe dieser Linearkombination *vorhergesagten Kriteriumswert*.

${}_c b_1, \dots, {}_c b_n$ Sind die sogenannten *Regressionskoeffizienten oder -gewichte*.

◆ **Interpretation der sogenannten Regressionskoeffizienten:**

- ${}_c b_1$ gibt z.B. an, um welchen Wert y'_c wächst, wenn x_1 um eine Einheit zunimmt und alle anderen Werte x_2, \dots, x_n konstant bleiben.)
- Diese Gewichte werden durch ein mathematisches Optimierungsverfahren so bestimmt, dass die Summe der *quadrierten Abweichungen der vorhergesagten Kriteriumswerte, der y'_c , von den tatsächlichen Kriteriumswerten, den y_c , ein Minimum ergibt.*

◆ **Verdeutlichung des statistischen Charakters einer so getroffenen Vorhersage:**

- y'_c stellt nur eine mehr oder weniger genaue Schätzung des tatsächlichen Kriteriumswertes dar. Deshalb ist ein Vertrauensintervall um den geschätzten Wert herum anzugeben, in dem der tatsächliche Kriteriumswert mit einer bestimmten Wahrscheinlichkeit liegt.

- **Übergang von klinischen zu statistischen Aussagen:** Erst durch die Ersetzung der subjektiven Zusammenhangsvermutungen durch empirisch geprüfte und bewährte Aussagen sind statistische Vorhersagen möglich.

2. Welchen Einfluss haben Ausbildung und Berufserfahrung auf die Genauigkeit diagnostischer Urteile? Lässt sich die Genauigkeit durch spezielles Training verbessern? Was folgt für die diagnostische Praxis aus der Befundlage?

- **Ausbildung und Berufserfahrung** haben einen gewissen Einfluss auf die Genauigkeit diagnostischer Urteile (Ausgangswerte: ›Naive‹: 52%, ›Mittlere‹ & ›Experten‹: 65%).
- Durch **Training** lässt sich die Genauigkeit diagnostischer Urteile nur eingeschränkt – auf vorher ›Naive‹ - verbessern (um 6%). Auf Graduierte der Psychologie und ›Experten‹ konnte kein Einfluss eines Trainings festgestellt werden.
- Menschen sind schlechte Informationsverarbeiter durch die begrenzte Informationsverarbeitungskapazität des Kurzzeitgedächtnisses und seines geringen Speichervermögens. Dies lässt sich durch Erfahrung und Training nicht wesentlich verbessern.
- Diagnostische Urteile erfordern i.d.R. die Aggregation einer Vielzahl unterschiedlicher Informationen (z.B. Laborwerte, Testergebnisse, biographische und anamnestiche Daten).
- Daher sollte der Diagnostiker überall da, wo es möglich ist, ersetzt werden durch ein aktuarisches bzw. statistisches Vorgehen bei der Informationsverarbeitung. Ihm sollten **formale Hilfsmittel für die Datenaggregation** zur Verfügung gestellt werden (z.B. statistische Verfahren wie BAYES'scher Algorithmus, lineare Regressionsmodelle oder aktuarische Verfahren (Erwartungstabellen))
- ◆ Eine **Studie von GOLDBERG** (1968) zeigt, dass Ausbildung und Berufserfahrung nur begrenzten Einfluss auf die Genauigkeit diagnostischer Urteile haben.

Gruppe	Beginn	nach 17 Wochen Training
Naive	52,00%	58,00%
Mittlere	65,00%	65,00%
Experten	65,00%	65,00%

- ◆ Ergebnisse:
 - Unterschiede zwischen Experten und Naiven nicht sehr groß.
 - Durch Training keine weitere Verbesserung; nur durch Wertetraining: dauerhafte Zunahme der Urteilsgenauigkeit auf knapp 70%.
 - ◆ **Wertetraining** (erfolgreich).
 - Zurverfügungstellung der numerischen Werte der Formel ($L + Pa + Sc - Hy - Pt$) für jedes einzelne Profil und die optimalen cutting scores.
 - Mitteilung, dass die Formel zu einer Genauigkeit von etwa 70% führt und für Extremwerte gültiger ist als für Werte nahe dem cutting score.
 - Es wurde ihnen freigestellt, auch andere Wege zu nutzen, um die durch die Formel erreichte Gültigkeit noch zu verbessern • kriteriale Information.
 - ◆ **erfolglose Trainings:**
 - Gruppentraining: 2 Urteiler arbeiten zusammen und einigen sich auf ein Urteil.
 - Generalisierungstraining: Training an noch nicht verwendeten Profilen, das v.a. früher gezeigte Schwächen beseitigen sollte.
 - Formeltraining: Formel zur Verbesserung ihrer Genauigkeit als Indikator für die Skalen, die v.a. zu berücksichtigen wären.

3. Woher beziehen die Befürworter des statistischen Ansatzes (z.B. Sarbin oder Lundberg) ihre Sicherheit, dass ihr Ansatz dem klinischen Ansatz auf jeden Fall ebenbürtig, meist sogar deutlich überlegen sein wird?

