

## **Lösungen zu den Übungsaufgaben aus dem Lehrbuch „Empirische Methoden in der Psychologie“ von Markus Pospeschill**

### **Kapitel 1**

#### **1. Was zeichnet eine wissenschaftliche Methode aus?**

Eine wissenschaftliche Methode sollte kommunizierbar, normativ und intersubjektiv kontrollierbar sein.

#### **2. Welche Funktionen übernehmen Variablen?**

Es werden u.a. unterschieden: abhängige Variable, unabhängige Variable, Moderator, Mediator, Kontrollvariable/Kovariate, Residualvariable, latente Variable, manifeste Variable.

#### **3. Welche Kriterien sind an wissenschaftliche Hypothesen anzulegen?**

Wissenschaftliche Hypothesen sind im Hinblick auf ihren empirischen Gehalt, ihre Generalisierbarkeit und ihre potenzielle Falsifizierbarkeit zu bewerten.

#### **4. Welche Hypothesenarten gibt es?**

Es werden Punkt- vs. Bereichshypothesen, gerichtete vs. ungerichtete Hypothesen sowie wissenschaftliche und statistische Hypothesen unterschieden.

#### **5. Wie sieht ein Konditionalsatz aus?**

Ein Konditionalsatz besteht aus einem Bedingungs- bzw. Antezedenzteil und einem Folge- bzw. Konsequenzteil.

#### **6. Welche Funktion übernehmen Hypothesen in der Psychologie?**

Hypothesen besitzen induktive (hypothesenerkundende, explorative) und deduktive (hypothesenprüfende, inferentielle) Funktion.

#### **7. Wozu benötigt man statistische Hypothesen?**

Statistische Hypothesen dienen der Quantifizierung intendierter Aussagen, haben Bezug auf eine Grundgesamtheit und sind vom Charakter her probabilistische

Aussagen mit einer vordefinierten Schwankungsbreite.

**8. Sind Wahrscheinlichkeitsmodelle durch Einzelfälle falsifizierbar/verifizierbar?**

Nein, Stichprobenaussagen sind nicht auf den Einzelfall übertragbar, da der Messfehler für den Einzelfall nicht bekannt ist.

**9. Wie löst man das Problem der Falsifizierbarkeit?**

Das Problem der Falsifizierbarkeit wird mit dem Konzept der statistischen Signifikanz gelöst.

**10. Zu welchen Aussagen ist man berechtigt, wenn es nicht gelingt, die Nullhypothese abzulehnen?**

Kann die Nullhypothese nicht zurückgewiesen werden, bedeutet dies streng genommen nur, dass sie nicht abgelehnt werden kann. Damit ist sie aber nicht bestätigt.

**11. Woran hängt die Interpretation einer Beziehung zwischen Variablen als Kausalbeziehung?**

Kausale Interpretationen sind abhängig vom Untersuchungsdesign (z.B. Experiment), inhaltlichen Erwägungen und der Voraussetzung, dass kein sinnvoller Austausch zwischen Wenn- und Dann-Teil möglich ist.

**12. Welchen Kriterien müssen Begriffe in den empirischen Wissenschaften genügen?**

Begriffe müssen Bezug zur erfahrbaren Realität aufweisen, Aussagen müssen einen empirischen Bezug haben und prinzipiell widerlegbar sein.

**13. Welchem Hauptprinzip folgt der kritische Rationalismus?**

Begriffe und Aussagen sind an der Erfahrung überprüfbar und können daran scheitern.

**14. Was zeichnet eine sog. Allaussage/nomologische Aussagen aus?**

Allaussagen besitzen (im Gegensatz zu Aussagen mittlerer Reichweite) keine raumzeitliche Einschränkung.

**15. Welche Aspekte werden beim Werturteilsstreit unterschieden?**

Unterschieden werden der Entdeckungs-, der Begründungs- und der Verwertungszusammenhang.

**16. Welchen Charakter besitzen deduktiv-nomologische Erklärungen in den Humanwissenschaften?**

Sie besitzen den Charakter empirisch abgesicherter Theorien.

