

Single-Case Expert

Eine Software für verteilungsfreie
Zeitreihenanalysen von Einzelfällen

Leon Reicherts

Schriften zur psychosozialen Gesundheit

Leon Reicherts

Single-Case Expert

Eine Software für verteilungsfreie
Zeitreihenanalysen von Einzelfällen



Impressum

CIP-Titelaufnahme der Deutschen Bibliothek

Leon Reicherts

Single-Case Expert

Eine Software für verteilungsfreie Zeitreihenanalysen von Einzelfällen.

Goßmannsdorf/Ochsenfurt: ZKS-Verlag

Alle Rechte vorbehalten

© 2017 ZKS-Verlag

Technische Redaktion: Meike Kappenstein

Lektorat: Tony Hofmann

Cover-Design: Leon Reicherts

ISBN 978-3-947502-00-4

Anschrift:

ZKS-Verlag / Verlag für psychosoziale Medien

Winterhäuser Str. 13

97199 Goßmannsdorf/Ochsenfurt

Kontakt:

info@zks-verlag.de

www.zks-verlag.de

Tel./Fax (09331) 2001509

Danksagung

An dieser Stelle möchte ich allen danken, die mich bei meiner Bachelorarbeit, die der vorliegenden Publikation zugrunde liegt, unterstützt und begleitet haben.

Ein großer Dank geht an Prof. Dr. Martin Huber für die ausgezeichnete Betreuung der Arbeit und für die wichtigen Ratschläge sowie den konstruktiven Austausch. Für die Hilfe bei der Analyse der statistischen Verfahren und ihrer Berechnungslogik möchte ich Dr. Christoph Leuenberger herzlich danken. Ein Dankeschön geht auch an die Doktorandinnen am Departement für Informatik Rédina Berkachy und Layal Pipoz für ihre technische Unterstützung bei der Entwicklung des Programmcodes. Weiter möchte ich Prof. Dr. Philippe Genoud für seine methodischen und statistischen Vorschläge danken. Ein spezieller Dank gilt auch meinem Vater, Prof. em. Dr. Michael Reicherts, für viele spannende und hilfreiche Diskussionen während des Studiums und speziell während der Realisierung dieses Projekts.

Abstract

Die Webapplikation *Single-Case Expert (SCE)* ist ein neuartiges methodisch-statistisches Hilfsmittel zur Durchführung und Auswertung von Einzelfallanalysen.

Einzelfallanalysen haben zum Ziel, Zustände oder Verhaltensweisen eines Individuums bzw. sozialen Aggregates auch in ihrer Abfolge zu beschreiben oder – wenn möglich – zu erklären. Neben dem Entwickeln und Erproben entsprechender Hypothesen (für die Bildung neuer Theorien) ermöglicht der im Kontext dieser Arbeit wichtige *quantitative Ansatz*, noch mehr als der *qualitative*, die Schaffung neuer *Evidenz*, sofern die jeweils angemessenen Ansätze und Methoden in der Planung, Durchführung und Interpretation der Untersuchungen angewendet werden. Diese Methoden werden zusammen mit Fallbeispielen aus verschiedenen Domänen ausführlich in dem neuen Sammelband über Einzelfallanalysen (Reicherts & Genoud, 2015) beschrieben. Die Webapplikation *SCE* ist als „Werkzeugkasten“ zu dieser Publikation zu verstehen, speziell zum 7. Kapitel, wo einige für die Einzelfallanalyse wichtige non-parametrische statistische Verfahren vorgestellt werden.

SCE schlägt ausgehend von den eingegebenen Daten passende Methoden vor und erlaubt eine benutzergeführte und weitgehend automatisierte Durchführung der im Buch beschriebenen Methoden. Die Verfahren dienen dazu, die Zeitreihe auf ihre *Zufälligkeit*, auf *zyklische Variation* und auf *Trends* in der *zentralen Tendenz* und der *Dispersion*, sowie auf *Niveau-Unterschiede* zu untersuchen. Davon ausgehend lassen sich Hypothesen *bilden* oder *unterstützen*, oder Interventionen können in ihrer Wirksamkeit untersucht werden.

