

## II. Itemanalyse Deskriptive Analyse von Items

- 1) Schwierigkeit
- 2) Itemvarianz
- 3) Trennschärfe
- 4) Itemselektion und Revision
- 5) Testverteilung und Normalisierung

Itemanalyse = Erst Testgruppe von ca. 20 Personen für erste grobe Analyse, z.B. auf Mehrdeutigkeit, Unverständlichkeit. Dann: Eine für die Zielgruppe möglichst repräsentative Stichprobe bearbeitet den Testentwurf um empirische Daten über die Items zu gewinnen. Auf der Grundlage der Daten werden Kennwerte bestimmt, die Entscheidungen über Items zugrunde liegen -> Schwierigkeit, Itemvarianz, Trennschärfe

### 1. Schwierigkeitsanalyse

Wie viel Gewicht liegt auf einem Item?

= Itemschwierigkeit gibt an, wie groß der Anteil an Personen ist, die das Item „im Sinne des Merkmals“ (bei Leistungstest) beantwortet haben.

#### Leistungstests

**Achtung:**

1. Schwierigkeit ist nicht zu verwechseln mit Leistung des Einzelnen
2. Numerische Höhe des Schwierigkeitsindex kennzeichnet eigentlich die „Leichtigkeit“ des Items: Je höher der Wert, desto mehr Personen haben das Item gelöst, desto leichter ist also das Item

$$p = \frac{N_R}{N}$$

$p$  = Itemschwierigkeit  
 $N_R$  = Zahl der Probanden, die die Aufgabe im Sinne des Merkmals beantwortet haben  
 $N$  = Zahl der Probanden, die das Item beantwortet haben  
 Wertebereich liegt also zwischen 0 und 1

$p = 0$  | Niemand hat das Item im Sinne des Merkmals (korrekt) beantwortet

$p = 1$  | Alle haben das Item im Sinne des Merkmals (korrekt) beantwortet

Unterscheidung zwischen **Speedtests** (Geschwindigkeit) und **Powertests** (Niveautests)

#### Speedtest

R = richtig

F = falsch

A = ausgelassen

U = unbearbeitet

Proband	Item 1	Item 2	Item 3	Item 4
Proband 1	R	R	F	U
Proband 2	R	F	U	U
Proband 3	A	R	A	R
$n_R$	2	2	0	1
$n_F$	0	1	1	0
$n_A$	1	0	1	0
$n_U$	0	0	1	1
$P_i$	0.667	0.667	0	1.00

$$p = \frac{N_R}{N}$$

$$N_R = n_R$$

$$N = n_R + n_F + n_A$$

- Item wird nur dann als unbearbeitet markiert, wenn danach kein Item mehr bearbeitet wurde, d.h. wenn danach ein Item bearbeitet wurde, wird das unbearbeitete als ausgelassen markiert
- Unbearbeitete Items werden in der Schwierigkeitsanalyse nicht mitberechnet, da sie aus Zeitgründen nicht bearbeitet wurden (-> Speedtest)
- Korrigierter Schwierigkeitsindex durch Inangriffnahmekorrektur (Wurde Item überhaupt bearbeitet?)
- Randomisierung der Items!

**Powertest**

$$p = \frac{N_R}{N}$$

$$N_R = n_R$$

$$N = n_R + n_F + n_A$$

Proband	Item 1	Item 2	Item 3	Item4
Proband 1	R	R	F	A
Proband 2	R	F	A	A
Proband 3	A	R	A	R
$n_R$	2	2	0	1
$n_F$	0	1	1	0
$n_A$	1	0	2	2
$P_i$	0.667	0.667	0	0.333

- Korrigierter Schwierigkeitsindex durch Ratekorrektur

- Bei Leistungstest sind Itemschwierigkeiten von **.20 bis .80** anzustreben.
- Itemstreuung ist bei mittlerer Schwierigkeit am größten
- Itemschwierigkeiten sollen variieren, am besten normalverteilt, die meisten Items mittelschwer
- Variation der Schwierigkeiten ermöglicht Differenzierung in Extrembereichen
  - Items, die bei Personen mit mittlerer Merkmalsausprägung optimal differenzieren, haben bei Personen mit extremer Merkmalsausprägung eine sehr geringe Streuung und können so in den sehr hohen und sehr leichten Bereichen nicht gut differenzieren