- Aus **mathematischen Gründen gilt das statistische Verfahren als optimal** (Berechnung der Regressionsgewichte so, dass die Summen der Abweichungsquadrate ein Minimum ergibt).
- LUNDBERG bestritt ganz allgemein die Wissenschaftlichkeit der am Einzelfall orientierten **klinischen Vorgehensweise**. Er, aber auch SARBIN, sahen in ihr lediglich eine Vorstufe zur statistischen und im eigentlichen Sinne wissenschaftlichen Vorhersagemethode.
- ◆ Die **klinische Methode**.
 - Klinisches Prognostizieren sei nur eine informelle, vergleichsweise unsystematische und grobe Version des statistischen Vorgehens.
 - Kliniker gehen im Prinzip genauso vor wie die Statistiker: Auch er könne nicht umhin, den Einzelfall bestimmten Klassen zuzuordnen und aufgrund bestimmter – allerdings von ihm selbst subjektiv geschätzter – Wahrscheinlichkeiten eine Vorhersage zu machen.
 - Es wird **implizit ein Urteil aufgrund der Modellierung der Daten in Form einer linearen Regressionsgleichung** vorgenommen, wobei die **Regressionsgewichte ebenfalls intuitiv geschätzt** werden; bzw. es erfolgt eine Zuordnung zu Klassen und darauf basierende subjektive Wahrscheinlichkeitsschätzungen.
- ◆ Die **statistische Methode**.
 - Die statistische Methode berechnet die **Regressionsgewichte** so, dass **die Summen der Abweichungsquadrate ein Minimum ergibt**. Sie werden unter Einhaltung mathematischer Optimalitätsgesichtspunkte bestimmt. Jedes Abweichen von diesen Gesichtspunkten muss zu einem Absinken der Vorhersagegenauigkeit führen.
- Schlussfolgerung.
 - Der **Kliniker kann bei seiner intuitiven Schätzung der Regressionsgewichte bestenfalls so gut sein wie die statistische Methode**, wenn es ihm gelingt, die optimale Gewichtung zu treffen, in allen anderen Fällen wird er schlechter abschneiden, weil er suboptimal gewichtet. Schon aus theoretischen Gründen wäre demnach der Kliniker dem Statistiker gegenüber im Nachteil.
 - WESTMEYER unter Bezugnahme auf SARBIN (1941): „Das, was die statistische Methode explizit und unter Bezugnahme auf empirisch überprüfte Zusammenhänge tue, mache der klinisch vorgehende Prognostiker implizit und unter Rückgriff auf ungeprüfte und deshalb suboptimale Zusammenhangsvermutungen.“

4. In welcher Weise trägt die Unterscheidung zwischen dem Entdeckungs- und dem Begründungszusammenhang einer diagnostischen Aussage zur Klärung der Kontroverse ›Kliniker‹ und ›Statistiker‹ bei?

- Die **klinische Methode** hat ihre Berechtigung vor allem im Rahmen des **Entdeckungszusammenhangs** hat: Sie kann dazu beitragen, dass Individuum besser zu verstehen und ist von zentraler Bedeutung im Prozess der Hypothesenbildung.
- Die klinische Methode ist dagegen im Hinblick auf den **Begründungszusammenhang** der **statistischen Methode** unterlegen: Implizit liegt ihr [der klinischen Methode] ein statistisches Rationale zugrunde; die entsprechenden Gewichte (Regressionskoeffizienten) werden aber nur intuitiv geschätzt, daher kann sie in der Vorhersage des Kriteriums bestenfalls genauso gut sein wie die statistische Methode, i.d.R. wird sie aber schlechter sein. Die **statistische Methode ist im Hinblick auf den Begründungszusammenhang der klinischen überlegen**.
- Dadurch, dass auch der klinischen Methode implizit ein statistisches Rationale zugrunde liegt, ist der Gegensatz zwischen ›klinischer‹ und ›statistischer‹ Methode nicht ganz klar durchzuhalten.
- ◆ **Entdeckungszusammenhang.**
 - Informationserhebung
 - Verständnis des Individuums
 - Einblicke in die Persönlichkeit
 - Prozess der Hypothesenbildung
- ◆ **Begründungszusammenhang.**
 - die Informationsverarbeitung und Datenaggregation zur Vorhersage des Kriteriums
 - Begründung einer Vorhersage
 - bezieht sich auf die Richtigkeit von Vorhersagen
- **Hintergrund:** HORST et al. (1941) befassten sich mit der Behauptung von ALLPORT, statistische Methoden seien zwar für die Vorhersage von Populationscharakteristika geeignet, versagen aber am Einzelfall. Ihre Schlussfolgerung:
 - Die **klinische bzw. fallorientierte Methode** sei ein brauchbares Verfahren, um ein Individuum besser zu verstehen. Sie sei unentbehrlich in Situationen, in denen anerkannte funktionale Zusammenhänge zwischen Prädiktoren und Kriterium fehlen und im Prozess der Hypothesenbildung von zentraler Bedeutung.
 - Ihr liege aber **implizit ein statistisches Rationale** – v.a. im Hinblick auf den Begründungszusammenhang – zugrunde, so dass ein wirklicher Gegensatz nicht bestehe.

5. Welche zur statistischen Vorgehensweise alternative Form der Datenkombination und Urteilsbildung verteidigt – zumindest als Denkmöglichkeit – Meehl (1954)? Finden Sie seine Argumente überzeugend? Wie ließe sich prüfen, ob diese Form tatsächlich dem Vorgehen „klinischer“ Diagnostiker zugrunde liegt?

- MEHHL [Herleitung: s.u.] vertritt die Ansicht, dass im Rahmen der klinischen Methode auf der Basis von Persönlichkeitstheorien und diagnostischen Informationen **ideographische Theorien über das einzelne untersuchte Individuum** konstruiert werden könnten.
- Aus dieser unter Umständen deterministischen Theorie ließen sich wiederum Konsequenzen in bezug auf das bei dieser Person zu erwartende Kriteriumsverhalten ableiten.
- Die Argumente erscheinen zumindest plausibel – und ein derartiges Vorgehen ist in der Praxis durchaus üblich (Bsp.: ›Komparative Kasuistik‹ von JÜTTEMANN).
- Ob Kliniker tatsächlich ein Vorgehen über die Entwicklung strukturell-dynamischer Hypothesen wählen, ließe sich mittels **empirischer Vergleiche** untersuchen.
- Untersuchungen auf der Grundlage des **Linsenmodells** / der **Linsengleichung (von TUCKER)** zu prognostischen Aktivitäten haben ergeben:
 - Prognostizieren lässt sich im Rahmen dieses Modells ausgezeichnet paramorph beschreiben.
 - Hinweise auf konfigurale Formen der Datenverarbeitung wurden zwar gefunden, die Überlegenheit dieser nichtlinearen Modelle, sofern sie überhaupt nachgewiesen werden konnte, war aber vernachlässigenswert.
 - Nach GOLDBERG (1968) wird ein einfaches lineares Modell gewöhnlich in der Lage sein 90-100% der reliablen Urteils-Varianz der klinischen Urteilsbildung aufzuklären.