Einzelfallanalysen, für Forschung und evidenzbasierte Praxis sehr wichtig, aber in beiden Feldern noch nicht ihrer Relevanz entsprechend verbreitet, sollen mit diesem Tool speziell für Praktiker erleichtert werden, da man gerade hier laufend mit Einzelfällen zu tun hat, diese aber oft nicht angemessen analysieren kann, weil die nötigen methodischen Wissensgrundlagen und Werkzeuge fehlen.

Der vorliegende Text ist eine angepasste und erweiterte Version der gleichnamigen Bachelorarbeit, die an der Universität Freiburg (Schweiz), Departement für Ökonomie, Lehrstuhl für angewandte Ökonometrie von Prof. Dr. Martin Huber betreut und im August 2016 angenommen wurde.



Release Version (v1.0)

www.single-case.expert

Inhalt

1	Einführung.....	7
1.1	Einzelfälle in Wissenschaft und Praxis.....	7
1.2	Wissensarten, Hypothesen und ihre Überprüfung.....	7
1.3	Formen und Designs der Einzelfallanalyse.....	9
1.4	Die Quantitative Einzelfallanalyse.....	9
1.5	Die Idee eines Expertensystems für die Analyse von Einzelfällen.....	10
2	Statistische Grundlagen.....	11
2.1	Deskription, Inferenz und Exploration.....	11
2.2	Skalenniveaus.....	11
2.3	Parametrische und non-parametrische Methoden.....	12
3	Statistische Verfahren in <i>Single-Case Expert</i>	13
3.1	Unterstützte Skalenniveaus.....	13
3.2	Deskriptiv-statistische Methoden.....	14
3.3	Inferenzstatistische Methoden.....	15
4	Das Tool <i>Single-Case Expert</i>	23
4.1	Programmierung von <i>Single-Case Expert</i>	23
4.2	Verwendung von <i>Single-Case Expert</i>	28
5	Exemplarische Auswertungen von Fallbeispielen.....	28
5.1	Fallbeispiel 1: Stress im Alltag.....	29
5.2	Fallbeispiel 2: Förderung der sozialen Kontaktaufnahme.....	30
5.3	Fallbeispiel 3: Frau G. – Intervention bei Angstsymptomen.....	32
5.4	Fallbeispiel 4: Beleuchtungskonzept für verbessertes Wohlbefinden.....	38
5.5	Fallbeispiel 5: Neue Suchmaschine für einen B2B-Onlineshop.....	43
5.6	Anmerkung zur Interpretation der Inferenzstatistiken.....	49
5.7	Aspekte des Wissenszuwachses.....	49
6	Zusammenfassung und Ausblick.....	50
7	Abbildungsverzeichnis.....	53
8	Literaturverzeichnis.....	54
9	Anhang: Extraktion der Hilfestellungen und Informationen.....	57

1 Einführung

1.1 Einzelfälle in Wissenschaft und Praxis

Es gibt berühmte Beispiele, in denen Einzelfälle zu wichtigen Hypothesen oder zu ihrer Revision geführt haben: Der Fall des H. M. (Milner, 1966; z. n. Haymoz, Ledermann & Martin-Soelch, 2015) der nach Ablation des Hippocampus schwere Störungen seines Erinnerungsvermögens zeigte; die *Vergessenskurve*, die Ebbinghaus (1885) im Selbstversuch entdeckte; die Revision der modernen genetischen Abstammungstheorie durch einen einzigen Schädelfund in einer Höhle in Israel (Hershkovitz et al., 2015): Danach war der „moderne Mensch“ aus Afrika (*Eva0* und *Adam0*) nicht der einzige Vorfahre, von dem wir abstammen, sondern der nördliche Neandertaler hat sich mit ihm gemischt, und unser Chromosomensatz enthält auch Elemente von ihm.