## Persönlichkeitstests

**Vorbereitung**

- umpolen (invertieren) negativ formulierter Items
- transformieren: -3 -2 -1 0 1 2 3 → 0 1 2 3 4 5 6
  - Nullpunkt wird bei der Schwierigkeitsberechnung benötigt

**Berechnung**

$$p = \frac{\sum_{i=1}^n X_{im}}{n \cdot (k-1)}$$

$k$  = Anzahl der Stufen einer Skala (hier  $k=7$ )

$(k-1)$  = Relativierung der Mittelwerte an Anzahl der Stufen

$\sum_{i=1}^n X_{im}$  = Summe der Ausprägungswerte aller Personen

Im Beispiel:

$$p = \frac{\sum_{i=1}^n X_{im}}{n \cdot (k-1)}$$

$k = 7$  (0 bis 6)  
 $n = 4$

Spalten- summe	6	13	12	: 24	(4 x (7 - 1))
p	.25	.54	.50		

- Bei Persönlichkeitstests sind Items schon inhaltlich variiert. Trotzdem ist eine Variation in der Schwierigkeit möglich, aber nicht notwendig

## 2. Itemvarianz

= Differenzierungsfähigkeit eines Items hinsichtlich der untersuchten Probandenstichprobe

- Die Itemschwierigkeit begrenzt die mögliche Ausprägung der Itemvarianz
  - Sehr schwere und sehr leichte Items geben weniger Varianz

### Verteilungseigenschaften

- Ausnutzung der Bandbreite
  - Sind alle möglichen Antwortmöglichkeiten ausgenutzt worden?
- Schiefe und Kurtosis
  - Verteilung der Itemantworten – Kolmogorov-Smirnov-Test auf Normalverteilung
- Varianz
  - z.B. bei 5-stufiger Skala: Varianz von 1 optimal

### Berechnung

Bei intervallskalierten Items:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Bei dichotomen Items:

$$s^2 = p(1 - p)$$

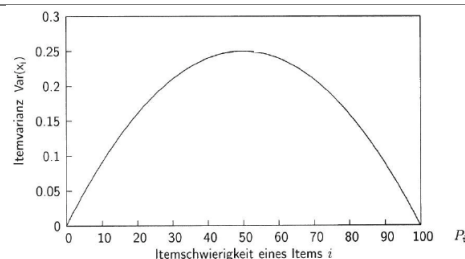
[p = Schwierigkeit]

### Zusammenhang zwischen Schwierigkeit und Itemvarianz

#### Intervallskaliertes Antwortformat

- Die Schwierigkeit eines Items begrenzt die maximal mögliche Itemvarianz.
- Itemvarianz kann bei mittlerer Schwierigkeit die höchsten Werte erreichen.
- Je extremer die Schwierigkeit, desto geringer ist das Potential des Items, eine hohe Varianz zu erreichen

#### Dichotomes Antwortformat



### 3. Trennschärfe

- Korrelation zwischen den **Itemwerten** (eines Items) und den **Testwerten** der Probanden
- *in welchem Ausmaß die durch das Item erfolgte Differenzierung der Probanden in Löser und Nichtlöser mit derjenigen durch die Skala als Ganzes übereinstimmt*

#### Korrelation

##### Kovarianz:

$$\text{COV}_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}$$

$$\text{max} \longrightarrow s_x \cdot s_y$$

Kovarianz einer Variablen mit sich selbst:

$$\text{COV}_{xx} = s_x^2$$

##### Produkt-Moment-Korrelation

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x \cdot s_y \cdot n}$$

$x_i$  = Itemwert des Probanden

$y_i$  = Testwert des Probanden i

- Punktbiseriale Korrelation
  - ↳ Bei dichotomen Antworten  
(1 intervallskaliert (Testwert) & 1 natürlich dichotome Variable (Itemantwort))
- Biseriale Korrelation
  - ↳ Bei dichotomen Antworten bei „eigentlich“ vorliegendem Kontinuum  
(1 intervallskaliert (Testwert) & 1 künstlich dichotome Variable (Itemantwort))