**17. Was ist eine Theorie?**

Eine Theorie ist ein erklärendes Netzwerk bezüglich der relativen Bedeutung von UVs und AVs mit der Bewährung an realen Beobachtungen. Ferner ist eine Theorie ein System widerspruchsfreier Aussagen zu einem Gegenstandsbereich oder auch eine Gesetzmäßigkeit mit Axiomen (Postulate, die nicht aus anderen Sätzen abgeleitet werden können).

**18. Wie kann man die Allgemeingültigkeit einer Theorie überprüfen?**

Zur vorläufigen Bestätigung von Theorien werden systematische Replikationen herangezogen.

**19. Was kennzeichnet das Korrespondenzproblem bei falsifizierenden Untersuchungsergebnissen?**

Das Korrespondenzproblem entsteht durch eine unangemessene Operationalisierung der Konstrukte, durch mangelnde Validität und Reliabilität.

**20. Wie „rettet“ man Theorien?**

Theorien können durch Exhaustion gerettet werden, z.B. durch Hinzunehmen zusätzlicher Und-Komponenten.

## **Kapitel 2**

### **1. Wie wird mit dem Problem der Dimensionalität umgegangen?**

Beim Problem der Dimensionalität gelten das Parsimonitätsprinzip (einfachere Konzepte sind günstiger als komplexe) und das Begründungsprinzip (jede Messung ist zu begründen).

### **2. Welche Funktion besitzt eine dimensionale Analyse?**

Abgeleitet aus dem Erkenntnis- und Verwertungsinteresse arbeitet die dimensionale Analyse diejenigen Einzelheiten der Realität heraus, die untersucht werden sollen.

### **3. Welche Funktion besitzt eine semantische Analyse?**

Die semantische Analyse legt die Bedeutung von Begriffen fest.

### **4. Wozu dienen Korrespondenzregeln?**

Es ist zu klären, ob der Sachverhalt (definiert durch den theoretischen Begriff) auch tatsächlich vorliegt.

### **5. Wie lassen sich Begriffe definieren?**

Begriffe können nominal (als rein begriffliche Zuordnung) oder real (als Wesensbeschreibung) definiert werden.

### **6. Warum müssen Begriffe im Kontext wissenschaftlicher Verwendung analytisch und operational definiert werden?**

Die Definition erfolgt analytisch, um Semantik und Verwendung zu explizieren, und operational, um die Messung über explizite Indikatoren zu realisieren.

### **7. Welche Arten von Indikatoren gibt es?**

Unterschieden werden definatorische, (intern und extern) korrelative sowie schlussfolgernde Indikatoren.

### **8. Warum verwendet man häufig multiple Indikatoren?**

Multiple Indikatoren dienen der Steigerung der Validität einer Messung.

**9. Warum werden Messungen aggregiert?**

Die Aggregation von Messungen dient der Ausmittelung von Zufallsfehlern und der Erhöhung der Messgenauigkeit.

**10. Was ist das Ziel der Operationalisierung?**

Ziel der Operationalisierung ist die Angabe einer Messvorschrift und des Messniveaus.

**11. Welche Varianten von Operationalisierungen gibt es?**

Varianten von Operationalisierungen sind Häufigkeit, Reaktionszeit, Reaktionsdauer, Reaktionsstärke und Wahlreaktionen.

**12. Was unterscheidet logische von empirischer Validität?**

Die logische Validität bezieht sich auf den Bedeutungsumfang von Begriff, Indikator und Operationalisierung und die empirische Validität fragt danach, was dabei konkret erfasst wird (bzw. ob auch im Sinne der logischen Validität geantwortet wird).

**13. Was bedeutet Messen in der Psychologie?**

Messen in der Psychologie ist die Zuordnung von Zahlen zu Objekten unter Erhalt der Relationen. Dabei sollen die Relationen zwischen den Objekten (empirisches Relativ) mathematisch korrekt abgebildet werden (numerisches Relativ); diese Abbildung muss strukturverträglich sein. Bei Erfüllung der Bedingungen entsteht eine Skala.

