

Von der Formulierung zur Bewertung von Hypothesen

Seminar Forschungsmethoden I

P. Wilhelm

1. Kennzeichen wissenschaftlicher Hypothesen
2. Arten von wissenschaftlichen Hypothesen
3. Von der Forschungshypothese zur statistischen Prüfung
 - 3.1 Voraussetzungen für die Überprüfung von Hypothesen
 - 3.2 Operationalisierung
 - 3.3 Logik der Signifikanzprüfung
4. Von der Signifikanzprüfung zur Bewertung der Forschungshypothese
 - 4.1 Entscheidung über die Testhypothese
 - 4.2 Simultankontrolle der Determinanten des Signifikanztests
 - 4.3 Entscheidung über die Statistische Vorhersage
 - 4.4 Bewertung der Operationalen Hypothese
 - 4.5 Bewertung der Forschungshypothese
5. Zusammenfassung

Literatur

Fragen



1. Kennzeichen wissenschaftlicher Hypothesen

Was ist eine wissenschaftliche Hypothese?

- theoretisch begründete Vermutungen bzw. Annahmen über Zusammenhänge von Sachverhalten
- provisorische Antworten auf ein wissenschaftliches Problem

Kennzeichen einer wissenschaftlichen Hypothese:

- allgemeingültige über den Einzelfall hinausgehende Behauptungen
- Formalstruktur eines Konditionalsatzes (wenn-dann-Satz, je-desto-Satz, Wenn-Teil \approx Antezedenz \approx u.V, Dann-Teil \approx Konsequenz \approx a.V.)
- potentiell falsifizierbar
- *vorläufig*
- *überprüfbar*

Generierung von Hypothesen

Induktiv: von Einzelereignissen zu allgemeineren Vermutungen (z.B. Pawlow, Freud)

Deduktiv: aus einer Theorie werden spezielle (neue) Vermutungen abgeleitet

3.1 Voraussetzungen für die Überprüfung von Hypothesen

- Widerspruchsfreiheit
- Operationalisierbarkeit
- prinzipielle Widerlegbarkeit
- Detaillierte Formulierung vor der Datenerhebung (H. in Panzerschrank)