#### Part-Whole-Korrektur

= Korrektur von **Autokorrelationen**

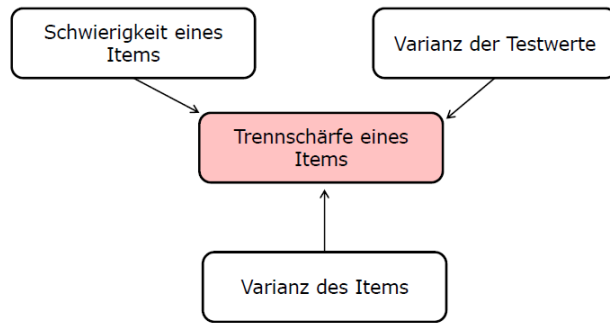
*(Bei Korrelation von Itemwert mit Testwert, korreliert man auch immer Itemwert mit sich selbst, da Testwert aus Itemwerten zusammengesetzt wird. Bei wenig Items ist Autokorrelation größer)*

- man berechnet die Trennschärfe nicht mit dem Testwert  $y_i$ , sondern mit dem um den Itemwert (des gerade betrachteten Items) bereinigten Testwert
- Einfluss geringer, wenn (1) viele Items und (2) homogene Items zur Berechnung herangezogen werden

#### Interpretation

- Trennschärfe nahe bei 1:
  - ↳ Item differenziert (bezogen auf das Merkmal) in ähnlicher Weise wie der Gesamttest.  
**Gute Trennschärfe = .40 bis .70**
  - ↳ Trennschärfe = 1 -> identische Items, nicht facettenreich
- Trennschärfe nahe bei 0:
  - ↳ Kein Zusammenhang bezüglich der Differenzierung des Items und der Differenzierung durch den Gesamttest.
- Trennschärfe nahe bei -1:
  - ↳ Item differenziert in nicht intendierte Richtung. Invertiert? Itemformulierung?

## Zusammenhang mit anderen Kennwerten



*Items mit geringer Trennschärfe nicht direkt entfernen, sondern vorher Abhängigkeiten von anderen Kennwerten beachten*

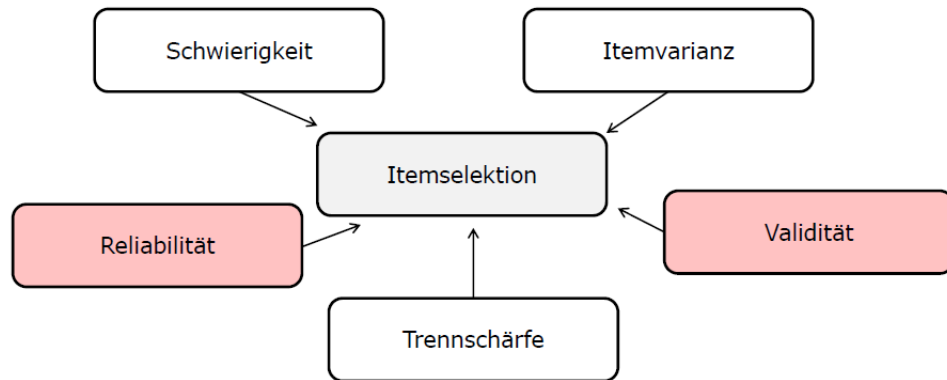
- Trennschärfe und Schwierigkeit: meist umgekehrt u-förmige Beziehung
  - Sehr hohe und sehr niedrige Schwierigkeit -> geringe Trennschärfe
- Verteilungsform von Item und Testwert
  - Sind Item und Testwert in gleicher Weise (!) (egal ob beide normalvert., rechtsschief, etc.) verteilt, wird die Korrelation (Trennschärfe) dadurch nicht eingeschränkt
  - Sind Item und Testwert unterschiedlich verteilt, wird Korrelation eingeschränkt
- Streuung der Testwerte
  - Homogene Stichprobe / geringe Varianz auf Merkmal -> geringe Trennschärfe
  - z.B. bei Varianzeinschränkung
- Streuung der Itemwerte
  - Gleich: Homogene Stichprobe / geringe Varianz auf Merkmal -> geringe Trennschärfe
  - Hängt wiederum von Itemschwierigkeit ab ([s.o.](#))