**14. Welche Problembereiche werden beim Messen unterschieden?**

Das Repräsentationsproblem bezieht sich auf die homomorphe Abbildung von empirischem und numerischem Relativ, das Eindeutigkeitsproblem entscheidet über zulässige Transformationsregeln der resultierenden Skalen und das Bedeutsamkeitsproblem fragt nach der sinnvollen Anwendung mathematischer Operationen.

**15. Welches Problem spielt vor allem bei der Wahl des Skalenniveaus eine entscheidende Rolle?**

Bei der Wahl des Skalenniveaus ist vor allem das Bedeutsamkeitsproblem relevant, da es über sinnvolle Interpretationen statistischer Kennwerte entscheidet.

**16. Was unterscheidet eine explorative von einer populationsbeschreibenden und hypothesenprüfenden Untersuchung?**

Explorative Untersuchungen dienen der Hypothesenerkundung, populationsbeschreibende Untersuchungen der Schätzung von Populationsparametern und hypothesenprüfende Untersuchungen der Entscheidung über Hypothesen.

**17. Welche Untersuchungsart besitzt die höchste interne Validität, welche die höchste externe Validität?**

Höchste interne Validität besitzt das Laborexperiment, höchste externe Validität das Feldexperiment.

**18. Was unterscheidet Experimente von Quasi-Experimenten?**

Quasi-Experimente besitzen eine eingeschränkte interne Validität, da die Personen den Gruppen nicht zufällig (randomisiert) zugewiesen werden können.

**19. Welche Rolle spielt die Stichprobenwahl bei explorativen Studien?**

Bei explorativen Studien kommen nur Kollektive ohne spezifische probabilistische Ziehungstechnik zum Einsatz.

## **Kapitel 3**

### **1. Was versteht man unter Evaluationsforschung?**

Evaluationsforschung ist eine bewertende Begleitforschung, gekoppelt mit der Erfolgs- und Wirkungskontrolle von Maßnahmen unter Gewährleistung möglichst hoher interner Validität.

### **2. Was unterscheidet technologische von wissenschaftlichen Theorien?**

Technologische Theorien dienen der Gewinnung praktischen Wissens, während wissenschaftliche Theorien ein stringentes Gefüge über Ursache und Wirkung darstellen. Bei grundlagenwissenschaftlichen Theorien ist der direkte praktische Nutzen nachrangig.

### **3. Was ist einer Evaluationsforschung vorgeschaltet?**

Der Evaluationsforschung ist eine Interventionsforschung vorgeschaltet mit dem Ziel, eine Maßnahme oder eine spezifische Intervention zu entwickeln.

### **4. Was unterscheidet summative von formativer Evaluation?**

Bei einer summativen Evaluation ist die Maßnahme bereits abgeschlossen, während bei einer formativen Evaluation die Maßnahme begleitet wird.

### **5. Was unterscheidet bei formativen Evaluationen die Prävalenz- von der Inzidenzrate?**

Die Prävalenzrate entscheidet über das Ausmaß (z.B. der Betroffenen) und die Inzidenzrate über die zeitlich-räumliche Verteilung (z.B. betroffene Bereiche) einer Maßnahme.

### **6. Um welche Untersuchungsart handelt es sich bei Evaluationsstudien?**

Evaluationsstudien sind quasiexperimentelle Feldstudien.

### **7. Was versteht man unter der Ausschöpfungsqualität?**

Die Ausschöpfungsqualität bezeichnet den Anteil betroffener vs. nichtbetroffener Teilnehmer an der Zielpopulation.

## **Kapitel 4**

### **1. Was ist beim Auszählen/Bilden von Kategorien zu beachten?**

Beim Auszählen bzw. Bilden von Kategorien ist das Genauigkeitskriterium (präzise Indikatoren), das Exklusivitätskriterium (nur eine Kategorie pro Objekt) und das Exhaustivitätskriterium (alle Objekte werden zugeordnet) zu beachten.

### **2. Was ist ein Index?**

Ein Index ist ein errechnetes Maß aus Messwerten verschiedener Indikatorvariablen (im Gegensatz zu einem Skalenwert, der eine Zusammenfassung über homogene Items darstellt).

