Fragebogen 2

Einführung in die Psychologische Diagnostik, Westmeyer
Ausarbeitung: Johannes Geffers [auf Grundlage vieler Vorarbeiten...]

1. Welches sind die Charakteristika der probabilistischen Testtheorie im Unterschied zur klassischen?

- Der zentrale Unterschied zwischen KTT und PTT liegt in der **Unterscheidung latenter und manifester Variablen** (ROST & SPADA 1982) die in der KTT nicht explizit gemacht wird. Die PTT setzt vor der Messung an und begründet diese erst.
- Modelle der probabilistischen Testtheorie haben **empirischen Gehalt**. Sie sind überprüfbar.
- **Homogenität der Items.** Es werden nur solche Items zugelassen, deren Itemcharakteristik-Kurven gleichartig verlaufen und sich nicht schneiden.
- **Stichprobenunabhängigkeit von Skala und Items.** Wenn ein Itemsatz homogen ist, dann ist ein Test stichprobenunabhängig in einem besonderen Sinne. Genannt wird diese Eigenschaft auch ›spezifische Objektivität‹ – d.h.: Innerhalb der Population, für die Modellkonformität festgestellt ist, fallen für einen Probanden die Item- und Personenparameter immer gleich aus, gleichgültig welche Items er bearbeitet.
- **Lokale stochastische Unabhängigkeit der Items.** Die Eintretenswahrscheinlichkeit mehrer Ereignisse hängt nur ab von der Eintretenswahrscheinlichkeit jedes einzelnen Ereignisses und nicht von ihrer Kombination.
- **Separierbarkeit von Item- und Personenparameter.** In der klassischen Testtheorie hat ein *Item ›keine Schwierigkeit an sich‹*, sondern nur in Bezug auf eine Population. Im RASCH-Modell sind *Item- und Personenparameter unabhängig voneinander konzipiert*: »Man berechnet also die Parameter für verschiedene Stichproben unter der Vorannahme, daß innerhalb desselben Itemvektors der Itemparameter gleich und innerhalb derselben Probanden-Gruppe der Personenparameter gleich ist.« (FISSENI 1997, 179)
- Modelle enthalten Auftretenswahrscheinlichkeit von manifestem Verhalten.
- **Testscore als erschöpfende Statistik.** Die Anzahl der gelösten Items gibt die durch den Test erhobene Information ›vollständig‹ wieder – es ist unerheblich welche Items und in welcher Reihenfolge (Lösungsvektor) gelöst wurden.

	Klassische Testtheorie (KTT)	***Item-Response-Theorie (IRT)***
Messwert	**Ausdruck der Fähigkeit** (manifeste Variabel). Die klassische Testtheorie beschäftigt sich mit der Frage, wie aus einer Anzahl von Verhaltensbeobachtungen von Versuchspersonen in bestimmten Situationen auf die wahre Ausprägung eines Persönlichkeitsmerkmals geschlossen werden kann.	**Indikator einer latenten Variable.** Der Item-Response-Theorie liegt die Vorstellung zugrunde, dass eine latente, nicht beobachtbare Eigenschaft (›latent trait‹) der Personen das Testverhalten steuert, so dass *vom Testverhalten auf den Ausprägungsgrad einer latenten Eigenschaft* (Verhaltensdisposition) geschlossen werden kann.
Zusammenhang Messwert-Merkmal *und* ***Verhältnis von Beobachtungs-ebene & Konstruktebene***	**deterministisch.** Der Testwert einer Vp wird als Realisation einer Zufallsvariablen mit dem »wahren Testwert« als Erwartungswert betrachtet. Der Testwert nur eine mehr order weniger genaue Schätzung des wahren Wertes dar. **Gleichsetzung.** Unter Berücksichtigung des Messfehlers wird die Beobachtungsebene mit der Konstruktionsebene gleichgesetzt. Die beobachtete Variable (das Testverhalten) und die gemessene Variable (die Eigenschaft oder Fähigkeit der Vp) werden gleichgesetzt; (Summenscore = Messwert).	**probabilistisch.** Auch bei genauer Kenntnis des Ausprägungsgrades der latenten Eigenschaft kann das Testverhalten nicht mit absoluter Sicherheit vorhergesagt werden, sondern nur mit einer gewissen Wahrscheinlichkeit • *probabilitische, stochastische* Modelle. **Trennung.** Trennung zwischen Beobachtungsebene und theoretisch angenommener Konstruktebene. Die beobachtete Variable wird lediglich als Indikator einer latenten Variablen, die mit dem Test erfasst werden soll, betrachtet • daher »*latent-trait-theory*«.

	Klassische Testtheorie (KTT)	***Item-Response-Theorie (IRT)***
Aufgabenschwierigkeit	**Durchschnittliche Lösungswahrscheinlichkeit.** Ein Item ist leicht oder schwer in Abhängigkeit davon, ob viele oder wenige (nicht-) fähige Personen in der Stichprobe sind. $P=\frac{N_R}{N}$ P = Schwierigkeitsindex N_R = Anzahl ›Richtiglöser‹ N = Gesamtzahl d. ProbandInnen	**Funktion von Aufgabenschwierigkeit und Personenfähigkeit.** Die Wahrscheinlichkeit P einer (richtigen) Reaktion a der v-ten Vp auf das i-te Item ist eine Funktion der latenten Variablen der Fähigkeit des Pbn und der Itemschwierigkeit. $P(A_{vi})=f(\xi_v,\sigma_i)$ ξ_v = Fähigkeit des Schülers v σ_i = Schwierigkeit des Items i
Art der Theorie	**Fehlertheorie.**	**Messtheorie.**
Stichprobenabhängigkeit	**Testscore abhängig von der Stichprobe.** Der Testscore wird als Schätzung für den Ausprägungsgrad der Verhaltensdisposition genommen. Dem sich daraus ergebenden Problem der Stichprobenabhängigkeit der Meßwerte soll durch sehr strenge Anforderungen an die Repräsentativität der Stichprobe nachgekommen werden. Die einzelnen Meßwerte bleiben zwar stichprobenabhängig, aber da jede Stichprobe gleiche Verteilungen der Itemschwierigkeiten aufweist, bleiben die anhand verschiedener Itemstichproben gewonnenen Tesergebnisse invariant.	**Testscore unabhängig von der Stichprobe.** Im Modell von Rasch werden *spezifisch objektive* Vergleiche ermöglicht; d.h. der Unterschied der Fähigkeit ξ_v und ξ_w je zweier Personen kann unabhängig davon bestimmt werden, welche Aufgaben eines modellkonformen Itempools dafür herangezogen werden, bzw. umgekehrt, der Vergleich je zweier Aufgaben bezüglich ihrer Schwierigkeiten σ_i und σ_j ist unabhängig davon, welche Personenstichprobe dafür verwendet wird.
Grundgleichungen	Die Lösung einer Aufgabe entspricht dem ›wahren Wert‹ unter Berücksichtigung des Messfehlers: $x_{vi}=\tau_{vi}+\epsilon_{vi}$ x_{vi} = Messung bzw. beob. Wert τ_{vi} = ›wahrer Wert‹ ϵ_{vi} = Fehlerwert	Die Wahrscheinlichkeit der zu beobachtenden Reaktion soll auf latente Variablen zurückgeführt werden: $P(A_{vi})=f(\xi_v,\sigma_i)$ ξ_v = Fähigkeit des Schülers v σ_i = Schwierigkeit des Items i Parameter, die nur aufgrund von Stichprobendaten geschätzt werden können Die Wahrscheinlichkeit P einer (richtigen) Reaktion a der v-ten Vp auf das i-te Item ist eine Funktion der latenten Variablen der Fähigkeit des Pbn und der Itemschwierigkeit.
Standardmessfehler *und* ***Reliabilität***	**Der Standardmessfehler gilt für alle Messwerte einer bestimmten Population.** Problem: Der Standardmessfehler wird als Konstante behandelt, obwohl er tatsächlich eine Kurve ist (niedriger in der Nähe des Mittelwerts, zu den Extremen hin größer). **Eine durchschnittliche Reliabilität.**	**Der Standardmessfehler unterscheidet sich in Bezug auf die einzelnen Messwerte, ist aber in Bezug auf verschiedene Populationen gleich.** **Unterschiedliche Reliabilitäten für unterschiedliche Fähigkeitsniveaus.** • Reliabilität als Funktion der Fähigkeit schätzen