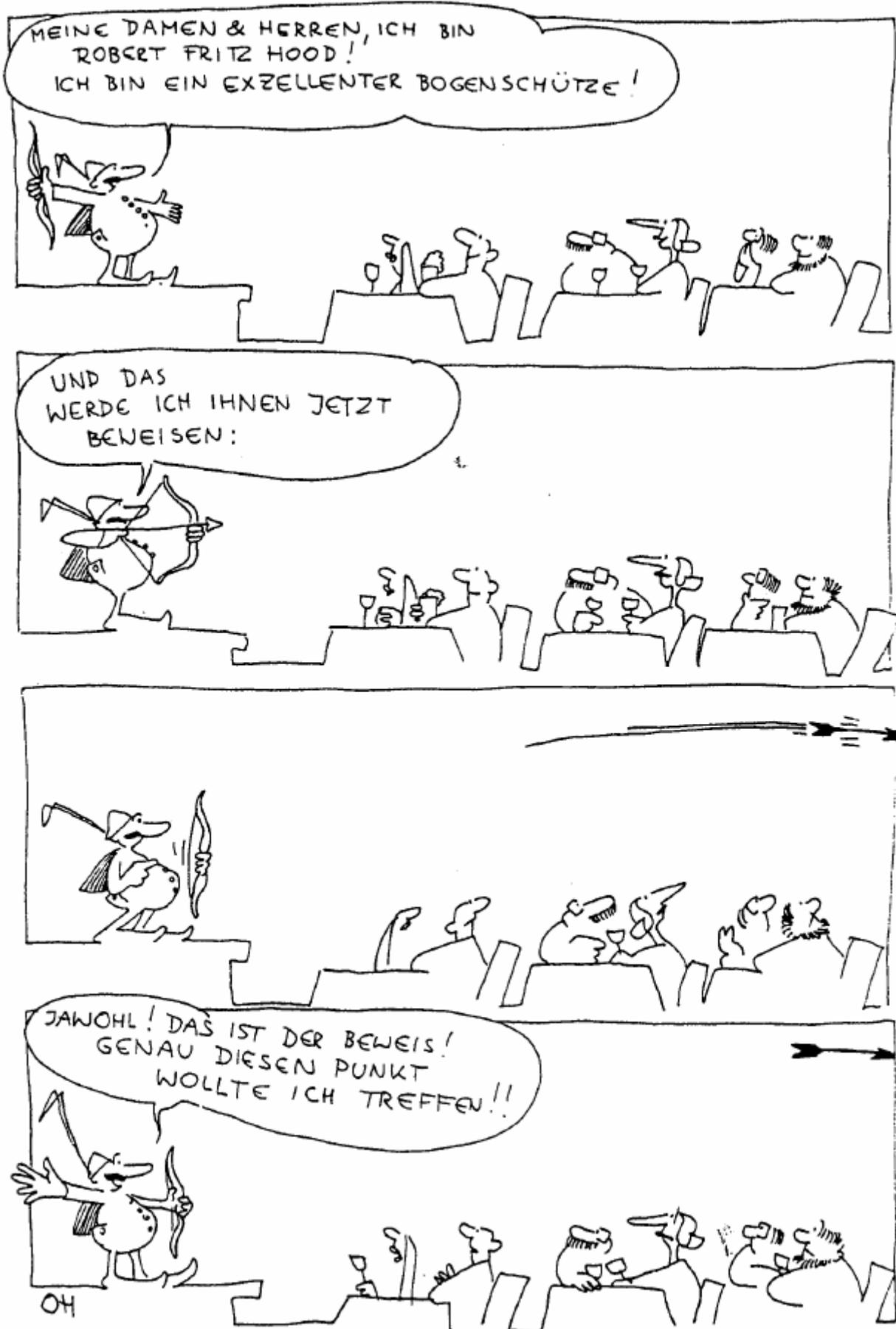


Abbildung 2: Warum eine Hypothese vor der Prüfung aufgestellt werden muß.

3. Von der Forschungshypothese zur statistischen Prüfung

Forschungshypothesen (theoretisch-inhaltliche H.)
enthalten nicht direkt beobachtbare theoretische Konstrukte und
Relationen zwischen diesen Konstrukten

Operationalisierung und Versuchsplan
(Bedeutungsverlust)



Operationale Hypothesen (empirisch-inhaltliche H.)
enthalten operational definierte beobachtbare Begriffe und
Definition der Relation zwischen diesen Begriffen

Transfer auf statistische Ebene, weil es eine Diskrepanz zwischen den
erhobenen Daten und den wahren Werten gibt



Statistische Vorhersage
z.B. Vorhersagen von Zusammenhängen zwischen Variablen
oder Mittelwertsunterschieden...
macht implizit zusätzliche Annahmen (z.B. Skalenniveau)

Wahl des statistischen Prüfverfahrens



Testhypothesen H_0 , H_1
Signifikanztest ermittelt Wahrscheinlichkeit mit der empirisches Ergebnis
auftreten könnte, wenn in der Population H_0 gilt.
macht implizit zusätzliche Annahmen
(Voraussetzungen der Signifikanzprüfung)

3.2 Operationalisierung

Theoretischen Konstrukten müssen empirische Indikatorvariablen (beobachtbare Begriffe) zugeordnet werden. Diese sind mehr oder minder gute Indikatoren für das theoretische Konstrukt und bilden dieses mehr oder minder fehlerbehaftet ab.

Beispiel: Einige Möglichkeiten der Operationalisierung von Aggression: Anzahl der von einem Beobachter registrierten aggressiver Handlungen in einem bestimmten Zeitraum, Anzahl und Schwere begangener Gewaltstraftaten, Anzahl aggressiver Inhalte im Rorschach-Test, Ausprägung in der FPI-R-Skala spontane Aggressivität, Einschätzung der Aggressivität auf einer Ratingskala durch den Klassenlehrer usw.