Die Wichtigkeit dieser Einzelfälle ist unbestritten, und dennoch hat die systematische Einzelfallanalyse in Wissenschaft und Praxis noch längst nicht den Stellenwert, der ihr eigentlich zukommen müsste.

Die Einzelfallanalyse bezieht sich auf eine einzelne Untersuchungseinheit (eine Person, eine Gruppe oder eine Gesellschaft), die als Ganzes betrachtet wird und deren Zustände und Verhaltensweisen im Verlauf analysiert werden. Die vorliegende Arbeit beschäftigt sich mit *quantitativen* bzw. *quantifizierenden* Vorgehensweisen zur Untersuchung derartiger Einzelfalldaten (siehe Abschnitt 1.4). Wichtige Vertreter der quantitativen Einzelfallmethodik sind in der angelsächsischen Psychologie Hersen und Barlow (1976) oder Kratochwill (1986). Im deutschen Sprachraum sind die Arbeiten von Westmeyer (1979), der Sammelband von Petermann (1996) oder in jüngerer Zeit Köhler (2008) zu nennen.

Der neue Sammelband von Reicherts und Genoud (2015), der dieser Methodik neue Impulse geben möchte, war ausschlaggebend für die Entwicklung dieser Applikation, speziell das 7. Kapitel zu den non-parametrischen Methoden (Reicherts, M., Genoud, P. & Reicherts, L., 2015). Im Text wird in diesem Sinne von dem „Referenzwerk“ oder dem „Referenzkapitel“ gesprochen.

Andere Formen, die in *SCE* nicht behandelt werden, sind z. B. Einzelfallbeschreibungen oder Fallstudien mit rein qualitativem, deskriptivem, oder didaktischem Charakter (z. B. Kasuistiken oder *Case Studies*; siehe auch Reicherts und Genoud, 2015, S. 20 f.).

1.2 Wissensarten, Hypothesen und ihre Überprüfung

Um die verschiedenen Problem- bzw. Aussagebereiche zu differenzieren, ist es nützlich, verschiedene Wissensarten zu unterscheiden (nach Bunge, 1967, 1985; Perrez & Patry, 1982): (a) *Faktenwissen*, (b) *Bedingungswissen* und (c) *Veränderungswissen*. Faktenwissen betrifft das *Know That* mit einer Struktur „A ist/hat X“, Bedingungswissen, das *Know Why*, beinhaltet Aussagen der Form „Wenn-dann“ oder „Je mehr, desto“, das Veränderungswissen *Know How* erlaubt Aussagen im Sinne sog. Technologischer Regeln „B per A“ oder „Um B unter den Bedingungen S zu erreichen, empfiehlt sich A zu tun“. Beispiel:

- (1) Die Zahl der Konflikte eines Paares ist im Vergleich zur vorausgehenden Periode höher (*Faktenwissen*).
- (2) Die Zahl der Paarkonflikte ist höher, weil Kommunikationsprobleme zwischen den Partnern vorliegen (*Bedingungswissen*).

- (3) Um die Zahl der Paarkonflikte zu reduzieren, empfiehlt sich eine Verbesserung der Kommunikationstechniken (*Veränderungswissen*).

Für alle Wissensarten bzw. für die mit ihnen verbundenen Aussagen stellt sich die Frage nach ihrer Gewinnung und (Über-)Prüfung, auch im Kontext von Einzelfällen; sei es in der Forschung – sei es in der Praxis. Die damit verbundenen Hypothesen – als begründete Annahmen – lassen sich folgendermaßen systematisieren (nach Reicherts & Genoud, 2015, S. 32):

- (1) Singuläre Hypothese Bezieht sich auf ein einzelnes Objekt und behauptet das Vorhandensein eines Merkmals.
- (2) Existenz-Hypothese Behauptet das Vorhandensein von wenigstens einem bestimmten Objekt, das ein bestimmtes Merkmal aufweist.
- (3) Universelle Hypothese Behauptet für alle Objekte einer bestimmten Gesamtheit, dass sie ein bestimmtes Merkmal aufweisen.
- (4) Aggregatshypothese Behauptet für eine Gesamtheit von Objekten ein bestimmtes, meist statistisches, Merkmal.