2. Zwischen welchen Größen wird in einer Itemcharakteristik ein funktionaler Zusammenhang ausgedrückt?

- Eine **Itemcharakteristik** gibt den Zusammenhang zwischen **Itemparameter** ([Aufgaben-]Schwierigkeit) und **Personenparameter** (Fähigkeit) auf der einen Seite zur **Lösungswahrscheinlichkeit** andererseits wieder. Diese ist eine Funktion von Personen- und Itemparameter.

$$P(A_{vi}=1)=f_i(\xi_v) \qquad (3)$$

- Die Funktion f_i wird Itemcharakteristik genannt. Sie hat den Index i, da im allgemeinsten Fall jedes Item eine andere Funktion (Itemcharakteristik) haben kann.

3. Was ist unter lokaler stochastischer Unabhängigkeit zu verstehen? Worauf ist die trotz lokaler Unabhängigkeit zu beobachtende mehr oder weniger hohe Korrelation zwischen den Lösungswahrscheinlichkeiten einzelner Items zurückzuführen? Was ist also mit „lokal" gemeint?

1. Das **Lösungsverhalten bei einem Item ist unabhängig von dem bei anderen Items**, es hängt nur von der Personenfähigkeit ab.
2. Wenn das **Lösungsverhalten bei Items untereinander korreliert**, dann **aufgrund der unterschiedlichen Fähigkeitswerte** der untersuchten Personen.
3. **Lokal** bezieht sich auf **einen bestimmten Ort (locus) auf dem Fähigkeitskontinuum**. Haben alle Personen den selben Fähigkeitswert, verschwindet die Korrelation.

- Wichtige Annahme, weil die relative Häufigkeit der Beobachtung eines Ereignisses nur dann einen Schätzwert für die Wahrscheinlichkeit des Ereignisses darstellt, wenn die Beobachtungen unabhängig voneinander gemacht werden, sich also nicht gegenseitig beeinflussen.

4. Welche Annahmen liegen dem probabilistischen Testmodell von Rasch zugrunde? Welche Konsequenzen hat es?

- **Annahmen.**

  0. aus Frage 1:
     i. Den *beobachteten* (**manifesten**) Reaktionen liegen *nicht beobachtbare* (**latente**) Eigenschaften zugrunde (• probabilistischer Zusammenhang).
     ii. **Personen-** & **Itemparameter**. **Lösungswahrscheinlichkeit** ist eine logistische Funktion aus Fähigkeit *und* Aufgabenschwierigkeit.
     iii. Grundgleichung des Modells:

$$P(A_{vi})=\frac{e^{(\xi_v-\sigma_i)}}{[1+e^{(\xi_v-\sigma_i)}]} \qquad (8)$$

     ξ_v = Fähigkeitsparameter der VP v in bezug auf Aufgaben des vorgelegten Typs
     σ_i = Schwierigkeitsparameter der Aufgabe i in bezug auf Personen der untersuchten Population
     e = Eulersche Zahl, gleich 2.72

  1. Der **Test ist homogen** (Eindimensionalität, alle Items erfassen dasselbe Merkmal).
     - Homogenität von Items bedeutet, dass alle Items dasselbe Merkmal erfassen oder dass das Lösungsverhalten einer Person hinsichtlich aller Items durch denselben Fähigkeitsparameter charakterisiert werden kann. Als homogen bezüglich eines Tests bezeichnet man dabei jene Personen, bei denen der Test dasselbe Merkmal anspricht.
  2. Die **Itemcharakteristiken** sind **monoton steigend**.
  3. Es wird **lokale stochastische Unabhängigkeit** vorausgesetzt.
     - Das Lösungsverhalten bei einem Item ist unabhängig von dem bei anderen Items, es hängt nur von der Personenfähigkeit ab. Wenn das Lösungsverhalten bei Items untereinander korreliert, dann aufgrund der unterschiedlichen Fähigkeitswerte der untersuchten Personen.
  4. **spezifische Objektivität** der Vergleiche: Das Ergebnis des Vergleiches zweier Personen oder der Items hängt nicht vom benutztem Messinstrument ab, unabhängig von der Item-

auswahl.