Die Operationalisierung führt zwangsläufig zu einer Reduktion des Gehalts der theoretischen Konstrukte. Die Operationalisierung hat Auswirkungen auf den Versuchsplan sowie umgekehrt der Versuchsplan die Operationalisierung beeinflusst.

Kriterien für angemessene Operationalisierung:

- Operationalisierung aller in der Forschungshypothese enthaltenen theoretischen Konstrukte
- Übertragung der Relation zwischen den Konstrukten
- Auswahl möglichst valider Indikatoren

Die Operationalisierung erfolgt unter Bezugnahme auf das **Hintergrundwissen** (bekannte Gesetzmäßigkeiten, technischen Voraussetzungen, weitere Vorannahmen, experimentelle Techniken, empirische Resultate usw.) Dieses Wissen wird ungeprüft als richtig vorausgesetzt. Somit gibt es neben den expliziten Hypothesen, die geprüft werden, viele implizite Hypothesen, die nicht überprüft werden.

Für ein theoretisches Konstrukt sind verschiedene Operationalisierungen möglich.
--

3.3 Logik der Signifikanzprüfung

Grundlage: Zentraler Grenzwertsatz

Unter der Annahme die Null-Hypothese sei richtig, läßt sich die Verteilung der Stichprobenkennwerte bestimmen, die angibt, mit welcher Wahrscheinlichkeit bestimmte Stichprobenergebnisse erwartet werden.

Mit dieser Verteilung wird das konkrete Stichprobenresultat verglichen (Signifikanztest).

Der Signifikanztest sagt nichts über die Hypothesen aus, sondern nur etwas über die Wahrscheinlichkeit von statistischen Kennwerten bei Gültigkeit der Nullhypothese

4. Von der Signifikanzprüfung zur Bewertung der Forschungshypothese

4.1 Entscheidung über die Testhypothese

Signifikanztests ermöglicht Abschätzung der Fehlerrisiken bei Entscheidung über Annahme oder Ablehnung der H_0

Wahrscheinlichkeit für richtige und falsche Entscheidung beim Signifikanztest

Entscheidung für	Wahrer, aber unbekannter Sachverhalt	
	H_0 gilt	H_1 gilt
H_0	$1-\alpha$	β -Fehler
H_1	α -Fehler	$1-\beta$ (Teststärke)

α = Wahrscheinlichkeit einer falschen Entscheidung gegen H_0
(Fehler der Übereifrigkeit)

β = Wahrscheinlichkeit einer falschen Entscheidung für H_0
(Fehler der Blindheit)

4.2 Simultankontrolle der Determinanten des Signifikanztests

Festlegung des α -Fehler-Niveaus (per Konvention $\alpha \leq .05$, $\alpha \leq .01$),
bei ungerichteten Hypothesen -> zweiseitige Signifikanzprüfung
bei gerichteten Hypothesen -> einseitige Signifikanzprüfung (erhöht die Teststärke)

Schätzung der β -Fehlerwahrscheinlichkeit: Setzt voraus, daß Populationsverhältnisse näher spezifiziert werden und die Effektgröße bestimmt werden kann. In den meisten Fällen ist dies jedoch nicht möglich.

Dann sollte der Mindestbetrag der **Effektgröße** hypothetisch festgelegt werden, damit der β -Fehler für diese hypothetische Effektgröße bestimmbar wird. **Kriterium für das Festlegen einer Mindesteffektgröße ist die praktische Bedeutsamkeit der Effektgröße.** (Wenn die Stichprobe groß genug ist können auch minimale, praktisch völlig irrelevante Unterschiede statistisch signifikant werden).