- Zusammenhang zwischen Schwierigkeit und Trennschärfe beachten
- Berücksichtigung der Homogenität/Heterogenität des erfassten Merkmals
- .4 bis .7 gibt als gute Trennschärfe
- Richtwert Trennschärfe  $\geq .30$

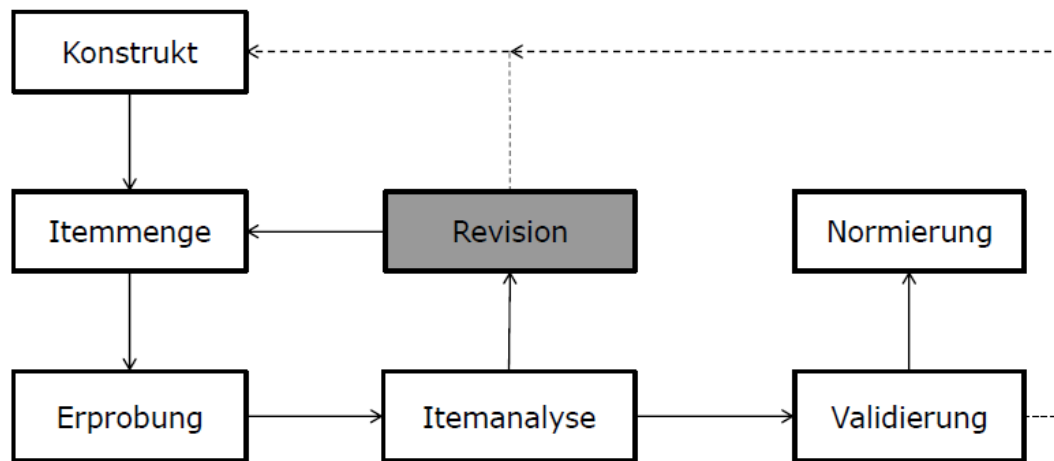
- Homogenität bezieht sich auf die Beschaffenheit einer Skala und nicht direkt eines einzelnen Items.
- Homogenität liegt vor, wenn bezogen auf die Items einer Skala formale und inhaltliche Einheitlichkeit vorliegt.
- Homogene Skalen beschreiben eng umschriebene Merkmalsaspekte.

#### 4. Itemselektion und Revision

= Berücksichtigung aller Kennwerte simultan unter Einbezug des Einsatzzweckes des Messinstrumentes



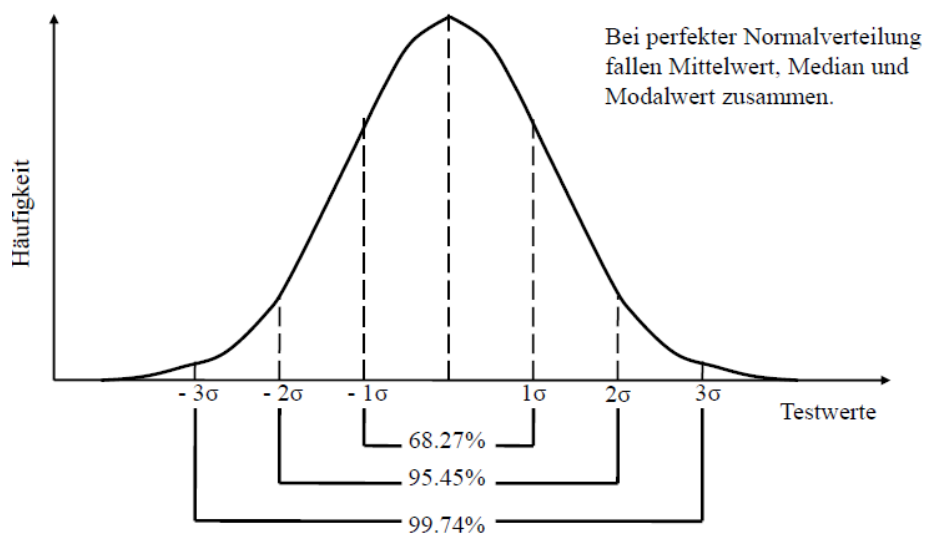
- Schwierigkeit, Itemvarianz, Trennschärfe sind nur Richtwerte.
- Am Ende können z.B. inhaltsvalide Items trotz schlechter Werte im Test behalten werden.
- Individualität je nach Item -> Subjektivität