Einzelfallanalysen eignen sich insbesondere zur Formulierung und Überprüfung von singulären Hypothesen und Existenzhypothesen (siehe hierzu auch die Fallbeispiele im 5. Kapitel, v. a. Abschnitt 5.7). Eine Untersuchung von universellen Hypothesen ist nur möglich, wenn die Einzelfallanalysen hinreichend oft *systematisch repliziert werden*. Damit ist die wissenschaftstheoretische Problematik der *Generalisierbarkeit* angesprochen: Während z. B. eine Hypothese, die für einen Einzelfall behauptet, dass eine bestimmte Kommunikationstechnik das Interaktionsverhalten dieser Person verbessert, sich gut mit einer Zeitreihenanalyse (quasi-experimentelle „Interventionsanalyse“) untersuchen lässt, kann eine entsprechende universelle Hypothese, die einen positiven Einfluss dieser Kommunikationstechnik bei allen Personen behauptet, nur durch mehrfache und systematische Replikationen überprüft werden. Und selbst in diesem Fall kann die gewonnene Evidenz nicht abschließend „bestätigt“, sondern nur vorläufig gestützt werden (z. B. Westmeyer, 1996; Perrez & Patry, 1982; siehe auch Abschnitt 5.7). Im Unterschied dazu sind bei Aggregatshypothesen *grundsätzlich* Gruppenuntersuchungen notwendig, um die gesuchten Parameter (z. B. zentrale Tendenz oder Dispersion) einer bestimmten Gesamtheit statistisch fundiert (z. B. durch bestimmte Stichprobenziehung und -größe) zu ermitteln.

Auch im Einzelfall sollte nicht nur die Datengewinnung nach bestimmten Kriterien erfolgen: geeignete Versuchsplanung (Design), Indikatorbildung bzw. Operationalisierung, objektive, reliable und valide Datenerfassung (Reicherts & Genoud, 2015, S. 25; „Komponenten von Versuchsplänen“). Auch die Untersuchung bzw. Prüfung der Hypothesen sollte mit möglichst angemessenen Methoden erfolgen: mit präzisen, möglichst erschöpfenden Indikatoren und trennscharfen, für die statistische Entscheidung sensitiven Methoden. Dabei spielen insbesondere die zu wählenden Tests für die inferenzstatistische Entscheidung eine Rolle (siehe Abschnitt 2.3). Projiziert man die drei Wissensarten auf die beiden großen Bereiche des Forschungskontextes einerseits und des Anwendungskontexts andererseits, lässt sich die Bedeutung der quantitativen Einzelfallanalyse in Sozial- und Verhaltenswissenschaft in nachfolgender Tabelle veranschaulichen (vereinfacht übernommen aus Reicherts & Genoud, 2015, S. 28).

Wissensarten & ihre Aussagenstruktur	Forschungskontext	Anwendungskontext
<i>Tatsachenwissen</i>	1	2
„A ist/hat X“ „Alle Individuen von G haben X“	Einzelfallanalysen für die Vorbereitung von systematischen Gruppenuntersuchungen zum Tatsachenwissen	<i>Einzelfalldiagnose</i> , Fallkontrolle Erfassung von Verhalten und Ereignissen
<i>Bedingungswissen</i>	3	4
„Wenn A – dann B“ „Je A – desto B“	Einzelfallanalytische Experimente: z. B. systematische (Vor)Untersuchungen im Labor im Hinblick auf die Entwicklung kausaler Gesetzmäßigkeiten bzw. Theorien bzw. zu heuristischen Zwecken der Formulierung von Hypothesen im <i>Entstehungszusammenhang</i> von Bedingungswissen	Funktionale (Verhaltens-)Analyse Untersuchung von individuellen Bedingungen (Situationen, Ereignissen) und Konsequenzen (Verhalten und Erleben) Gewinnung <i>individueller</i> „Gesetzmäßigkeiten“ bzw. funktionaler Hypothesen für den Einzelfall
<i>Veränderungswissen</i>	5	6
„B (Ziel) per A (Intervention)“	Einzelfallanalytische Experimente, z. B. systematische <i>Wirksamkeitsuntersuchungen</i> neuer Interventionen mit Replikation	Vorexperimentelle Einzelfallanalyse mit <i>A-B-Design</i> Bestimmung individueller Interventionsziele und Analyse von Veränderungen in Funktion bestimmter Interventionen