### **3. Wie lassen sich Indizes bilden und was unterscheidet sie?**

Indizes können ungewichtet (additiv, mit möglicher Kompensation), multiplikativ (ohne Kompensation) und gewichtet additiv (mit Gewichtungsfaktoren) gebildet werden.

### **4. Wie können Gewichtungen von Indizes entstehen?**

Gewichtungen von Indizes entstehen normativ (z.B. über Expertenurteile) oder statistisch (z.B. über Faktorladungen oder Regressionsgewichte).

### **5. Was unterscheidet eine normative von einer analytischen Indexbildung?**

Eine normative Indexbildung erfolgt über Expertenratings, während eine analytische Indexbildung über eine Statistik (z.B. Faktorenanalyse) erfolgt.

### **6. Nach welchen Methoden kommt man zu ranggeordneten Daten?**

Ranggeordnete Daten können entstehen durch die Methode der sukzessiven Intervalle / Q-Sort-Verfahren, durch Dominanz-Paarvergleich oder Ähnlichkeits-Paarvergleich.

### **7. Was sind die Probleme bei Paarvergleichen?**

Probleme bei Paarvergleichen können durch zirkuläre Triaden oder intransitive Urteile entstehen.



### **8. Wie können ordinale Daten in Intervalldaten umgerechnet werden?**

Mittels des *Law of Categorical and Comparative Judgement* von Thurstone werden ordinale Daten (1) in relative Häufigkeiten umgerechnet, (2) in z-Werte transformiert, (3) zu Summen- und Mittelwerten verrechnet, um schließlich (4) Werte für Empfindungsstärken zu erhalten.

### **9. Welche Unterschiede gibt es bei Ratingskalen?**

Es gibt ein- vs. mehrdimensionale, grafische vs. numerische sowie bi- vs. unipolare Ratingskalen.

### **10. Was ist die optimale Anzahl von Kategorien bei Ratingskalen?**

„Optimal“ für eine Ratingskala sind 7 Kategorien, ggf. mit Ausweichkategorie.

### **11. Wann sind forcierte vs. nicht-forcierte balancierten vs. unbalancierten Skalierungen vorzuziehen?**

Mit einem forcierten Rating wird eine Einschätzungstendenz erzwungen, bei nicht-forcierten Ratings bleibt eine Rückzugsposition in der Mitte, bei einer balancierten Skala sind die Bewertungsrichtungen ausgeglichen, bei einer unbalancierten Skala ist eine der Bewertungsrichtungen differenzierter.

### **12. Welches Skalenniveau besitzen Ratingskalen?**

Ratingskalen besitzen kein vordefiniertes Skalenniveau. Auch kann das Skalenniveau nicht direkt bestimmt werden (Einzelreizmethode).

### **13. Wann wirkt eine Verletzung des Skalenniveaus gravierend?**

Die Verletzung des Skalenniveaus wirkt dann gravierend, wenn keine sinnvolle Interpretierbarkeit einer Messung anhand des angenommenen Skalenniveaus möglich ist (z.B. Verletzung der Äquidistanz, wie sie für Intervallskalen angenommen wird).

### **14. Welche Effekte können bei der Beantwortung von Ratingskalen entstehen?**

Typische Fehlerquellen sind u.a. der Halo-Effekt, der Milde-Effekt, der

Normanpassungs-Effekt, die Rater-Ratee-Interaktion sowie der Primacy-Recency-Effekt.

### **15. Was kennzeichnet eine Likert-Skala?**

Eine Likert-Skala entsteht durch ein mehrstufiges Verfahren, bei dem aus einem Pool von Statements ungeeignete Items mittels Itemanalyse eliminiert, aus den verbleibenden Items eine endgültige Skala zusammengestellt und aus der Addition der Einzelurteile eine Gesamtpunktzahl pro Proband ermittelt wird.