- ◆ **Vorgehen** (Linsengleichung von TUCKER):
 - **Prüfung**, ob die Vorhersagen des Klinikern durch ein lineares Modell abbildbar ist.
 - Mit Hilfe der **Linsenmodellgleichung** von TUCKER ist feststellbar, ob die Informationsverarbeitung auf konfigurale, also nicht-lineare Art vor sich geht.
 - Die Gleichung zerlegt die Validität des Diagnostikers.

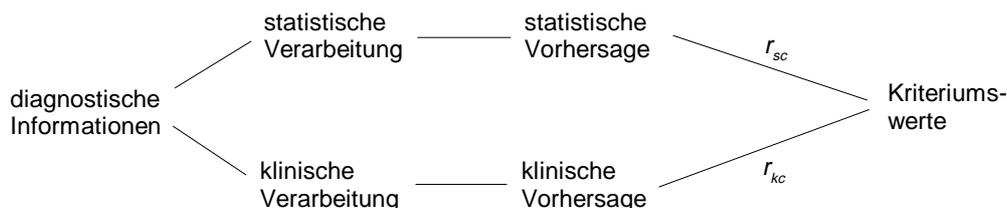
$$r_{y_c y_j} = G R_c R_j + C \sqrt{1 - R_c^2} \cdot \sqrt{1 - R_j^2}$$

- Die Gleichung enthält eine nichtlineare Komponente der Urteilsgenauigkeit **C**.
 - **C** entspricht der Korrelation zwischen dem Residualwertwerten des Kriteriums und den Residualwerten der Vorhersage des Urteilers, nachdem die lineare Komponenten beim Kliniker und beim Kriterium abgezogen worden sind:

$$C = r_{(y_c - y'_c)(y_j - y'_j)}$$
- **MEEHLS** (1954) Buch *Clinical versus statistical prediction* enthält v.a. eine theoretische Analyse der klinischen Vorgehensweise. Diese steht im starken Kontrast zu der zuvor skizzierten These von SARBIN und LUNDBERG.
- **Vorwurf: Vermischung von Entdeckungs- und Begründungszusammenhang.** Es sei unzulässig, dass aktuarische oder statistische Modell aus dem Begründungszusammenhang, in dem es mit dem klinischen Vorgehen verglichen werden könne, auf den Entdeckungs-zusammenhang zu übertragen
 - ◆ **statistisches Vorgehen.**
 - arbeitet mit empirischen Verallgemeinerungen
 - operiert auf der Ebene einfacher Input-Output-Relationen
 - ◆ **klinisches Vorgehen.**
 - Kliniker können zu ihrem Urteil auf völlig andere Weise kommen als Statistiker.
 - Vorhersagen auch auf Basis einer Persönlichkeitstheorie (Kenntnis der internen Mechanismen (Lampen-Apparat)).
 - Konstruktion einer idiographischen Theorie für das einzelne Individuum aufgrund der diagnostischen Informationen im Rahmen einer solchen Persönlichkeitstheorie.
 - Ableitung von Konsequenzen auf das bei dieser Person zu erwartende Kriteriumsverhalten.
 - Über diese fallbezogene Erfindung strukturell-dynamischer Hypothesen sei es dann möglich, jedes Individuum für sich genauer zu erfassen, als es unter Bezugnahme auf eine Durchschnittsnorm bezogene Gleichung erfolgen kann.
- Das klinische Vorgehen nicht gleich dem statistischen (nur suboptimal), sondern grundlegend anders.

6. Unter welchen Umständen und in welcher Weise lassen sich klinische und statistische Vorhersagen miteinander vergleichen?

- Vorhersagen lassen sich im Hinblick auf den **Begründungszusammenhang** miteinander vergleichen. **Dieselben diagnostischen Informationen** müssten einmal zu einer **statistischen** (d.h. auf empirisch überprüfte Zusammenhänge basierende) und einmal zu einer **klinischen Vorhersage** des Kriteriums verarbeitet werden. Durch **Korrelation der jeweils vorhergesagten mit den tatsächlichen Kriteriumswerten** lässt sich die Überlegenheit einer der beiden Methoden demonstrieren (Vorgehen nach MEEHL: Vergleich der Validität klinischer und statistischer Vorgehensweise).
- Voraussetzung ist, dass die entsprechenden Kriteriumsinformationen für die statistische Vorhersage vorhanden sind!



7. Welche Ergebnisse haben die von Meehl analysierten Vergleichsuntersuchungen erbracht? Haben sich aus den Untersuchungen von Ganzach irgendwelche neuen Aspekte ergeben?

◆ Ergebnisse von MEEHL:

- In der Mehrzahl der Fälle war die **statistische Vorhersagemethode der klinischen Überlegen**.
- In einer geringen Zahl an Fällen waren beide Methoden gleich gut.
- In einem – methodisch fragwürdigen – Fall war die klinische Methode besser als die statistische.
- Seit 1966 sind eine Vielzahl weiterer Untersuchungen erschienen, ohne dass sich an den Schlussfolgerungen etwas geändert hat. Die Überlegenheit des statistischen Vorgehens wurde nur noch nachdrücklicher dokumentiert und die Hoffnung, die sich mit den theoretischen Überlegungen von Meehl verband, nachhaltig enttäuscht.