Im 5. Kapitel wird in den Fallbeispielen 4 und 5 dargestellt (siehe insbesondere Abschnitt 5.7), wie mit beobachtenden Designs Tatsachenwissen und zunächst hypothetisches Bedingungswissen gewonnen werden kann, bzw. wie mittels Fallkontrolle untersucht wird, ob solches, bereits vorhandenes Wissen auch im interessierenden Fall Gültigkeit hat. In Fallbeispiel 5 (siehe Abschnitt 5.5) wird zudem gezeigt wie mit einem *A-B-Design* (in einer Vorstudie gewonnenes) Bedingungswissen erhärtet und Veränderungswissen gebildet werden, und wie auf diese Weise die Wirksamkeit der Intervention geprüft werden kann.

1.3 Formen und Designs der Einzelfallanalyse

Die verschiedenen Formen der Einzelfallanalyse zur Exploration und Überprüfung bestimmter Verhaltensweisen, Zustände oder Ereignisse verlangen entsprechende *Versuchsanordnungen*. Diese umfassen mehrere Komponenten (Reicherts & Genoud, 2015, Teil II): neben der zugrundeliegenden Fragestellung, die Situation und den Kontext der Erfassung, die untersuchten Variablen, die Datenstruktur (abhängige & unabhängige Variable, Kontrollvariablen, Skalenniveaus), das Design im engen Sinne (z. B. (*quasi-experimentelles* vs. *Beobachtungsdesign*), die Datenquellen und Erhebungsmethoden sowie die Methoden der Datenaufbereitung und statistischen Analyse. Im Mittelpunkt dieser Arbeit steht die *statistische Analyse*: diese soll durch die Entwicklung des Tools *SCE* unterstützt werden.

1.4 Die Quantitative Einzelfallanalyse

Der quantitativ-systematische Einzelfallansatz, der in *SCE* im Mittelpunkt steht, hat seit den 1970er Jahren zunehmend an Bedeutung gewonnen. Wichtige Impulse dazu lieferte die *ARIMA-Modellierung* von Box & Jenkins (1976). Neben diesem neueren parametrischen Ansatz existieren non-parametrische

Methoden schon seit längerer Zeit (z. B. Kendall & Bradford, 1953). Auch diese Methoden wurden seit den 70er Jahren vermehrt für die Einzelfallanalyse eingesetzt.

Nach Petermann (1996) kann die quantitative Einzelfallanalyse durch drei Bestimmungsstücke charakterisiert werden:

- (1) Die Einzelfallanalyse betrachtet eine einzelne Untersuchungseinheit.
- (2) Es geht um die Untersuchungseinheit als ganze; handelt es sich um Aggregate, werden nicht deren Bestandteile betrachtet.
- (3) Die Untersuchungseinheit bzw. ihr Verhalten kann hinsichtlich (a) ihres natürlicherweise auftretenden Verhaltens oder (b) ihrer Veränderungen infolge einer unabhängigen Variable betrachtet werden. Petermann unterscheidet nach (a) *deskriptiver* und (b) *explikativer* Einzelfallanalyse.