### **16. Welche Probleme können im Zuge der Multidimensionalen Skalierung entstehen?**

Typische Probleme bei der Multidimensionalen Skalierung sind lokale vs. globale Minima, degenerierte Objektkonfigurationen, Stresswerte nahe Null, das Dimensionsproblem sowie das Interpretationsproblem.

### **17. Was ist vor Anwendung einer Multidimensionalen Skalierung zu beachten?**

Da es sich um ein Verfahren mit einer Datenverdichtung handelt, muss die Anzahl der Eingabedaten größer sein als die der Ausgabedaten (9 oder mehr Objekte für 2, 11 bei 3 Dimensionen).

### **18. Es soll eine zweidimensionale Skala entstehen, die später faktorenanalytisch ausgewertet werden soll. Welche Kriterien sind bei der Skalenkonstruktion zu beachten?**

Für eine zweidimensionale Skala sollten mindestens 5 Items pro Skala vorliegen und die Normalverteilung des Merkmals bzw. der Skala sollte gegeben sein.

## **Kapitel 5**

### **1. Mit welchem Ziel wird eine Exploration betrieben?**

Die Exploration dient als Vortest oder Vorstudie zur Klärung inhaltlicher und methodischer Fragen und als Instrumententest zur Prüfung der Funktionstüchtigkeit technischer Apparaturen und verwendeter Untersuchungsmaterialien.

### **2. Was unterscheidet eine theorie- von einer methodenbasierten Exploration?**

Bei der theoriebasierten Exploration geht es um Vergleiche, Integrationsversuche, Formalisierungen, Modellbildungen sowie die Bildung von Metatheorien, während bei der methodenbasierten Exploration Methodenvergleiche, Methodenvariationen und die Anwendung von Methoden als Denkwerkzeuge im Vordergrund stehen.

### **3. Welche Techniken und Verfahren werden bei Exploration quantitativer Daten herangezogen?**

Die Exploration quantitativer Daten erfolgt über die grafische Aufbereitung von Daten, eine explorative Datenanalyse oder explorative Verfahren wie die Cluster- oder Faktorenanalyse.

### **4. Was versteht man unter explorativem Signifikanztesten?**

Als exploratives Signifikanztesten bezeichnet man durchgeführte Signifikanztests zum Zwecke einer ersten Eindrucksbildung (Größenordnung oder Ausmaß) von Effekten, ohne dass diese damit bestätigt werden.

### **5. Wann wird die Exploration qualitativer Daten als „nonreaktiv“ bezeichnet?**

„Nonreaktiv“ ist eine Datenerhebung dann, wenn diese von den Probanden entweder nicht bemerkt oder zuvor ohne eine erkennbare Forschungsabsicht vorgenommen wurde.

## **Kapitel 6**

### **1. Was sind Voraussetzungen für die Ziehung großer Stichproben?**

Für die Ziehung großer Stichproben müssen konstante Ziehungskriterien und eine konstante Größe gegeben sein. Große Stichproben sollten zudem nicht der Kompensation mangelnder Reliabilität des Messinstrumentes dienen.

### **2. Wann sind große Stichproben indiziert?**

Große Stichproben sind anzuraten bei heterogenen Merkmalen, bei (erwarteten) kleinen Effekten, bei einer differenzierten Merkmalsbetrachtung (Moderatoren), bei angestrebter Repräsentativität und bei Verwendung verteilungsgebundener Verfahren.

### **3. Wann sind kleine Stichproben ausreichend?**

Kleine Stichproben sind ausreichend bei homogenen Merkmalen, bei (erwarteten) großen Effekten, bei einheitlichen Merkmalen, bei Untersuchung definierter Teilstichproben und bei Verwendung verteilungsfreier Tests oder Randomisierungstests.

### **4. Was unterscheidet globale von lokaler Repräsentativität?**

Globale Repräsentativität berücksichtigt alle Merkmale einer Grundgesamtheit, lokale Repräsentativität nur spezifische Merkmale der Grundgesamtheit.