Festlegung des β -Fehler-Niveaus: Je nach Wichtigkeit der $H_1 \leq .20$ (Empfehlung von Cohen (1988). Bislang fehlt allerdings verbindliche Konvention) -> damit steht auch die **Teststärke** = $1-\beta$ fest (Wahrscheinlichkeit einen tatsächlich bestehenden Unterschied zu entdecken).

→ Ermittlung des **optimalen Stichprobenumfangs** (Cohen, 1988)

Unreflektierte Festlegung des Stichprobenumfangs ->
 β -Fehler bleibt unkontrolliert. Meist haben nur große Effekte eine realistische Chance signifikant zu werden.

Veränderung des optimalen Stichprobenumfangs in Abhängigkeit von den anderen Determinanten:

benötigte Stichprobenumfang ist kleiner \Downarrow , wenn Effekt größer wird \Uparrow ,
Stichprobenumfang \Uparrow , wenn α -Fehler-Niveau \Downarrow ,
Stichprobenumfang \Uparrow , wenn Teststärke \Uparrow

Zur Berechnung der Power gibt es das Computer Programm "G-Power" von Erdfelder, E., Faul, F., & Buchner, A. (Universität Trier), das ihr Euch kostenlos unter folgender Internetadresse besorgen könnt: <http://www.psych.uni-duesseldorf.de/abteilungen/aap/gpower3/download-and-register>
Die Autoren haben auch ein Internet-Manual dazu verfasst, das ihr unter der gleichen Adresse findet.

Den Daten werden bestimmte Eigenschaften unterstellt = **Voraussetzungen, die bei dem jeweiligen Signifikanztest erfüllt sein müssen**

(z.B. t-Test, Intervallskalenniveau, Normalverteilung, Varianzhomogenität).

Überprüfung, ob das mit dem Signifikanztest implizit unterstellte Datenmodell tatsächlich den Daten entspricht. -> Nur unter dieser Bedingung sind die Fehlerrisiken abschätzbar.

Bortz (1993) gibt für die jeweiligen Signifikanztests an, gegenüber welchen Verletzungen der Voraussetzungen sie robust sind und welche alternativen Verfahren dann gerechnet werden sollten (siehe auch aktuellen Ausgabe von Bortz & Schulz, 2010; Eid, Gollwitzer & Schmitt, 2010 oder amerikanische Statistiklehrbücher).

Probleme: Unzureichendes Skalenniveau, Voraussetzungen sind z.T. nicht überprüfbar (Normalverteilung in der Population, Additivität der Effekte)

4.3 Entscheidung über die Statistische Vorhersage

Statistische Vorhersage
 Fehlerrisiko abhängig davon ob SV H_0 oder H_1 entspricht,
 welches Prüfverfahren gewählt wurde,
 wie viele Tests gerechnet wurden

Wahrscheinlichkeit für richtige und falsche Entscheidung über die Gültigkeit der Statistischen Vorhersage auf der Grundlage des Signifikanztests:

Wenn SV = H_1 : führt ein signifikantes Ergebnis zur Annahme der SV. (Z.B. Es besteht ein Unterschied zwischen Mittelwerten der Treatment- und Kontrollgruppe bei der Postmessung)

Entscheidung:	Wahrer, aber unbekannter Sachverhalt	
	SV gilt nicht	SV gilt
gegen SV	$1-\varepsilon$	φ (entspricht β)
für SV	ε (entspricht α)	$1-\varphi$

ε = Wahrscheinlichkeit einer **falschen Entscheidung für SV** (Übereifrigkeit)

φ = Wahrscheinlichkeit einer **falschen Entscheidung gegen SV**
(Blindheit)

Wenn $SV = H_0$, führt ein signifikantes Ergebnis zur Ablehnung der SV.
(Beispiel: $SV = H_0$: Es besteht kein Unterschied zwischen Mittelwerten der Post- und Follow-up Messung in d. Treatmentgruppe)

Entscheidung:	Wahrer, aber unbekannter Sachverhalt	
	SV gilt nicht	SV gilt
gegen SV	$1-\varepsilon$	φ (entspricht α)
für SV	ε (entspricht β)	$1-\varphi$