- »Der Vergleich zweier Personen bzw. einer Person mit einem Kriterium hinsichtlich eines Merkmals ist dann spezifisch objektiv im Sinne von Rasch (1966), wenn das Ergebnis des Vergleichs nicht von dem für den Vergleich benutzten Meßinstrument abhängt. Auf den hier interessierenden Fall bezogen: Der Vergleich ist dann spezifisch objektiv, wenn er unabhängig von der in den Test aus einer Klasse homogener Items aufgenommenen Itemauswahl ist.« (ROST & SPADA 1982, 75)
- Ein Vergleich einer Person mit einer anderen Person oder einem Kriterium kann nur dann als spezifisch objektiv angesehen werden, wenn auch die Testresultate der anderen mituntersuchten Probanden das Ergebnis dieses Vergleichs nicht beeinflussen.
- ➢ Der <u>Vergleich zweier Personen erfordert keine Bezugnahme auf die gleiche Itemstichprobe</u> (z.B. Pbn mit hoher Fähigkeit / Pbn mit niedriger Fähigkeit • Auswahl einer je geeigneten Itemstichprobe). • adaptives Testen

5. **wechselseitige Stichprobenunabhängigkeit der Parameterschätzungen**.
   - »Wenn das Modell einem Datensatz adäquat ist, d.h. wenn Gleichung (8) die Ergbnisse hinreichend gut beschreibt, was mit Hilfe des Modellgeltungstests prüfbar ist, so ist das Ergebnis der Schätzung der Itemparameter unabhängig von der Verteilung der Personenparameter in der Analysestichprobe. Umgekehrt ist das Ergebnis Personenparameterschätzung unabhängig von der gewählten Itemstichprobe.« (ROST & SPADA 1982, 76) – Stichprobenabh. da kein Bezug auf Mittelwert und SD
6. Der **Testscore** (=Anzahl der gelösten Aufgaben) **stellt eine erschöpfende Statistik für die Fähigkeiten einer Person** dar.
   - Der <u>Testscore schöpft alle im Test über das Merkmal enthaltene Informationen aus</u>, d.h. <u>alle Lösungsvektoren sind gleichwertig</u>, es ist egal welche Aufgaben gelöst werden; diese Annahme wird in der PTT implizit vorausgesetzt. RASCH fordert die Gültigkeit dieser Annahme explizit, nach ROST und SPADA ist das Rasch-Modell das einzige, welches mit dem Vorgehen verträglich ist (PPs, Ips).

◆ <u>**Konsequenzen**</u>.

1. **empirisch prüfbar**.
   - In der **KTT** wird die Grundgleichung $x = T + e$ unterstellt; sie ist <u>nicht empirisch prüfbar</u> und kann daher als Modell in der Praxis nicht scheitern.
   - In der **PTT** (probabilistischen Testtheorie) haben die <u>Testmodelle einen empirischen Gehalt</u>, sie sind damit prüfbar und können ›an der Realität scheitern‹. Geprüft wird, ob die Daten mit dem Modell kompatibel sind. Items, die nicht modellkonform sind, werden ausgeschieden.

5. Warum ist im einfachen Rasch-Modell die Lösungswahrscheinlichkeit eines Items, das so schwierig wie die Person fähig ist, 0.50? Gibt es probabilistische Testmodelle, bei denen in diesem Fall die Lösungswahrscheinlichkeit größer als 0.50 ist? Welche?

$$P(A_{vi}) = \frac{e^{(\xi_v - \sigma_i)}}{[1 + e^{(\xi_v - \sigma_i)}]}$$

$P(A=1)$ = Lösungswahrscheinlichkeit
ξ_v = Fähigkeitsparameter (PP) der VP v in bezug auf Aufgaben des vorgelegten Typs
σ_i = Schwierigkeitsparameter (IP) der Aufgabe i in bezug auf Personen der untersuchten Population
e = Eulersche Zahl, gleich 2.72

i. Das ist eine logische Konsequenz der Modellgleichung des RASCH-Modells. Der Term $e^{\xi - \sigma}$ ergibt für den Fall, daß IP und PP gleich sind 1, womit die Gleichung den Wert ½ ergibt.

ii. Das **3-Parameter-Modell** von BIRNBAUM enthält einen sogenannten <u>Rateparameter</u>. Er soll berücksichtigen, dass z.B. durch die Vorgabe von multiple-choice-Aufgaben Lösungen auch durch Raten zustande kommen können. So ist die Lösungswahrscheinlichkeit selbst für $\xi \rightarrow -\infty$ größer Null.

6. Was wird innerhalb des Rasch-Modells unter spezifischer Objektivität, was unter Stichprobenunabhängigkeit verstanden?

- Der Vergleich zweier Personen P und Q ist dann **spezifisch objektiv**, wenn das Ergebnis dieses Vergleichs unabhängig ist von der dem Vergleich zugrundegelegten Itemmenge.
- Die Schätzung des Personenparameters ist unabhängig davon, welche **Itemstichprobe** dem Vergleich zugrundegelegt wird und vice versa.
- Wenn ein Itemsatz homogen ist, dann ist ein Test stichprobenunabhängig in einem besonderen Sinne. Genannt wird diese Eigenschaft auch ›spezifische Objektivität‹ – d.h.: Innerhalb der Population, für die Modellkonformität festgestellt ist, fallen für einen Probanden die Item- und Personenparameter immer gleich aus, gleichgültig welche Items er bearbeitet.
- ausführlich: s.o.

7. Welches sind die zentralen Charakteristika kriteriumsorientierter Tests? Worin unterscheiden sie sich insbesondere von klassisch konstruierten Tests?

- **Kriteriumsorientierte Tests.**
  - Kriteriumsorientierte Tests sind auf eine **Idealnorm** (Lehrzielnorm) **bezogen**. Die Probanden werden unabhängig von der Einordnung anderer Probanden in Bezug auf ein Fähigkeitskontinuum oder durch Zuordnung zu einer Klasse eingestuft.
  - Die wesentliche Unterscheidung zu klassisch konstruierten Tests besteht in der **Kontentvalidität** der kriteriumsorientierten Tests. *Ein Test ist kontentvalide, wenn der wahre Wert der VPn sich mit Hilfe einer repräsentativen Auswahl von Aufgaben aus einer Grundmenge von Aufgaben schätzen lässt*. Kriteriumsorientierte Tests enthalten Aufgaben aus einer zuvor wohldefinierten Grundmenge von Aufgaben, sei es, dass sämtliche Aufgaben der Grundmenge im Test enthalten oder dass repräsentative Stichprobe von Aufgaben gebietet werden.
  - **Problem ›absoluter‹ Messung** – sie sagt (logischerweise) nichts über die Relativität des Ergebnisses aus. Beispiel: Wenn ich durch absolute Messung exakt die Körpergröße eines Menschen kenne, weiß ich noch nicht, ob er relativ groß oder klein ist.
  - Kriteriumsorientierte Tests treten also nicht an die Stelle der sozial relativierenden Tests, sondern ergänzen sie.