◆ Ergebnisse von GANZACH (1995: Reanalyse eines Datensatzes, der von MEEHL 1954 erhoben wurde). [13 klinische Psychologen, 16 klinische Psychologen in Ausbildung; 861 MMPI-Profile, Diagnosestellung (psychotisch oder nicht)]:

- Datensatz enthält Hinweise, dass die Urteile der 29 Psychos unter Verwendung nichtlinearer Strategien gebildet werden.
- Frühere Versuche durch nichtlineare Regressionsgleichungen diese vermutete nichtlineare Beziehung zwischen Prädiktor (MMPI-Profile) und Kriterium (Diagnose) abzubilden erbrachten keine konsistenten Ergebnisse.
- GANZACH schlägt ein anderes (neues) mathematisches Modell vor zur Erhöhung des Anteils der aufgeklärten Varianz an der Gesamtvarianz: Das **Scatter-Modell**. Ergebnisse:
 - Das **Scatter-Modell** ergibt eine bessere Modellanpassung als das lineare Modell (bei 24 v. 29 Urteilen).
 - Ein komplexes, nichtlineares, d.h. konfigurales Modell erwies sich in der Vorhersage des Kriteriums dem einfachen linearen Regressionsmodell als überlegen (bessere Modellanpassung, bessere Varianzaufklärung).
 - Geringer Unterschied in der Varianzaufklärung zwischen diesem komplexen Modell und dem einfachen linearen Modell: das lineare Modell konnte 94% der Urteilsvarianz des komplexen Modells aufklären.
 - Bestätigung von GOLDBERGS Aussage, dass sich 90-100% der Urteilsvarianz in den klinischen Vorhersagen durch ein lineares Modell abbilden lassen.
 - Die Daten weisen auf eine **konsistente Nutzung einer nichtlinearen Strategie** hin: disjunktive Urteilsstrategie, d.h. die Kliniker gewichten die pathologischen Skalenergebnisse stärker.
 - GANZACH selbst sieht seine Ergebnisse zwar als von geringem praktischen Wert (Zuwachsvalidität), aber von großer Aussagekraft hinsichtlich der den **Urteilen zugrundeliegenden kognitiven Prozessen**.

■ GANZACH schlägt ein anderes **mathematisches Modell** vor zur Erhöhung des Anteils der aufgeklärten Varianz an der Gesamtvarianz.

◆ **Lineare Entscheidungsstrategie**.

- Wie in der multiplen Regression werden verschiedene Prädiktoren mit einer unterschiedlichen Gewichtung aufaddiert

$$\hat{y} = \alpha + \sum_{i=1}^m \beta_i X_i$$

◆ **Nichtlineare Strategien:**

- Diese werden unterschieden in disjunktive und konjunktive Strategien. Bei beiden ist das Gewicht eines Prädiktors abhängig von seiner Ausprägung relativ zu den anderen Prädiktoren (Fokus auf die Unterschiede zwischen einzelnen Werten bei einem Patienten, d.h. um sein Profil)
 - **disjunktive Strategie**. Das Urteil basiert primär auf dem am höchsten ausgeprägten Prädiktor.
 - **konjunktive Strategie**. Das Urteil basiert primär auf dem am niedrigsten ausgeprägten Prädiktor.
- ◆ Bsp.: Evaluation des Schweregrades der Störungen von 2 psychisch kranken Patienten auf der Basis von 2 gleichermaßen wichtigen Testscores.

- Beide Patienten haben den gleichen Mittelwert, aber einer hat zwei moderate Werte, während der andere einen hohen und einen niedrigen hat
 - kompensatorische Strategie der Urteilsbildung: Beide Patienten werden etwa gleich bewertet, da Kompensation der einen Variablen durch die anderen im Regressionsmodell möglich ist.
 - disjunktive Strategie: Das Urteil richtet sich nach dem am höchsten ausgeprägten Prädiktor.
 - konjunktive Strategie: Das Urteil richtet sich nach dem am niedrigsten ausgeprägten Prädiktor.
- GANZACH entwickelt 4 verschiedene Modelle, die alle sowohl disjunktive als auch konjunktive Strategien modellieren können.
- **Die Stärke seines Ansatzes: konjunktive, lineare und disjunktive Entscheidungsstrategien werden als ein auf einer Dimension befindliches Kontinuum dargestellt** werden. Dies geschieht, indem
 - der nichtlineare Teil der Regressionsgleichung entweder ein negatives Gewicht erhält (• konjunktive Strategie) oder ein positives Gewicht (• disjunktive Strategie),
 - Nimmt die Gewichtung des Terms den Wert Null an, nimmt der gesamte nichtlineare Teil der Regressionsgleichung ebenfalls den Wert Null an (• lineare Strategie).
 - Der Betrag des Gewichts gibt an, in welchem Ausmaß nichtlinear geurteilt wird.
- ◆ Eines der 4 Modelle: Das **Simple Scatter Modell**.
 - Es wird ein sogenannter Scatter Term (SXT) gebildet, mit dem das oben genannten Gewicht (•) multipliziert und zu der bekannten linearen Gleichung hinzuaddiert wird.
 - In den Scatter Term gehen die Abweichungen der 11 Skalenausprägungen von ihrem Mittelwert ein.
 - Das kann dazu führen, dass ein einziger hoher oder niedriger Skalenwert dem Scatter Term einen hohen Wert gibt.
$$\hat{y} = \alpha + \sum_{i=1}^m \beta_i X_i + \delta SXT \quad SXT = \sqrt{\sum_{i=1}^m (X_i - \bar{X})^2} \quad \bar{x} = \sum_{i=1}^m \left(\frac{X_i}{n} \right)$$
 - Bei disjunktivem Urteil nimmt • einen positiven Wert an und erhöht somit den Wert von y (• Diagnose: psychotisch).
 - Bei konjunktivem Urteil nimmt • einen negativen Wert an. Der Wert des Scatter Terms wird von dem linearen Anteil abgezogen und senkt also den Wert von y (• Diagnose: nicht psychotisch)
- Dieses Prinzip gilt grundsätzlich für alle 4 Modelle, wird aber jeweils etwas abgewandelt!