### **5. Wie entsteht eine geschichtete Stichprobe?**

Eine geschichtete Stichprobe resultiert aus Subpopulationen, die durch Schichtungsmerkmale definiert werden. Aus jeder Schicht wird dann eine Zufallsstichprobe gezogen. Dazu muss die Größe der Subpopulationen bekannt sein. Die Schichten sollten möglichst homogen, aber untereinander bezüglich des Merkmals unterschiedlich sein.

### **6. Wie entsteht eine Klumpenstichprobe?**

Bei einer Klumpenstichprobe werden Teilpopulationen vollständig erhoben. Dazu müssen die Klumpen bekannt und gleich groß sein. Die entstehenden Klumpen

sollten dabei möglichst ähnlich, die Objekte bezüglich des Merkmals aber unterschiedlich sein.

## Kapitel 7

### 1. Warum kann ein statistischer Signifikanztest keine Aussagen über die Wichtigkeit eines Ergebnisses machen?

Ein statistischer Signifikanztest schätzt nur die Wahrscheinlichkeit für ein beobachtetes Stichprobenresultat unter der Annahme einer exakten Nullabweichung in der Population. Ob das Stichprobenergebnis die Population auch wirklich beschreibt, wird nicht bewertet. Auch erlaubt der Test keine Aussage darüber, ob sich das Ergebnis bei weiteren Stichproben wieder einstellt.

### 2. Warum sollten im Idealfall $\alpha$ - und $\beta$ -Fehler festgelegt werden?

Nur bei Festlegung von  $\alpha$ - und  $\beta$ -Fehler kann entschieden werden, welche der möglichen Fehlentscheidungen (zugunsten der Null- oder der Alternativhypothese) verhängnisvollere Konsequenzen besitzt.

### 3. Was unterscheidet ein *Cross-sectional*- von einem *Cross-lagged-Design*?

Ein *Cross-Sectional-Design* ist ein Querschnitts-Design (mit einem Messzeitpunkt), während ein *Cross-Lagged-Design* ein längsschnittliches Design darstellt (mit zwei oder mehr Messzeitpunkten).

### 4. Was unterscheidet eine Dummy- von einer Effekt- sowie einer Kontrastkodierung?

Dummykodierung (0, 1), Effektkodierung (0, 1, -1) und Kontrastkodierung (0, 1, -1, 0,5, -0,5) dienen der Transformation von Prädiktoren in Indikatorvariablen im Kontext einer multiplen Regression.

### 5. Warum lassen sich Zusammenhänge nicht kausal interpretieren?

Für Zusammenhänge lassen sich theoretisch unendlich viele Kausalmodelle konstruieren und daher nicht bestätigen.

### 6. Was charakterisiert einen randomisierten Zwei-Gruppen-Plan?

Randomisierte Zwei-Gruppen-Pläne dienen als Grundlage echter Experimente und zeichnen sich durch eine Treatment- und eine Kontrollgruppe aus, die bezüglich der

Veränderung in der AV verglichen werden. Beide Gruppen entstehen durch Zufallszuweisung aus einer Zufallsstichprobe, um Störvariablen zu kontrollieren.

### **7. Welche Arten von Interaktionen gibt es bei Mehr-Gruppen-Plänen?**

Interaktionen bei Mehr-Gruppen-Plänen können sich ordinal, hybrid oder disordinal darstellen.

### **8. Welche zentralen Größen werden bei der Untersuchung von Veränderungshypothesen unterschieden?**

Zentrale Größen bei Veränderungshypothesen sind Zeiteffekte, epochale Effekte und Generations- bzw. Kohorteneffekte.

### **9. Was unterscheidet indirekte von direkten Veränderungsmessungen?**

Eine indirekte Veränderungsmessung betrachtet das Ausmaß an Veränderungen zwischen Messungen, während die direkte Veränderungsmessung den Unterschied einer Einmalerhebung betrachtet.

### **10. Warum ist die Reliabilität von Differenzwerten umso geringer, je höher die Reliabilität der Einzeltestungen?**

Steigt die Korrelation zwischen den Messungen, sinkt die Reliabilität der Differenz. Grund ist die Zunahme wahrer Varianz der zwei Testungen. Die Differenz spiegelt dann nur noch die Fehlervarianz wider.