- **Klassisch konstruierte Tests.**
  - Klassisch konstruierte Tests sind hingegen auf eine **Realnorm bezogen**. Die Beurteilung druch eine komparative Messung erfolgt relativ zur Stichprobe/Population *oder* im Vergleich zwischen früherer und aktueller Leistung (individuelle Realnorm).
  - **Schwierigkeiten** für die KTT: Wenn alle Personen beispielsweise in Folge eines Treatments den gleichen wahren Wert erreichen, bricht der Test zusammen, da die Varianz der wahren Werte gleich null ist

    $var(\tau)=0$. Damit wird aber auch die Reliabilität gleich null:

    $r_{xx}=\frac{[var(\tau)]}{[var(x)]}$. Da die Validität in folgender Beziehung

    $r_{xy}\leqslant\sqrt{r_{xx}}$ zur Reliablität steht, wird auch die Validität gleich null.

    $Varianz=0 \quad \bullet \quad Reliabilität=0 \quad \bullet \quad Validität=0$

    - Auch die Trennschärfen der Aufgaben streben nach Null. D.h. für pädagogisch angestrebte Bedingungen versagen klassisch konstruierte Tests. Grund: Zweck klassisch konstruierter Tests ist die komparative Messung.

- **Forderungen & Folgen der Kontentvalidität.**
  - Elemente müssen zum **definitorischen Bereich** gehören.
  - Items müssen repräsentative Stichprobe aus dem gesamten Bereich oder dem gesamten Itemuniversum sein.

- Werte, die mit einem Test dieser Art gewonnen sind, lassen sich direkt interpretieren mit Bezug auf die Grundmenge von Aufgaben (das ›item universe‹ bzw. die ›behavior domain‹).
- Der **wahre Wert** des Probanden ist bei einem kontentvaliden Test derjenige Wert, den der Proband erhalten würde, wenn er die ganze Grundgesamtheit von Aufgaben beantworten würde. Er kann mit Hilfe einer repräsentativen Stichprobe geschätzt werden.
- Vorgang bei Validitätsüberprüfung: Interraterreliabilität.

8. Warum ist es schwierig im Bereich kriteriumsorientierter Test mit der klassischen Testtheorie zu arbeiten?

- Mit der KTT lässt sich nur sehr schwer im Bereich kriteriumsorientierter Tests arbeiten und zwar wegen folgender Gesichtspunkte:
  - **Varianz der beobachteten Werte = 0**. Es ergäbe sich folgendes:

    $var(x)=0$ •

    - Reliabilität wird unbestimbar, da

      $r_{xx}=\frac{[var(\tau)]}{[var(x)]}$ ist und die Division durch null verboten ist. •

    - Validität wird unbestimmbar, da

      $r_{xy} \leqslant \sqrt{r_{xx}}$ gilt.
  - **Varianz der wahren Werte = 0**. In Folge eines Treatments könnten die wahren Werte alle den gleichen Wert – beispielsweise das Maximum – annehmen. In dem Maße, in dem eine Lehrprozedur erfolgreich ist (der Erwartungswert x richtiger von N Aufgaben für jeden Schüler genau bei N liegt, also für jeden 100% richtige Lösungen zu erwarten sind), streben die Reliabilitäten und die Validitäten eines entsprechenden Leistungstests gegen Null.
    - Varianz der wahren Werte = 0.

      $var(\tau)=0$ •

    - Reliabilität wird null, da

      $r_{xx}=\frac{[var(\tau)]}{[var(x)]}$ gilt. •

    - Damit wird auch die Validität null, da

      $r_{xy} \leqslant \sqrt{r_{xx}}$ gilt.
  - Ähnliches gilt für die **Trennschärfen** der Aufgaben: Diese streben nach 0, wenn die Schwierigkeitsindizes nach 1 gehen.
  - **Für diese pädagogisch wünschenswerten und angestrebten Bedingungen versagen klassisch konstruierte Tests.**

9. Sie haben keine Lust, immer bei einer Intelligenzfeststellung den ganzen HAWIK-III durchzuführen, und wollen deshalb eine Kurzform, die nur aus zwei Untertests besteht, erstellen. Bei einer Sichtung der Untertests entscheiden Sie sich für AW, GF und BE. BE soll auf jeden Fall aufgenommen werden, es ist aber noch nicht entschieden, ob AW oder GF als zweiter Untertest herangezogen wird.

Sie inspizieren deshalb die Interkorrelationstabelle. Beim HAWIK-III wurden (aufgerundet) folgende Korrelationen zwischen AW (Allgemeines Wissen), GF (Gemeinsamkeitenfinden), BE (Bildergänzen), GT (Gesamttest) bei Sechsjährigen gefunden:

cor (BE, AW) = 0.46 cor (BE, GF) = 0.34 cor (BE, GT) = 0.71

cor (AW, GT)= 0.68 cor (GF, GT) = 0.61

Es sollen nun die Standardwerte im Gesamttest geschätzt werden unter Bezugnahme auf den Untertest BE und einen der beiden Untertests AW und GF. Welchen der beiden Untertests AW oder GF würden Sie als zweiten Prädiktor für geeigneter halten? Begründen Sie ihre Entscheidung! (Hinweis: Berechnen Sie zur Klärung der Frage zunächst die entsprechenden Beta-Gewichte, und geben Sie die zugeordneten linearen Regressionsgleichungen an. Berechnen Sie dann die multiplen Korrelationen, und geben Sie auf der Basis eines Vergleichs der beiden sich dabei ergebenden Werte den geeigneteren Untertest an!)

- Für eine möglichst optimale Zusammenstellung von einzelnen Testverfahren ist es günstig, wenn **einzelne Untertests mit dem Kriterium möglichst hoch und untereinander möglichst niedrig korrelieren**. Das heißt es sollen möglichst **verschiedene Seiten des Kriteriums erfasst** werden. Die Methode ist die **multiple Regression**.
- Sollen **mehrere Prädiktorvariablen** gleichzeitig mit einer Kriteriumsvariablen in Beziehung gesetzt werden, wird eine **multiple Korrelation berechnet**. Mit der multiplen Regressionsrechnung wird eine Gleichung zur Vorhersage einer Kriteriumsvariablen bestimmt bei gleichzeitiger Berücksichtigung mehrerer Prädiktorvariablen.
- Das führt zu der Frage, wie die Anzahl der zu erhebenden Variablen minimiert werden kann, ohne auf relevante Informationen zu verzichten.
- Anhand der **multiplen Korrelationsstatistik** ist es möglich, Beziehungen zwischen 2 oder mehreren Prädiktorvariablen und einer einzelnen Kriteriumsvariablen zu analysieren. Das Ergebnis dieser Analyse besteht in einer **multiplen Regressionsgleichung**, die der Vorhersage von Kriteriumswerten dient, und im **multiplen Korrelationskoeffizienten R**.
- Für den Fall das y, x_1 und x_2 in z-Werten ausgedrückt vorliegen gilt:

  $y´ = b_1 x_1 + b_2 x_2$

  b_1 & b_2 • **Beta-Gewichte** [b_1 gibt an, um welchen Wert y´ wächst, wenn x_2 = konstant]

  x_1 & x_2 • **Prädiktoren**
- Die **Aufgabe der multiplen Korrelations- bzw. Regressionsrechnung** besteht darin, die Beta-Gewichte zu berechnen. Die Beta-Gewichte werden so bestimmt, dass die Regressionsgleichung die Kriteriumsvariable möglichst genau vorhersagt. Die Regressionsgleichung wird nach dem Kriterium der kleinsten Quadrate festgelegt, d.h. die Summe der quadrierten Abweichungen der tatsächlichen y-Werte von den geschätzten y-Werten ergibt ein Minimum.
- Der **multiple Korrelationskoeffizient R** erfasst den Zusammenhang zwischen k Prädiktorvariablen und einer Kriteriumsvariablen. R hat definitionsgemäß einen Wertebereich zwischen 0 bis 1.
  - R entspricht der bivariaten Korrelation zwischen der vorhergesagten und der tatsächlichen Kriteriumsvariablen.

    $$R_{y \cdot x_1 x_2} = r_{yy´} \qquad (8)$$
- **Berechnung der multiplen Korrelation**:

  $$R^2_{y \cdot x_1 x_2} = b_1 r_{yx_1} + b_2 r_{yx_2} \qquad (6)$$
- **Berechnung der Beta-Gewichte**:

  $$b_1 = \frac{[r_{yx_1} - r_{yx_2} \cdot r_{x_1 x_2}]}{[1 - r^2_{x_1 x_2}]} \qquad (4)$$

  $$b_2 = \frac{[r_{yx_2} - r_{yx_1} \cdot r_{x_1 x_2}]}{[1 - r^2_{x_1 x_2}]} \qquad (5)$$
- **Daraus folgt**: Einsetzen von (4) und (5) in (6) •

$$R^2_{y \cdot x_1 x_2} = \frac{[r^2_{yx_1} + r^2_{yx_2} - 2r_{yx_1} r_{yx_2} r_{x_1 x_2}]}{[1 - r^2_{[x_1 x_2]}]} \qquad (7)$$

- **Interpretation der Beta-Gewichte.**
  - Ein **positives Beta-Gewicht** besagt, dass eine Zunahme der entsprechenden Prädiktorvariablen zu einer Vergrößerung des vorhergesagten Kriteriumswertes beiträgt, und
  - Ein **negatives Beta-Gewicht**, dass eine Zunahme der entsprechenden Prädiktorvariablen zu einer Verkleinerung des Wertes der Kriteriumsvariablen führt.
  - Dem **Beta-Gewicht** ist zu entnehmen, welchen Beitrag ein einzelner Prädiktor im Kontext aller übrigen Prädiktoren zur Klärung der tatsächlichen Kriteriumsvarianz leistet.

- **Für die multiple Korrelation R zwischen y und x_1 und x_2 gilt**:

1. die multiple Korrelation ist um so größer, *je größer die Korrelation zwischen y und den einzelnen Prädiktoren* ist.
2. die multiple Korrelation ist um so größer, je *kleiner die Korrelation zwischen den Prädiktoren untereinander* ist.
3. die multiple Korrelation ist immer *größer oder wenigstens gleich der Korrelation zwischen y und den einzelnen Prädiktoren*.
   - Daraus folgt, dass durch die Berücksichtigung weiterer Prädiktoren bei einer Kriteriumsvorhersage keine Verschlechterung der Vorhersagegenauigkeit eintreten kann, da die multiple Korrelation wenigstens so hoch ist wie die höchste einfache Korrelation zwischen dem Kriterium und einem der Prädiktoren

10. Zwischen einem Prädiktor x_1 und einem Kriterium y besteht eine Korrelation von 0.4. Eine weitere Variable x_2 weist eine Nullkorrelation mit dem Kriterium auf, korreliert allerdings zu 0.5 mit dem Prädiktor x_1 . Eine dritte Variable x_3 hat ebenfalls eine Nullkorrelation mit dem Kriterium, aber eine negative Korrelation von –0.3 mit dem Prädiktor x_1 . Sie sollen nun außer dem Prädiktor x_1 eine weitere Variable, also x_2 oder x_3 , für eine Vorhersage des Kriteriums y heranziehen. Um was für eine Art von Variablen handelt es sich bei x_2 und x_3 ? Welcher würden Sie den Vorzug geben und warum? Geben Sie die entsprechenden linearen Regressionsgleichungen an, und berechnen Sie die zugehörigen Regressionskoeffizienten (Beta-Gewichte)!

$$cor(y, x_1) = 0.4 \qquad cor(y, x_2) = 0.00 \qquad cor(y, x_3) = 0.00$$
$$cor(x_1, x_2) = 0.5 \qquad cor(x_1, x_3) = -0.3$$

- Die Aufgabe besteht darin, nach Vorliegen der Endergebnisse anhand der Beta-Gewichte und Validitäten zu prüfen, ob Suppressionseffekte wirksam waren. Hierbei ist v.a. darauf zu achten, ob das Beta-Gewicht einer Variablen gegenüber ihrer Validität deutlich erhöht ist.

- x_2 und x_3 sind **Suppressor-Variablen**. Sie sind unkorreliert mit dem Kriterium aber hoch korreliert mit dem Prädiktor.
  - Eine **Suppressorvariable** ist eine Variable, die den Vorhersagebeitrag einer (oder mehrerer) anderer Variablen erhöht, indem sie irrelevante Varianzen in der (den) anderen Prädiktorvariablen unterdrückt.
  - Die Suppressorvariable gibt sich in der multiplen Regressionsgleichung durch ihr **negatives Beta-Gewicht** zu erkennen.
- Die multiple Korrelation $R_{y \cdot x_1 x_2} = 0.46$ bedeutet, dass gegenüber der einfachen Korrelation zwischen y und x_1 die Berücksichtigung der Suppressorvariable einen Zuwachs von 0,06 gebracht hat.

- Die multiple Korrelation $R_{y \cdot x_1 x_3} = 0.42$ bedeutet, dass gegenüber der einfachen Korrelation zwischen y und x_1 die Berücksichtigung der Suppressorvariable einen Zuwachs von 0,02 gebracht hat.