8. Geben Sie jeweils ein Beispiel für acht Vorhersagemethoden, die Sawyer (1966) unterscheidet! Welche Ergebnisse hat Sawyers Analyse erbracht, und welche Konsequenzen lassen sich daraus für die diagnostische Praxis ableiten?

Vorhersagemethode	Datenerhebung	Datenverarbeitung	Beispiel
1. rein klinische Vorhersage 20%	urteilend	klinisch	unstandardisiertes Interview ohne Verwendung obj. Tests, klin. Urteil
2. Eigenschaftsrating 46%	urteilend	statistisch	anhand eines unstandardisierten Interviews Einschätzung auf einer obj. Persönlichkeitsdimension; dann: statistische Kriteriumsvorhersage aufgrund eines nachgewiesenen Zusammenhangs zwischen Kriterium und Dimension
3. Profilinterpretation 38%	mechanisch	klinisch	aus obj. Tests gewonnene diagnostische Informationen werden nach festgelegten Regeln zu einem Profil vereinigt, das in eine klinische Vorhersage eines Kriteriums eingeht.
4. rein statistische Vorhersage 62%	mechanisch	statistisch	obj. Testdaten werden zu einer linearen Regressionsgleichung der Kriterienwerte herangezogen
5. klinische Kombination 26%	urteilend und mechanisch	klinisch	diagnostische Informationen aus obj. Tests und einem unstandardisierten Interview werden zum klin. Urteil verarbeitet
6. mechanische Kombination 75%	urteilend und mechanisch	statistisch	Diagnostische Informationen aus obj. Tests und standardisiertem Interview werden nach festgelegten emp. Prüfbareren Regeln zur statistischen Vorhersage des Kriteriums herangezogen.

7. klinische Synthese 50%	urteilend und mechanisch	Zunächst statistisch, dann klinische Verarbeitung der Ergebnisse der ursprünglichen Datenerhebung unter Berücksichtigung der Ergebnisse der statistischen Datenverarbeitung	Zunächst: Statistische Verarbeitung der diagnostischen Informationen aus unstandardisiertem Interview; Dann: Berücksichtigung dieses Ergebnisses und der ursprünglichen diagnostischen Informationen bei klinischen Kriteriums- vorhersage
8. mechanische Synthese 75%	urteilend und mechanisch	Zunächst klinisch, dann statistische Verarbeitung d. Ergebnisse der ursprgl. Datenerhebung unter Berücksichtigung d. Ergebnisse d. klin. Datenverarbeitung	Zunächst: klein. Verarbeitung der diagn. Informationen aus obj. Tests u. unstandardisiertem Interview; Dann: Berücksichtigung dieses Ergebnisses u. der urspr. diagnostischen Informationen bei der statistischen Kriteriums- vorhersage

◆ Interpretation & Konsequenzen.

- Schon die **rein statistische Vorhersage** ist erfolgreicher als alle Vorhersagemethoden, die eine klinische Verarbeitung diagnostischer Informationen involvieren.
- Die **mechanische Kombination ist der mechanischen Synthese ebenbürtig.** Offenbar trägt das in die mechanische Synthese eingehende Resultat der klinischen Informationsverarbeitung nicht zu einer Verbesserung der Vorhersageleistung bei.
- Die **mechanische Kombination ist der rein statistischen Vorhersage überlegen.** Die Einbeziehung urteilend erhobener diagnostischer Informationen zusätzlich zu den objektiven (mechanisch erhobenen) Daten führt offenbar zu einer Verbesserung der Vorhersagegenauigkeit.
- Für die diagnostische Praxis lässt sich daraus die **Konsequenz** ableiten, dass der **Kliniker ein schlechter Informationsverarbeiter** ist, **im Rahmen der Datengewinnung aber durchaus seinen Platz haben sollte.**

◆ J. SAWYERS Vorgehen.

- wesentlich differenzierteres Vorgehen als bei MEEHL
- unterscheidet wie MEEHL zwischen statistischer und klinischer Informationsverarbeitung
- Bestimmung des Erfolgswertes eines Vorgehens:

$$\text{Erfolgswert der Methode I} = \frac{\text{Anzahl der Vergleiche, in denen Methode I den übrigen Methoden überlegen ist,} + \frac{1}{2} * \text{Anzahl der Vergleiche, in denen Methode I den übrigen Methoden ebenbürtig ist,}}{\text{Anzahl der Vergleiche zwischen Methode I und den übrigen Methoden.}}$$

9. Vergleichen Sie die Ergebnisse von Grove et al. mit denen von Meehl und Sawyer!

- GROVE et al. (2000): Meta-Analyse zum Vergleich der Genauigkeit klinischer vs. mechanischer Datenaggregation für die Kriteriumsvorhersage; Analyse von 136 Studien.
- ◆ Zusammenfassung der **Ergebnisse der Studie** von GROVE.
 - Im Durchschnitt waren **mechanische Vorhersagetechniken etwa 10% genauer als klinische.**
 - In Abhängigkeit von der spezifischen Analyse übertrafen mechanische Vorhersagemethoden die klinischen substantiell (33-47% der Studien).
 - Obwohl die **klinische Vorhersage häufig genauso genau** war wie die mechanische, war sie nur in wenigen Studien (6-16%) substantiell genauer.
 - Die **Überlegenheit mechanischer Vorhersagetechniken war konsistent**, ungeachtet der Urteilsaufgabe, des Urteilstyps, der Erfahrung der Urteiler oder der Art der zu kombinierenden Daten.
 - Die klinische Vorhersage erwies sich als schlechte, wenn die Prädiktoren auch klinische Interviewdaten enthielten.
 - Ergebnis: Die **mechanische Vorhersage menschlichen Verhaltens ist der klinischen Vorhersagemethode ebenbürtig oder überlegen.** Im Durchschnitt ist sie um etwa 10% besser und das gilt für ein breites Spektrum an Situationen (allgemeine Medizin, Psychia-

trie, Persönlichkeit, Erziehungs- und Trainingssettings) und Diagnostikern (medizinisch ausgebildete Urteiler, Psychologen, unerfahrene Urteiler)