ε = Wahrscheinlichkeit einer **falschen Entscheidung für SV**
(Übereifrigkeit)

φ = Wahrscheinlichkeit einer **falschen Entscheidung gegen SV**
(Blindheit)

Zusammenfassung:
 Wenn $SV = H_1$: $\varepsilon = \alpha$ -Fehler und $\varphi = \beta$ -Fehler,
 d.h. ein signifikantes Ergebnis führt zur Annahme der SV.
 Wenn $SV = H_0$: $\varepsilon = \beta$ -Fehler und $\varphi = \alpha$ -Fehler
 d.h. ein signifikantes Ergebnis führt zur Ablehnung der SV.

Sind mehrere Signifikanztests nötig, um SV zu prüfen, werden die Beziehungen von ε und φ zu α und β komplexer.

Problem: Kumulation des α -Fehlers, wenn mehrere Tests durchgeführt werden, macht α -Fehler-Adjustierung (z.B. Bonferoni-Korrektur) erforderlich. Dies führt jedoch zu einer Verringerung der Teststärke (vgl. Bortz, 1993, S. 248f).

Eine Statistische Hypothese wird nicht logisch falsifiziert, sondern der/die Forscher/in trifft eine **Entscheidung** über deren Gültigkeit oder Ungültigkeit

4.4 Bewertung der Operationalen Hypothese

Bei der Überprüfung einer inhaltlichen Hypothese fließen eine ganze Reihe subjektiver Entscheidungen und Beurteilungen des Forschers/der Forscherin ein, deshalb ist es angemessener von einer **Bewertung bzw. Beurteilung** der Forschungshypothese zu sprechen

In die Bewertung der Operationalen Hypothese fließen ein:

- Bewertung der Entscheidung über die Gültigkeit der Statistischen Hypothese,
- Bewertung der internen und statistischen Validität

Bewertungskriterien, die simultan Signifikanzniveau und Effektgröße berücksichtigen

Bewertung der OH	Signifikanz	geplante Effektgröße
bewährt	*	erreicht
bedingt bewährt	*	unterschritten
bedingt nicht bewährt	n.s	erreicht
nicht bewährt	n.s	unterschritten

* $\alpha \leq .05$

4.5 Bewertung der Forschungshypothese

Bewertung der Operationalen Hypothese, sowie **Bewertung aller Validitätskriterien** zusammen:

1. Interne Validität

2. Validität der Operationalisierung

- Variablenvalidität (für u.V. und a.V.)
- Ableitungsvalidität der OH .

3. Statistische Validität

- Ableitungsvalidität der statistischen Vorhersage (Übertragung der relevanten Information und Relationen der OH auf SV),
- Ableitungsvalidität der Testhypothesen
- angemessenes Testverfahren
- Voraussetzungen

4. Externe Validität

Die Forschungshypothese kann streng genommen nicht einmal falsi-
--

fiziert werden. Denn wenn die erwarteten Folgen nicht eintreten, können dafür immer die Hilfsannahmen verantwortlich gemacht werden.

Zusammenfassung

Kriterien für das Erstellen von Hypothesen

1. H. nach gründlicher Vorarbeit erstellen
2. H. sollten theoretisch und empirisch gut fundiert sein
3. H. so präzise wie möglich formulieren
(gerichtet, Schätzung der Effektstärke)
4. Nur solche H. sollten geprüft werden (Wett-Kriterium)
5. H. vor der Datenerhebung festlegen (H. in den Panzerschrank)

Wichtige Aspekte bei der Transformation und Interpretation von Hypothesen

- Adäquate Zuordnung wenn eine Übertragung der Relationen von FH -> OH -> SH gelingt
- Überprüfung einer FH durch Testung einer statistischen Vorhersage ist indirekt.
- SH wird nicht logischen falsifiziert, sondern Forscher trifft Entscheidung über Gültigkeit oder Ungültigkeit.
- Wichtig ist die simultane Kontrolle aller Determinanten des Signifikanz- tests und die daraus folgende Festlegung des Stichprobenumfangs. Eine unreflektierte Festlegung des Stichprobenumfangs hat zur Konsequenz, daß β -Fehler unkontrolliert bleibt. Meist haben dann nur große Effekte eine realistische Chance signifikant zu werden.
- Die OH kann nicht isoliert, sondern nur zusammen mit dem Hintergrundwissen falsifiziert werden.