■ Wenn die Bedingung $r_{ys} = 0.00$ erfüllt ist, die Suppressor-Variable also nicht mit dem Kriterium korreliert, lassen sich die Formeln sehr vereinfachen. U.a.:

(9) $r_{ys} = 0.00$ folgt für (4)

(10) $b_1 = \frac{[r_{yx_1} - r_{ys} \cdot r_{x_1 s}]}{[1 - r^2_{x_1 s}]} = \frac{[r_{yx_1}]}{[1 - r^2_{x_1 s}]}$ folgt für (5)

(11) $b_2 = \frac{[r_{ys} - r_{yx_1} \cdot r_{x_1 s}]}{[1 - r^2_{x_1 s}]} = - \left[\frac{[r_{yx_1} \cdot r_{x_1 s}]}{[1 - r^2_{x_1 s}]} \right]$ (negatives Beta-Gewicht)

für (6)

(12) $R^2_{[y \cdot y_1 s]} = b_1 r_{yx_1} + b_2 r_{ys} + b_2 r_{ys} = b_1 r_{yx_1}$ uns daraus schließlich für (7)

(13)
$$R^2_{[y \cdot x_1 s]} = \frac{[r^2_{yx_1} + r^2_{ys} - 2 r_{yx_1} r_{ys} r_{x_1 s}]}{[1 - r^2_{[x_1 s]}]} = \frac{[r^2_{yx_1}]}{[1 - r^2_{x_1 s}]}$$

- Die multiple Korrelation zwischen y und $x_1 s$ kann nun sehr einfach bestimmt werden:

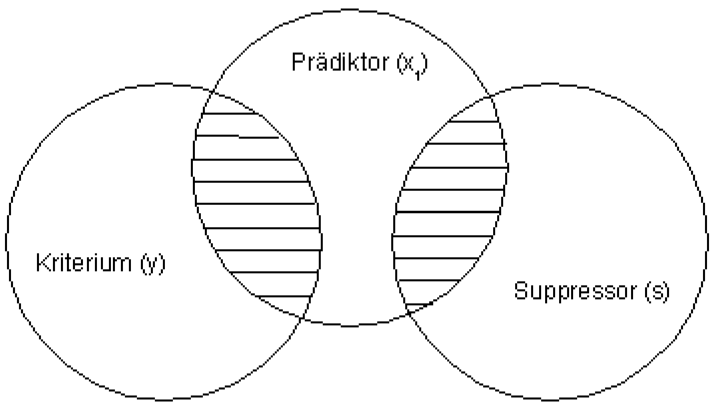
(14) $R_{[y \cdot x_1 s]} = \frac{r_{yx_1}}{\sqrt{1 - r^2_{x_1 s}}}$

- Ist $r_{ys} = 0$, dann ist die multiple Korrelation zwischen dem Kriterium und dem Prädiktor und der Suppressor-Variablen dann größer als die einfache Korrelation zwischen Kriterium und Prädiktor, wenn die Korrelation zwischen Prädiktor und Suppressor verschieden von 0 ist. Das Vorzeichen der Korrelation x_1 und s ist dabei bedeutungslos, da $r_{x_1 s}$ ja in (14) quadriert wird.
- $R_{[y \cdot x_1 s]}$ ist umso größer, umso höher Prädiktor und Suppressor miteinander korrelieren. Die zugeordnete multiple Regressionsgleichung:

1

(15) $y' = \left[\frac{r_{yx_1}}{[1 - r^2_{x_1 s}]} \right] \cdot x_1 - \left[\frac{r_{¿} \cdot [r_{x_1 s}]}{[1 - r^2_{x_1 s}]} \right] \cdot s$

◆ **Veranschaulichung des Suppressor-Effektes**:



Kreise: jeweilige Varianz der y-, x_1- und s-Werte. **Schraffierte Flächen**: gemeinsame Varianz von Kriterium und Prädiktor bzw. Prädiktor und Suppressor. **Über-**

schneidungen: Korrelation zwischen Prädiktor mit Kriterium und Suppressor mit Prädiktor. **Keine Überschneidung**: Nullkorrelation zwischen Suppressor und Kriterium

- Der **Varianzteil, den Prädiktor und Kriterium nicht gemeinsam haben, ist irrelevant für die Kriteriumsvorhersage aufgrund des Prädiktors und stellt unnötigen „Ballast“** dar. Eine Verbesserung der Kriteriumsvorhersage ist möglich, wenn dieser irrelevante Varianzanteil reduziert werden kann. Dies geschieht durch die Suppressor-Variable s, die einen Teil der Varianz „unterdrückt“, also gewissermaßen abzieht.Diese subtrahierende Funktion drückt sich in der multiplen Regressionsgleichung durch das negative Beta-Gewicht aus.
- In der psychologischen Diagnostik kommen v.a. *Antwortstile als Suppressor-Variablen* vor, die als Reaktionstendenzen zu einer Verfälschung von Testergebnissen führen können. Ihre Berücksichtigung kann die Kriteriumsvorhersage verbessern.
  - **Bsp.**: Der Prädiktor „Examensnote“ korreliert nur mäßig und der Prädiktor „Prüfungsangst“ praktisch gar nicht mit dem Kriterium „beruflicher Erfolg“. Eigentlich wäre zu erwarten, dass auch die multiple Korrelation beider Prädiktoren mit dem Kriterium nicht besonders hoch ausfällt. Dem ist aber nicht so. In der multiplen Korrelation werden die beiden Variablen so kombiniert, dass der für den beruflichen Erfolg irrelevante, auf Prüfungsangst beruhende Varianzanteil in der Variablen „Examensnote“ um den störenden Varianzanteil, der eine höhere Korrelation der Examensleistung mit dem beruflichen Erfolg verhindert hat. Kurz: Die Variable „Prüfungsangst“ ist in Kombination mit der Variablen „Examensleistung“ und dem Kriterium „beruflicher Erfolg“ eine Suppressorvariable.

11. Worin unterscheidet sich eine Moderatorvariable von einer einfachen Prädiktorvariablen? Geben Sie 2 Beispiele für Moderatorvariablen an.