- In diesem Punkt decken sich diese Ergebnisse mit den Aussagen von MEEHL und SAWYER.
- Aber: Die **Überlegenheit der mechanischen Methode ist nicht vollkommen einheitlich:**
- in der Hälfte der analysierten Studien war die klinische Vorhersage ungefähr genauso gut
- in einigen wenigen Fällen war die klinische Methode wesentlich genauer (8 Studien)
 - aber keine systematischen Ausnahmen
 - es konnten nicht viele Studiendesign-Variablen identifiziert werden, die einen Vorteil für die klinischen Vorhersagen vorhersagen
 - nur eine konsistente Eigenschaft: in 7 der 8 Studien hatten die Kliniker mehr Daten zur Verfügung als die mechanische Vorhersage (Aber: im Gesamtttest der 136 Studien machte es für die Überlegenheit der mechanischen Vorhersage keinen Unterschied, ob die Kliniker mehr Daten hatten oder nicht.)
- Wenn die Kliniker Zugang zu einem klinischen Interview hatten, erwies sich die Überlegenheit der mechanischen Vorhersage über die klinische im noch stärkeren Maße.
 - Mögliche Gründe: Empfänglichkeit der Menschen für viele Fehler im klinischen Urteil:
 - Ignorieren der Basisrate
 - Hinweisreizen keine optimalen Gewichte zuweisen
 - Regression zum Mittelwert nicht in Betracht ziehen
 - Kovariation richtig messen
 - Kein Feedback über die Genauigkeit ihrer Urteile

10. Erläutern Sie die Bedeutung des Linsenmodells von Brunswik für die empirische Untersuchung der klinischen Vorhersage! Welche Rolle spielt dabei die Linsenmodell-Gleichung?

- Das Linsenmodell von Brunswik stammt aus der Tradition des probabilistischen Funktionalismus und ist auf den diagnostischen Bereich übertragen worden.
- 1. Mit Hilfe des **Linsenmodells von Brunswik** lässt sich prüfen, in welchem Ausmaß sich die Vorhersagen des Kliniklers durch sein lineares Modell vorhersagen lassen, d.h., inwieweit sein Vorgehen beim Prognostizieren, die Art und Weise, wie er die Daten nutzt und in die Vorhersage einbezieht, durch ein lineares Modell abgebildet bzw. beschrieben werden kann
- 2. Mit Hilfe der **Linsenmodell-Gleichung von Tucker** ist feststellbar, ob die Informationsverarbeitung auf konfigurale, also nichtlineare Art und Weise vor sich geht.

■ Funktionen des Linsenmodells.

- Prüfung, ob die Vorhersagen eines Kliniklers durch ein lineares Modell vorhersagbar sind.
- empirische Analyse des klinischen Vorgehens
- Beschreibung des Modells:
 - objektive, linke Seite: beschreibt die **tatsächlichen Beziehungen** zwischen den diagnostischen Informationen und dem Kriterium

$$r_{y_i y_c} = \text{Korrelation zwischen Kriteriumswert und diagnostischen Infos, d.h., die empirischen Validitäten der diagnostischen Information } x_i \text{ in Bezug auf das Kriterium } y_c$$
 - subjektive, rechte Seite: beschreibt die **subjektive Repräsentation dieser Beziehungen im klinisch vorgehenden Diagnostiker**

$$r_{y_i y_j} = \text{Korrelation zwischen den diagnostischen Infos und der Vorhersage des Kliniklers d.h., in welcher Weise werden die diagnostischen Informationen durch den Klinikler genutzt}$$

◆ entsprechende Regressionsmodelle:

- lineares Regressionsmodell des Kriteriums:

$$y'_c = c b_1 x_1 + c b_2 x_2 + c b_3 x_3 + c b_4 x_4 \quad R_{(y_c \cdot x_1 \dots x_4)}$$

- lineares Regressionsmodell des Kliniklers:

$$y'_j = j b_1 x_1 + j b_2 x_2 + \dots + j b_4 x_4 \quad R_{(y_j \cdot x_1 \dots x_4)}$$

- Die **Leistung**, die Validität des Kliniklers lässt sich dann sehr einfach in Abhängigkeit von den diagnostischen Informationen, der Beziehung dieser Informationen untereinander (also ihrer Interkorrelation) und der Beziehung dieser Informationen zum Kriterium untersuchen.

■ Die Linsenmodell-Gleichung.

- **Zerlegung der Validität des Prognostikers** durch die Linsengleichung von TUCKER.

$$r_{y_c y_j} = G R_c R_j + C \sqrt{1 - R_c^2} \cdot \sqrt{1 - R_j^2}$$

$r_{y_c y_j}$ = Validitätskoeffizient des Urteilers: Korrelation zwischen den Vorhersagen des Kliniklers und des tatsächlichen Kriteriumswerten

G = lineare Komponente der Urteilsgenauigkeit: Korrelation zwischen den vorhergesagten Werten aus dem linearen Modell des Kliniklers und den vorhergesagten Werten aus dem linearen Modell des Kriteriums $r_{y_i y_c}$ (geschätzte Werte)

R_c = lineare Vorhersagbarkeit des Kriteriums: multiple Korrelation zwischen den diagnostischen Informationen und den Kriteriumswerten (tatsächliche Werte, basieren auf Empirie)

R_j = lineare Vorhersagbarkeit des Urteilers: multiple Korrelation zwischen den diagnostischen Informationen und den Vorhersagen des Kliniklers (Korrelation der tatsächlichen Werte)

C = nichtlineare Komponente der Urteilsgenauigkeit: Korrelation zwischen den Residualwerten des Kriteriums und den Residualwerten der Vorhersagen des Urteilers, nachdem die linearen Komponenten beim Klinikler und beim Kriterium abgezogen worden sind

- Untersuchungen zu prognostischen Aktivitäten zeigten:
 - **Prognostizieren** lässt sich im Rahmen dieses Modells ausgezeichnet paramorph¹ beschreiben!
 - Hinweise auf konfigurale Formen der Datenverarbeitung wurden zwar gefunden, die Überlegenheit dieses nichtlinearen Modells, sofern sie überhaupt nachgewiesen werden konnten, war vernachlässigenswert.
 - GOLDBERG (1968): 90-100% der reliablen Urteils-Varianz der klinischen Urteilsbildung kann durch in lineares Modell gewöhnlich aufgeklärt werden.