bei jeder Abänderung der Itemmenge muss eigentlich neu erprobt werden. Dieser Zyklus kann so oft durchlaufen werden bis Itemanalyse gut läuft.

#### 5. Testwertverteilung und Revision

Nach der Itemselektion kann die Verteilung der Testwerte (Gesamtwerte) betrachtet und beurteilt werden.

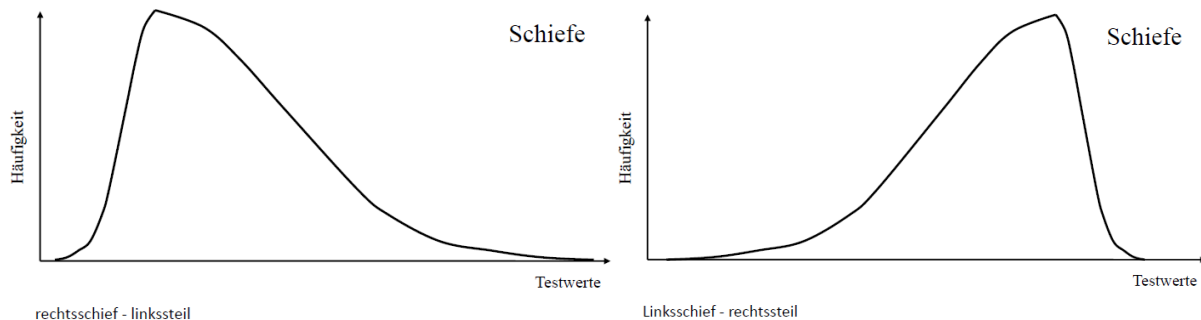


*Normalverteilung der Testwerte ist ideal*

## Maße der zentralen Tendenz

- Arithmetisches Mittel
- Modalwert
- Median

## Schiefe



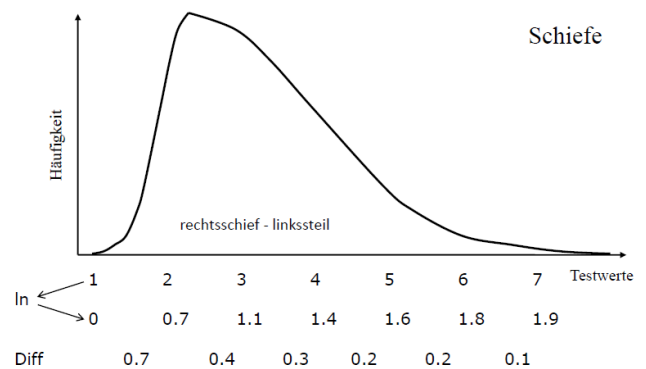
Schiefe ist problematisch, da die Ergebnisse der nachfolgenden Analyse verzerrt werden.

Gründe für Schiefe Verteilungen:

- Itemauswahl nicht ausgewogen hinsichtlich der Schwierigkeit
- heterogene Stichprobe (aus Unterstichproben zusammengesetzt/gepatched -> Substichproben sind zu heterogen)
- keine Normalverteilung des Merkmals in der Population gegeben (z.B. Optimismus)

Lösung:

- Optimierung der Itemauswahl
- Normalisierung (nichtlineare Transformation)
  - Logarithmieren:  $xv' = \ln xv$  (bei rechtsschiefer Verteilung)
  - Flächentransformation (Verteilungsunabhängig einsetzbar)



Logarithmieren für rechtsschiefe Verteilung