Ein zentrales Problem bei jeglicher quantitativen Analyse ist die zusammenfassende Beschreibung von Merkmalen im Sinne von Unterschieden, Zusammenhängen und Trends, für die unterschiedliche statistische Verfahren zur Verfügung stehen (siehe 3. Kapitel). Für die quantitativ-systematische Einzelfallanalyse soll mit diesem Tool ein neuartiger Ansatz der benutzergeführten Durchführung solcher Analysen verfolgt werden.

Zur Erhebung und Aufbereitung der Daten existieren verschiedene Formen: zum Beispiel die diagnostische Analyse eines Einzelfalles, die Zielerreichungsanalyse, die Veränderungsmessung durch Messwiederholung und die Analyse von Verlaufsstrukturen und Prozessen oder die Aggregation von Einzelfallanalysen (siehe Reicherts & Genoud, 2015, S. 25). Jede Form verlangt entsprechende Designs, welche bestimmte Daten generieren, die wiederum passender statistischer Methoden bedürfen.

Die Anwendung entsprechender statistischer Methoden ermöglicht die Generierung von Evidenz, was sowohl für die Forschung als auch für die Praxis essenziell ist. Durch die Entwicklung eines Tools zur statistischen Auswertung von Einzelfalldaten können passende Methoden auf anwenderfreundliche Weise zur Verfügung gestellt werden.

1.5 Die Idee eines Expertensystems für die Analyse von Einzelfällen

Um die Verwendung von Einzelfallanalysen in der angewandten Forschung und Praxis, und damit das Potential an Wissens(zu)gewinn und Interventionskontrolle zu unterstützen und zu verbessern, ist die Idee eines computergestützten „Expertensystems“ entstanden. Da die statistische Auswertungs-Methodik für viele Praktiker – und angewandte Forscher – das zentrale Problem bei der Durchführung von Einzelfallanalysen darstellt, möchte das Tool an diesem Punkt ansetzen. Als Webapplikation sollte *SCE* verschiedene *deskriptive* und *inferenzstatistische* Methoden (siehe 3. Kapitel) in einer grafischen und geführten Benutzeroberfläche bereitstellen und unterstützen. Es möchte insbesondere dem Mangel an relativ einfachen, zugänglichen Analysemöglichkeiten von Einzelfällen in der Praxis entgegenwirken.

Das Tool richtet sich zwar hauptsächlich an Praktiker, ist aber als Schnittstelle zwischen Forschung und Praxis zu verstehen, und kann z. B. auch für Studierende in den Sozial- und Wirtschaftswissenschaften nützlich sein. Diese Ausrichtung begründet die Hauptkriterien der Entwicklung dieser Applikation, die sich hauptsächlich auf die Qualität der Interaktion und die angemessene Interpretierbarkeit konzentrieren (siehe Abschnitt 4.1.3).

2 Statistische Grundlagen

2.1 Deskription, Inferenz und Exploration

Die Einzelfallanalyse zielt auf die Beantwortung verschiedener Fragen im Hinblick auf Zustände oder Verhaltensweisen eines Objektes und deren Veränderung (in Abhängigkeit von anderen Variablen). Solche Fragestellungen können sich auf (a) *Tatsachen-*, (b) *Bedingungs-* oder (c) *Veränderungswissen* beziehen (siehe Abschnitt 1.2):

- (a) Wie lässt sich die Verteilung/Ausprägung eines Zustands oder Verhaltens *beschreiben*?
- (b) Wie lässt sich die Verteilung/Ausprägung eines Zustands oder Verhaltens *erklären*?
- (c) Wie lässt sich die Verteilung/Ausprägung eines Zustands oder Verhaltens *verändern*?

Bei allen drei Fragestellungen dienen deskriptive Statistiken der Beschreibung bzw. der *Bildung* von Hypothesen, Inferenzstatistiken hingegen dienen der Beurteilung der überzufälligen, statistisch signifikanten Bedeutsamkeit bzw. der *Stützung* von Hypothesen.