11. Von welchen Größen hängt die Leistung eines Diagnostikers ab, wenn es nicht möglich ist, das Kriterium durch eine Linearkombination der diagnostischen Information vorherzusagen?

- Die Leistung des Diagnostikers hängt davon ab, inwieweit er in der Lage ist, die potentiellen Fehlerquellen zu umgehen (vgl. Material 1, S.1).
 1. **Mangelnde Kenntnis der Befundlage**, d.h.
 - unzureichende Berücksichtigung einschlägiger empirisch überprüfbarer und bestätigter Wissensgrundlagen,
 - Bezugnahme auf ›subjektives Wissen‹, das entweder für den Einzelfall nicht relevant oder nicht zutreffend ist.
 2. **Unvollständige Analyse und Beschreibung der Probleme.**
 3. **Unzureichende Entfaltung des Hypothesenraums.**
 - Zu frühe Beschränkung auf einige wenige diagnostische Hypothesen.
 - Untersuchung von für die Ausgangsfrage irrelevanten Hypothesen.
 4. **Mangelnde Kenntnis der in Frage kommenden informationserhebenden Verfahren**
 - Ungünstige Auswahl diagnostischer Verfahren

¹ JÄGER & PETERMANN (463): **paramorphe Repräsentationen** = die Beschreibung des diagnostischen Urteilsprozesses anhand einer Variante aus multiplen Regressionsgleichungen (HOFFMANN, 1960) • Betonung, dass der Vorgang der Datenintegration durch den Diagnostiker nicht notwendig so geschieht, wie in der Gleichung formuliert, dass also der Diagnostiker die einzelnen Daten gewichtet und addiert. Die Addition der Daten bedeutet, dass sie unabhängig voneinander in das Urteil eingehen. Besonders diese Eigenschaft der linearen Regressionsgleichungen widerspricht nach Aussage vieler Diagnostiker ihrer Urteilsbildung, die sie als **konfigural** beschreiben, d.h. die Informationen der zu beurteilenden Personen bilden spezifische Konfigurationen, welche die Bewertung und Integration einer diagnostischen Information nur unter Berücksichtigung der weiteren Informationen dieser Konfiguration erlauben.

- Problemunspezifische Anwendung von Testbatterien
 - Vernachlässigung von für die zu prüfenden Hypothesen einschlägigen Verfahren
 - Erhebung von für die Ausgangsfrage irrelevanten diagnostischen Informationen
5. **Unzureichende Beachtung der Durchführungsbedingungen** standardisierter diagnostischer Verfahren.
6. **Mangelnde Weiterverarbeitung erhobener diagnostischer Informationen** durch
- Verwendung veralteter Normwerte
 - Verzicht auf eine zufallskritische Absicherung der Befunde
 - Nicht gerechtfertigte Beschränkung auf rein klinische Formen der Informationsverarbeitung
7. **Inbeziehungsetzung von diagnostischen Hypothesen und für diese Hypothesen nicht einschlägigen diagnostischen Befunden**, damit
- unzutreffende oder relativ zur Wissenssituation unvollständige Diagnosen.
- **Empirische Erfahrung diagnostischer Urteilsbildung** (z.B. WIGGINS 1973).
1. Psychologiestudium; gezieltes systematisches Training; langjährige Berufserfahrung
 - keinen positiven/führenden Einfluss auf die Genauigkeit urteilender Messung und die Validität klinischer Vorhersagen
 2. Statistische Vorhersagemethoden sind den klinischen ebenbürtig oder überlegen.
 3. ›Kliniker‹ lassen sich i.d.R. sehr einfach paramorph modellieren.
 4. Bei neuen zu diagnostizierenden Fällen bewähren sich diese einfachen Modell i.d.R. wenigstens so gut wie die modellierten Diagnostiken selbst.
 5. Entscheidende Verbesserungen der Validität klinischer Vorhersagen sind i.d.R. dann erreichbar, wenn dem Diagnostiker Wissensgrundlagen zur Verfügung gestellt werden, die ihm eine Annäherung an die statistische (aktuarische) Vorgehensweise bei der Informationsverarbeitung erlaubt.
- **Daumenregeln** für den klinischen Praktiker.
1. Wenn **kriteriale Informationen existieren**, sind diese zu erheben und statistische Modelle der Informationsverarbeitung zu konstruieren (der Diagnostiker gewinnt Zeit, um seine Fähigkeiten bei der Datenerhebung sinnvoller zu nutzen).
 2. Wenn **kriteriale Informationen nicht existieren** und es **mehrere erfahrene Urteiler** gibt,
 - sind die zusammengesetzten Urteile den Vorhersagen zugrunde zu legen (setzt Vertretbarkeit der Kosten voraus).
 - ist zunächst die Kombination dieser Urteile zu bilden, dann ist der Diagnostiker zu suchen, deren Urteil am höchsten mit der Kombination der Urteile aller Diagnostiker korreliert. Diese Person ist mit Hilfe der linearen Regressionstechniken zu modellieren und das so entwickelte Modell der Vorhersage zugrunde zu legen (vorausgesetzt, die Kosten – s.o. – sind nicht vertretbar).
 3. Existieren **keine kriterialen Informationen** und es gibt **einen erfahrenen Diagnostiker**, dann ist das paramorphe Modell dieser Person den Vorhersagen zugrunde zu legen.