### **11. Welche Annahmen sind gerechtfertigt, wenn keine spezifischen Effektgrößen ermittelt werden können?**

Die erwarteten Effekte sollten in diesem Fall als klein, mittel oder stark eingeschätzt werden. Es kann auf entsprechende Klassifikationstabellen zurückgegriffen werden, um den Stichprobenumfang festzulegen.

### **12. Was sind die Bestimmungsgrößen für ein optimales $n$ ?**

Bestimmungsgrößen für einen optimalen Stichprobenumfang sind der verwendete Test, der  $\alpha$ -Fehler, die Teststärke ( $1-\beta$ ), die Effektstärke und die Art der Testung.

## Kapitel 8

### 1. Welches sind die zentralen Validitätsprobleme bei klinischen Forschungsdesigns?

Die interne Validität (um alternative Erklärungen hinsichtlich der Wirkung eines Treatments auszuschalten, Übungseffekte zu kontrollieren) und die externe Validität (um das Ergebnis auf andere Fälle bzw. Grundgesamtheiten übertragen zu können). Ggf. spielt die ökologische Validität eine besondere Rolle.

### 2. Welche Mindeststichprobengrößen werden bei parametrischen, nicht-parametrischen und Randomisierungstests vorausgesetzt?

Mindestgrößen bei parametrischen Tests sind  $n \geq 30$ , bei nicht-parametrischen Tests  $30 \leq n \leq 15$  und Randomisierungstests  $n \leq 10$ ; alle Angaben gelten pro Treatmentbedingung.

### 3. Was verbirgt sich hinter der Bezeichnung „Randomisierungstests“?

Bei einem Randomisierungstest bestimmt sich der Zeitpunkt der Verabreichung des Treatments per Zufall. Ferner wird die Prüfstatistik aus einer Zufallszuweisung der Messwerte auf die Bedingungen „ohne Treatment“ und „mit Treatment“ abgeleitet (bedingter Test).

### 4. Wann wirken Voraussetzungsverletzungen für Varianz- und Regressionsanalysen kritisch?

Voraussetzungsverletzungen für Varianz- und Regressionsanalysen sind dann kritisch, wenn multiple Verletzungen (kleines  $n$ , keine Normalverteilung, keine Varianzhomogenität, etc.) vorliegen und/oder die Daten autokorreliert sind.

### 5. Was unterscheidet die analytische von der computationalen Power?

Analytische Power ist bei klassischen parametrischen Tests erforderlich, wenn auf tabellierte (hypothetische) Prüfverteilungen zurückgegriffen wird; dabei müssen die empirischen Daten den Annahmen der Prüfverteilungen entsprechen.

Computationale Power ist erforderlich, wenn im Falle eines Randomisierungstests die Prüfverteilung aus den empirischen Daten selber erzeugt wird; die dadurch



entstehende Prüfverteilung ist dann für die empirischen Daten spezifisch.

### **6. Was versteht man unter Autokorrelationen und was ist der Effekt positiver Autokorrelationen?**

Als Autokorrelation wird die Abweichung einer Beobachtung vom Mittelwert durch Abweichungen vorangegangener Beobachtungen bezeichnet. Gegenmaßnahmen zum Effekt der Autokorrelation sind wechselnde Treatments, eine Anpassung des Signifikanzniveaus oder die Maximierung von Messintervallen.

### **7. Was ist die Problematik nicht-parametrischer Tests?**

Durch die Rangtransformation von Daten verlieren nicht-parametrische Tests an Teststärke. Dieser Effekt vergrößert sich bei verbundenen Rängen und sehr kleinen Stichproben (da die tabellierten kritischen Werte dann nur noch asymptotische Lösungen bieten).

### **8. Worin besteht die Problematik mangelnder Reliabilität von Einzelfällen?**

Nach der operanten Sichtweise wird die Replikation vorrangig zur Bestätigung des Treatmenteffekts angesehen. Problematisch ist hier allerdings, dass die Reliabilität notwendig aber nicht hinreichend für die interne Validität ist.