Untersuchungen haben *explorativen Charakter*, wenn im Vorfeld nur beschränkt Annahmen bestehen und/oder sie nicht das Ziel haben, die Zustände und Verhaltensweisen eines Objektes abschließend nachzuweisen oder zu begründen. Im gesamtwissenschaftlichen Kontext sind Einzelfallstudien nicht zuletzt aufgrund der meist kleinen individuellen Datenmenge oft explorativer Art.

2.2 Skalenniveaus

Skalenniveaus sind eine wichtige Eigenschaft von Merkmalen, die durch die Art der Messung definiert werden. Folgende Beschreibung orientiert sich an Bortz, Lienert und Boehnke (2008, S. 61 ff.). Die einzelnen Stufen unterliegen einer hierarchischen Ordnung und lassen sich, vereinfachend, anhand folgender Tabelle unterscheiden (entnommen aus Reicherts, M., Genoud, P. & Reicherts, L., 2015, S. 199):

Skalenniveau	Zulässige Aussagen	Stat. Maßzahlen/Methoden	Beispiele
Nominal	Gleichheit Verschiedenheit	Frequenzen Kontingenzkoeffizient Vierfelderkorrelation <i>Phi</i>	Geschlecht dichotomisierte Merkmale
Ordinal	Größer-Kleiner	Median Perzentil Rangkorrelation	Rohwerte psychol. Tests Ratings
Intervall	Gleichheit von Intervallen und Unterschieden	Mittelwert Varianz Produkt-Moment-Korrelation	Grad Celsius Standard-Testnormen
Verhältnis	Gleichheit von Summen, Vielfachen und Quotienten	geometrisches Mittel Variabilitätskoeffizient	Länge, Gewicht Zeitintervalle evozierte Potentiale

Die Messstruktur der analysierten Objekte lässt mit den jeweiligen empirischen und numerischen Relativ (abstrakte Beschreibungen der Skalenkonstruktion) wie folgt beschreiben.

2.2.1 Nominalskalen

Das *empirische Relativ* nominalskaliertener Merkmale (gleich/nicht gleich), wird vom *numerischen Relativ* durch Werte repräsentiert, die dieses identifizieren. Die Wahl der Zahlen ist eigentlich willkürlich, es geht lediglich darum, dass gleiche Objekte gleiche Zahlen erhalten.

In den Sozialwissenschaften sind nominalskalierte Daten häufig in Verbindung mit *dichotomen Variablen* relevant, wie zum Beispiel 1, 2 für zwei Arten/Ausprägungen eines Merkmals oder 0, 1 (*Binärkodierung*) für das Nicht-Eintreten bzw. Eintreten eines Ereignisses oder Verhaltens.

Für die korrekte Auswertung müssen dichotome Variablen in *SCE* immer in *Binärkodierung* eingegeben werden. Nachfolgend wird deshalb in der Regel von einer *binären Variable* ausgegangen.

2.2.2 Ordinalskalen

Für das *empirische Relativ* ordinalskaliertener Merkmale gilt eine sog. „*schwache Ordnungsrelation*“, nach welcher für jedes Objektpaar bestimmbar sein muss, welches der beiden Objekte dominiert. Dabei gilt auch die Transitivität: wenn *A* größer *B* und *B* größer *C*, dann *A* größer *C*. Die Werte des numerischen Relativs bilden so eine Rangordnung für alle Merkmalsausprägungen. Die Abstände zwischen den Objekten sind damit nicht gleich groß; es besteht keine Äquidistanz.

2.2.3 Intervallskalen

Intervallskalen beinhalten nicht nur die Information, ob ein Objekt ein anderes dominiert, sondern auch um wie viel es dominiert, z. B. größer ist. Das bedeutet, gleich große Unterschiede von zwei Objektpaaren des empirischen Relativs werden durch gleich große, sog. äquidistante, numerische Differenzen repräsentiert.

2.2.4 Verhältnisskalen

Verhältnisskalen verfügen im Gegensatz zu Intervallskalen zusätzlich über einen absoluten Nullpunkt. Dies erlaubt *Ähnlichkeits-Transformationen*, ohne dass die numerische Aussage ungültig wird. Die Aussage, *A* sei doppelt so groß wie *B*, ist hier im Gegensatz zu Intervallskalen, wie z. B. der Celsius-Skala möglich.