12. Was haben Nystedt und Magnusson (1973) untersucht, und welche Schlussfolgerungen lassen sich aus ihrer Untersuchung ziehen?

- ◆ **NYSTEDT & MAGNUSSON (1973)**: Untersuchung darüber, **inwieweit kriteriale Informationen (Interkorrelation und empirische Validität der Tests) und/oder die Urteilsgenauigkeit die Vorhersage der Diagnostiker verbessern** [Cue relevance and feedback in a clinical prediction task].
 - Vorgehen: **Datenanalyse erfolgte auf der Basis der Linsenmodell-Gleichung.**
 - Ergebnis: Die **kriterialen Informationen** (Interkorrelation und empirische Validität der Tests) **verbesserten die Vorhersage der Kliniker entscheidend**. Die Informationen über die Urteilsgenauigkeit hatten keine zusätzliche Verbesserung zur Folge.
 - Schlussfolgerung: **Diagnostiker haben sich mit Hilfe der kriterialen Information intuitiv einer linearen Regressionsgleichung angepasst**. Es gibt keine Anhaltspunkte, dass nichtlineares Vorgehen besser ist.

13. Wie ließe sich die Behauptung prüfen, ein Kliniker verarbeite seine diagnostischen Informationen nicht linear, sondern konfigural? Welche grundsätzlichen methodischen Probleme stehen aber dem Nachweis der Konfiguralität klinischer Urteilsbildung entgegen?

■ Überprüfung linear / konfigural:

- ◆ Mit Hilfe der **Linsenmodellgleichung von TUCKER** ist feststellbar, ob die Informationsverarbeitung nichtlinear, d.h. konfigural erfolgt.
 - Mit der Linsenmodellgleichung wird die **Validität des Prognostikers zerlegt**. Sie enthält eine **nichtlineare Komponente der Urteilsgenauigkeit, C**, diese entspricht der Korrelation zwischen den Residualwerten des Kriteriums und den Residualwerten der Vorhersagen des Urteilers, nachdem die linearen Komponenten beim Kliniker und beim Kriterium abgezogen worden sind.
 - Auf der Grundlage des Linsenmodell und der Linsengleichung sind inzwischen zahlreiche Untersuchungen prognostischer Aktivitäten in der psychologischen Diagnostik unternommen worden. Es hat sich gezeigt, dass sich Prognostizieren ausgezeichnet im Rahmen dieses Modells paramorph beschreiben lässt.
- ◆ **Lautes Denken.**
 - Diagnostiker auffordern, die Art und Weise seiner konfiguralen Urteilsbildung zu spezifizieren.
 - Auf dieser Grundlage müsste man die einzelnen Prädiktoren zu konfiguralen Variablen transformieren, indem man sie entsprechend den Anweisungen des Diagnostikers multiplikativ miteinander verknüpft und/oder quadriert, potenziert in die Vorhersage einbezieht.
 - [JÄGER & PETERMANN, S. 464]. Ob der Prozess der Informationsintegration mit algebraischen Modellen überhaupt beschrieben werden kann, wird v.a. von Simon et al. (1976) bestritten. Sie versuchen, die Verarbeitung der Informationen bei Problemlöseprozessen dadurch aufzuzeichnen, dass sie die Probanden instruieren, ihre Gedanken laut mitzuteilen. Aus den Protokollen des lauten Denkens werden die Strategien analysiert und so transformiert, dass sie von einem künstlichen Problemlöser nachvollzogen werden könne.
 Dass die Methode des lauten Denkens Einblick in die Denkprozess erlaubt ist von einigen bestritten worden, mit dem Argument, dass solche Prozesse – besonders routinierte – der bewussten Erfahrung nicht zugänglich seien. Aber auch dann, wenn sie vom Diagnostiker mitgeteilt werden können, ist ihre Analyse von seiner Fähigkeit abhängig, die einzelnen Regeln der diagnostischen Strategie sprachlich umzusetzen. Vermutlich sind bestimmte Regeln, etwa jene, die das Verstehen des Problems eines Klienten bewirken, der bewussten Erfahrung nicht direkt zugänglich und können erst später rekonstruiert werden. Darüberhinaus kann nicht ausgeschlossen werden, dass die Instruktion, alle Gedanken laut mitzuteilen, eine Art Selbstverpflichtung auslöst, so zu verfahren, wie es von einem guten Diagnostiker erwartet wird, etwa viele Hypothesen zu bilden und sie danach zu prüfen, ob die diagnostischen Daten diese Hypothesen stützen oder schwächen, zu entscheiden, ob weitere Daten notwendig sind und die Daten mit einer nachvollziehbaren Regel zu integrieren. Ob die typische diagnostische Urteilsbildung dadurch repräsentiert wird, ist eine offene Frage.

■ Grundsätzliche methodische Probleme beim Nachweise der Konfiguralität (nach GANZACH).

- Mangel an statistischer Power (bei der Modellierung von Nichtlinearität)
- Welches mathematische Modell ist angemessen?
- Schwierigkeit bei der mathematischen Modellierung von Nichtlinearität
- Je komplexer konfigurale Gleichung, desto größer die Anzahl der erforderlichen Probanden.
- MMPI:
 - cues der MMPI-Profile korrelieren hoch miteinander • dies verringert den Unterschied zwischen der Güte der Modellanpassung unterschiedlicher Modelle
 - auch unrichtige Spezifizierung der den Urteilen zugrunde liegenden nichtlinearen Prozesse kann eine Ursache für Probleme sein • z.B. ist es unwahrscheinlich, dass die Beziehungen zwischen allen Paaren der MMPI-Skalen nichtlinear sind (wie im Modell von WIGGINS & HOFFMAN angenommen wird).