Literatur (Das Skript stützt sich vor allem auf die Fett markierten Quellen, die anderen Titel erweitern und vertiefen das Thema)

Bortz, J. (1993). *Statistik für Sozialwissenschaftler (4. überarb. Aufl.)*. Berlin: Springer. (Einleitung, Kap. 4).

Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation (2., vollst. überarb. Aufl.)*. Berlin: Springer. (Kap 1, Kap. 2.3.5, 2.3.6, 2.3.8, Kap 8, Kap. 9.1, 9.2)

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2d ed.)*. Hillsdale, N.J.: Erlbaum.

Eid, M., Gollwitzer, M. & Schmitt, M. (2010). *Statistik und Forschungsmethoden*. Weinheim: Beltz.

Erdfelder, E. & Bredenkamp, J. (1994). Hypothesenprüfung. In T. Hermann & W. H. Tacks (Hrsg.), *Enzyklopädie der Psychologie, Forschungsmethoden der Psychologie, Bd. 1* (S. 605-648). Göttingen: Hogrefe.

Hager, W. (2000). Planung von Untersuchungen zur Prüfung von Wirksamkeits- und Wirksamkeitsunterschiedshypothesen. In W. Hager, J.-L. Patry & H. Brezing (Hrsg.), *Evaluation psychologischer Interventionsmaßnahmen. Standards und Kriterien: Ein Handbuch* (S. 202-239). Bern: Huber.

Huber, O. (1987). *Das Psychologische Experiment. Eine Einführung*. Bern: Huber.

Hussy, W. & Möller, H. (1994). Hypothesen. In T. Hermann & W. H. Tacks (Hrsg.), *Enzyklopädie der Psychologie, Forschungsmethoden der Psychologie, Bd. 1* (S. 475-507). Göttingen: Hogrefe.

Kontrollfragen

1. Welche Kriterien muss eine wissenschaftliche Hypothese erfüllen?
2. Welche Typen von Hypothesen sind nicht falsifizierbar, welche sind nicht verifizierbar?
3. Welche Möglichkeiten gibt es Hypothesen zu generieren?
4. Welche Voraussetzungen müssen erfüllt sein, damit eine Hypothese überprüft werden kann?
5. Operationalisierung
 - a) Warum müssen Forschungshypothesen operationalisiert werden?
 - b) Was sind die Kriterien für eine gelungene Operationalisierung?
 - c) Welche Probleme ergeben sich im Zusammenhang mit der Operationalisierung
6. Signifikanztest
 - a) Welche Faktoren beeinflussen die Signifikanzprüfung
 - b) Warum sollten diese Faktoren bereits in der Planungsphase einer Untersuchung festgelegt werden?
 - c) Nach welchen Kriterien sollten die Determinanten eines Signifikanztests festgelegt werden?
7. Warum sollte vor Anwendung eines bestimmten Signifikanztests überprüft werden, ob die Voraussetzungen für dessen Anwendung erfüllt sind?
8. Welches Problem ergibt sich, wenn mehrere Signifikanztests durchgeführt werden, um Hypothesen zu prüfen?
9. An welchen Stellen im Prozess der Überprüfung einer Hypothese trifft der/die Forscher/in Entscheidungen oder nimmt eine Bewertung vor?
10. Welche Aspekte sollte der/die Forscher/in beim letzten Interpretationsschritt – der Bewertung der Forschungshypothese - berücksichtigen?
11. Warum kann eine Forschungshypothese streng logisch betrachtet nicht einmal falsifiziert werden?