- **Moderatorvariablen** beeinflussen die Werte der Korrelation zwischen Prädiktor und Kriterium.
  - Die Moderatorvariable geht nicht in die Regressionsgleichung ein. Die Moderatorvariable moderiert den Zusammenhang zwischen Prädiktor und Kriterium, d.h. **der Beitrag, den ein Prädiktor zur Vorhersage des Kriteriums leistet ist abhängig vom Wert des Moderators**.
  - **Klassische multiple Regressionsgleichung**. (Für die Vorhersage anhand von Prädiktorvariablen – *ohne Einfluss von Moderatorvariablen*.)
    - In der klassischen multiplen Regressionsgleichung $y' = b_1 x_1 + b_2 x_2$ nimmt y' um den Wert b_1 zu, wenn x_1 um eine Einheit steigt und x_2 konstant bleibt. **Die Prädiktoren sind unabhängig voneinander**.
  - **Linear joint function**. (Berücksichtigung des Einflusses von *Moderatorvariablen*.)
    - In einer *linear joint function* $y' = b_1 x_1 + b_2 m + b_3 x_1 m \quad (x_2 = m = Moderator)$ nimmt y' um $b_1 + b_3 m$ zu, wenn x_1 um eine Einheit steigt und m konstant ist. **Die Unabhängigkeit der Prädiktoren geht hier verloren** ($b_1 x_1$ • Prädiktoreffekt; $b_2 m$ • Moderatoreffekt; $b_3 x_1 m$ • Interaktion).
  - Die **Einbeziehung von Moderatorvariablen** ist dann sinnvoll, wenn die Korrelation zwischen geschätzten und tatsächlichen Werten bei Berücksichtigung der Moderatoreffekte deutlich höher ist (sonst Moderator wie einfachen Prädiktor behandeln).
  - **Moderatoreffekte** lassen sich entdecken:
    - indem aus der Stichprobe Kriteriums, Prädiktor- und Moderatorwert erhoben werden;
    - anschließend versucht wird, die gefundenen Werte mit der *linear joint function* zu approximieren;
    - wenn sich dabei ein von 0 verschiedenes b_3 -Gewicht ergibt, liegt ein Moderatoreffekt vor.
  - **Beispiele**.
    - kontinuierliche Moderatoren: z.B. Alter oder Intelligenz.
    - sonstige: Sozialstatuszugehörigkeit, Geschlecht
  - **Analogie zur Varianzanalyse**.
    - $b_1 x_1$ entspricht dem Effekt des ersten Faktors (Prädiktor)
    - $b_2 m$ entspricht dem Effekt des zweiten Faktors (des Moderators)
    - $b_2 x_1 m$ entspricht dem Effekt der Wechselbeziehung zwischen den beiden Faktoren

(Prädiktor und Moderator)

12. Welche Vor- und Nachteile gegenüber dem Moderatoransatz bietet Ghisellis Konzept der Vorhersage der Vorhersagbarkeit?

- Der Ansatz von GHISELLI beschäftigt sich grundsätzlich mit der Frage, ob sich die **Vorhersagbarkeit eines Kriteriums aufgrund eines Prädiktors vorhersagen** lässt.
- Seine Grundidee war die Suche nach einem einfachen, ökonomisch anwendbaren **Maß** x_2, das die **Differenz D der vorhergesagten von den tatsächlichen Kriteriumswerten vorhersagen kann** (durch hohe Korrelation zwischen x_2 und D). Bei Kenntnis der Vorhersagbarkeitsvariablen x_2 kann von vornherein entschieden werden, ob eine Messung des Prädiktorwertes (Testanwendung) überhaupt sinnvoll ist (x_2 muss klein sein). Diese Kenntnis ist ein großer Vorteil gegenüber dem Moderatoransatz.
- Nachteilig ist, dass die **Personenzahl auf die Extremgruppen reduziert** wird und dass das Modell nur im Extrembereich gute Korrelationen erreicht, nicht aber im mittleren Bereich, wo die meisten Werte liegen (vgl. Beispiel).
  - Bei x_2 handelt es sich auch um eine **Moderatorvariable**, die allerdings nicht in eine linear joint function eingeht, sondern Personen nach ihrer Vorhersagbarkeit differenziert und damit ähnlich funktioniert wie eine Moderatorvariable auf Nominalskalenniveau.

- **Vorgehen**.
  1. **Entwicklung des Vorhersagbarkeitsmaßes**.
     i. **Berechnung: Kriteriumswert vs. tatsächlicher Kriteriumswert**. Für jeden Probanden wird die Abweichung des vorhergesagten Kriteriumswertes vom tatsächlichen Kriteriumswert errechnet (absoluter Betrag $D = | y - y' |$).
     ii. **Suche nach der Vorhersagbarkeitsvariablen**. Dann wird nach einer Variablen x_2 gesucht, die hoch mit D korreliert. Bei einem kleinen Wert von x_2 lässt sich das Kriterium gut vorhersagen. Bei einem großen Wert von x_2 lässt sich das Kriterium nicht gut vorhersagen. Ist die Ausprägung von x_2 also bekannt, lässt sich entscheiden, ob die Vorhersage lohnt oder nicht.
  2. **Kreuzvalidierung**. Im Rahmen der Kreuzvalidierung wird untersucht, ob sich bei der Berücksichtigung des Vorhersagbarkeitsmaßes eine Verbesserung der Vorhersage ergibt.
  - Beispiel: Vorhersage von Geselligkeit durch Intelligenz.
    - Bildung zweier Gruppen (hohes/ niedriges D) und Durchführung einer Aufgabenanalyse (welche Items sagen D am besten vorher = x_2).
    - Dann Überprüfung des neuen Prädiktors x_2 (Kreuzvalidierung). Dazu werden 2 Extremgruppen gebildet mit sehr hohem bzw. sehr niedrigem x_2 -Wert gebildet. Es ergeben sich unterschiedliche Korrelationen von Prädiktor und Kriterium.
- **Fazit**: Lohnt nur bei aufwendiger Datenerhebung. Bei einem hohen D-Maß • eventuell Verzicht auf Erhebung

13. Welche Ziele werden mit der Erweiterung des einfachen Regressionsansatzes durch die Einführung von Moderator-, Suppressor- und Vorhersagbarkeitsvariablen verfolgt?

- Ziele sind die **Verbesserung der Validität**, und zwar der kriteriumsorientierten Validität (Güte der Vorhersagbarkeit eines Kriteriums; d.h. größeren Anteil an der Gesamtvarianz des Kriteriums aufzuklären) und eine ökonomischere Testkonstruktion. Zusätzliche Prädiktoren bewirken niemals eine Verschlechterung der Vorhersagbarkeit!

- **Suppressorvariable**.
  - geht als Prädiktor in die multiple Regressionsgleichung ein;
  - korreliert nicht mit dem Kriterium, aber (hoch) mit dem Prädiktor;
  - **unterdrückt Varianz im Prädiktor**, die nicht relevant ist für Kriteriumsvorhersage und führt somit zu Verbesserung der Kriteriumsvorhersage;
    - Z.B. ›soziale Erwünschtheit‹ • etwa durch Unterskalen wie ›Lügenskala‹.
  - ist erkennbar an negativen Beta-Gewicht und Nullkorrelation ($r_{sy} = 0$).
- **Moderatorvariable**.
  - geht in die *linear-joint function* als Moderator und Interaktion zwischen Moderator und Prädiktor ein;
  - Einbeziehung **erhöht aufgeklärte Varianz** im Kriterium – ein Moderatoreffekt ist zu identifizieren, da sonst der ermittelte Zusammenhang fälschlicherweise auf die Prädiktoren zurückgeführt würde;
  - Erhöhung der differentiellen Validität;
  - höherer multipler Regressionskoeffizient.
- **Vorhersagbarkeitsvariable**.
  - kann **Anwendung oder Konstruktion eines Tests effizienter** machen (Selektion gut vorhersagbarer Personen);
  - vergleichbar einer Moderatorvariable auf Nominalskalenniveau.
- ABER: Aufwand bei allen oft größer als Nutzen.

14. Bei der Selektion von Bewerbern für das Studienfeld Medizin wollen wir davon ausgehen, dass 75% der Bewerber um einen Studienplatz für das Studium geeignet sind und dass die Selektionsrate etwa 20% beträgt.

a) Welchen Validitätswert kann unter diesen Randbedingungen ein Eignungstest maximal erreichen?

$$BR = 0.75 \qquad SR = 0.20 \qquad P(VP) = maximal\, 0.20$$

- Maximale Validität, d.h. ein maximales Phi (Φ), ist gegeben, wenn die Basisrate gleich der Selektionsrate ist.
- $BR = SR \rightarrow \Phi = 1$
- Je höher die Diskrepanz zwischen BR und SR, umso niedriger ist das maximale Φ . Unter diesen Bedingungen ist die maximale Validität gegeben, wenn VP (Valide Positive) = Selektionsrate.

b) Nehmen wir an, von den aufgrund ihres Ergebnisses in einem Eignungstest zugelassenen Studienbewerbern fallen 15% bei der 1. ärztlichen Vorprüfung durch. Welche Validität hat dann der Eignungstest gehabt?

- Die Anzahl der validen Positiven und falschen Positiven ändert sich: Es sind jetzt nur noch 15 % der anfangs valide Positiven tatsächlich valide Positiv. Die Anzahl der falschen Positiven steigt entsprechend.

c) 25.000 Studenten haben sich um einen Studienplatz beworben. Es soll berechnet werden, wie sich die Anzahl der Falschen Positiven verändert, wenn man statt einer reinen Zufallsauswahl einen Eignungstest mit der unter (b) berechneten Validität unter den oben genannten Randbedingungen zum Einsatz bringt.

d) Berechnen Sie den zu erwartenden Nutzen, der sich bei einer solchen Verwendung des Eignungstests ergibt, und vergleichen Sie ihn mit dem zu erwartenden Nutzen, der bei einer reinen Zufallsauswahl entsteht. Einem VP ent-

spreche ein Nutzen von +3, einem FP ein Nutzen von –2, einem FN entspreche ein Nutzen von –1 und einem VN ein Nutzen von +1. Die Kosten für den Eignungstest betragen für 25.000 Bewerber insgesamt U1 = 1000.

- ■ <u>nicht ausgearbeitet</u> – nur eine Sammlung aus dem Skriptmaterial!

- Der Validitätskoeffizient Φ kann nur dann den Wert 1.00 erreichen, wenn $BR = SR$. Je höher die Diskrepanz zwischen BR und SR, desto niedriger ist das maximale Φ.
- Bsp.: $CD(Correct\,Decisions) = P(VP) + P(VN) = 0.79$.
  - Die Proportion der korrekten Entscheidungen CD ist 79 %. Der Test verhilft also zu 79% richtigen Entscheidungen. Um beurteilen zu können, welchen Wert der Test in der Situation wirklich hat, muß man vergleichen, wie die Entscheidungen aussehen, wenn sie aufgrund einer <u>zufälligen Auswahlstrategie</u> getroffen werden, wenn also $\Phi = 0$.
    - Dann gilt: $P(VP) = BR \cdot SR$.
- Der Testeinsatz bringt um so bessere Ergebnisse, je mehr die BR sich im mittleren Bereich bewegt.
- Bei extremen BR ist die <u>inkrementelle Validität</u>, d.h. die Verbesserung der Proportionen der korrekten Entscheidungen über das Zufallsniveau hinaus, nur noch gering. Die inkrementelle Validität richtet sich nach der Anzahl der korrekten Entscheidungen

- ◆ Ein Test ist bei **extremen SR** <u>weniger effektiv bezüglich der Proportion der korrekten Entscheidungen</u> als bei mittleren SR.
  - ➢ $BR > SR$ • viele in P (FN)
  - ➢ $BR > SR$ • viele in P (FP)

- ➢ Bei Problemen der psychologischen Diagnostik ist der **Nutzen eines Ergebnisses** im allgemeinen nicht leicht zu bestimmen. Da aber <u>jede Entscheidung Konsequenzen nach sich zieht</u>, kann eine rationale Entscheidung nur gefällt werden, wenn versucht wird, die Konsequenzen zu antizipieren.
  - Angenommen, das Problem der Bestimmung des Nutzens einer diagnostischen Entscheidung sei gelöst, so bestimmt sich der **erwartete Gesamtnutzen** einer Entscheidung

$$EU = U_1 \cdot P(VP) + U_2 \cdot P(FP) + U_3 \cdot P(FP) + U_4 \cdot P(VN) - U_t$$

  - ◆ **Selektionsschema**.

Daten		*Entscheidung*		*Ergebnisse*	*Nutzen*
		Annehmen (accept)	D_A	Valide Positive (VP)	U_1
				Falsche Positive (FP)	U_2
Test A	Cut-off-Wert				
		Ablehnen (reject)	D_R	Falsche Negative (FN)	U_3
				Valide Negative (VN)	U_4

  - ◆ Nach diesem Modell sind <u>*nichtsequentielle (a)*</u> und <u>*sequentielle (b)*</u> **Entscheidungsstrategien** möglich:
    - a) <u>einmalige Durchführung eines oder mehrerer Tests</u> • endgültige Entscheidung über Annahme oder Ablehnung
    - b) <u>ein Teil der Individuen aufgrund eines ersten Tests</u> • Ablehnung oder Aufnahme; größerer Teil • ein oder mehrerer weiterer Tests
    - ➢ Die <u>sequentielle Strategie</u> ist bezüglich des erwarteten Nutzens jeder nichtsequentiellen Strategie <u>überlegen</u>.