Intervall- und Verhältnisskalen können unter dem Begriff der „*Kardinalskalen*“ zusammengefasst werden.

2.3 Parametrische und non-parametrische Methoden

Für die Einzelfallanalyse bieten sich bei der inferenzstatistischen Analyse häufig non-parametrische Verfahren an, da diese im Gegensatz zu parametrischen Verfahren weniger Anforderungen an die Daten stellen; sie sind sog. „*verteilungsfrei*“. Allerdings sind sie dadurch weniger sensitiv als die analogen parametrischen Methoden, sofern deren Voraussetzungen erfüllt sind, denn werden parametrische Methoden unter „*nichtparametrischen*“ Bedingungen angewendet, ist das Gegenteil der Fall, insbesondere bei kleineren Stichproben (Bortz et al., 2008, S. 45). Dementsprechend ist jeweils zu prüfen, ob parametrische Tests angewendet werden *dürfen*. Bestehen diesbezüglich Zweifel, ist die Wahl eines non-parametrischen Verfahrens methodisch meist sinnvoller.

Ein zentrales Problem der Einzelfallanalyse – im Gegensatz zur Gruppen- oder Aggregatstatistik – ist die serielle Abhängigkeit der gewonnenen intraindividuellen Daten in Verbindung mit der Autokorrelation. Oft sind Zeitreihen durch die zeitlichen Abhängigkeiten zwischen den einzelnen Messungen davon betroffen.

Da parametrische Methoden (z. B. *t-Test*) gleichverteilte und meist voneinander unabhängige Daten voraussetzen, eignen sie sich für Zeitreihen nicht, es sei denn, die serielle Abhängigkeit wird beispielsweise durch die Elimination der autoregressiven Komponente kontrolliert (z. B. im Rahmen von *ARIMA-Modellierung* der Zeitreihe; Box & Jenkins, 1976, vgl. auch Köhler, 2008).

Diese Problematik beschreiben auch Morell und Fried (2009) in ihrem Artikel zu Trendanalysen mithilfe non-parametrischer Tests in saisonalen Zeitreihen. Dort analysieren sie das Verhalten einiger non-parametrischer Tests im Zusammenhang mit autokorrelierten Daten. Sie rechtfertigen die Verwendung non-parametrischer Methoden einerseits durch die oben erwähnten geringeren Anforderungen an die Daten, andererseits durch ihre höhere Flexibilität bezüglich der Struktur der Testfunktion (Morell & Fried, 2009, S. 1). Mit anderen Worten, sie favorisieren non-parametrische Methoden, da diese flexibler mit Daten umgehen können, von denen man die Verteilung nicht kennt (und es (zu) aufwändig wäre, diese zu ermitteln). Dieser Ansatz ist auch für *SCE* bedeutsam: Da das Tool im Rahmen von Einzelfallanalysen *generell* einsetzbar sein soll, muss es unterschiedliche Datentypen und Verteilungen „behandeln“ können. Des Weiteren würde es für die Entwicklung und Benutzung einen erheblich größeren Aufwand bedeuten, ein Tool zur Verfügung zu stellen, welches parametrische Tests verwendet (z. B. gemäß der *ARIMA-Modellierung*). Dies würde eine deutlich komplexere und anspruchsvollere Modellierungs- und Analysearbeit verlangen, die für praktische Anwendungen (gegenwärtig) kaum in Frage kommt – und auch von großen Statistik-Programmen (z. B. SPSS) bisher nicht geleistet wird.

Auch die vorgeschlagenen non-parametrischen Verfahren haben jedoch – trotz ihrer inferenzstatistischen Möglichkeiten – vor allem *explorative* Funktion, und auch schlüssige Ergebnisse sollten deshalb vorsichtig interpretiert werden (vgl. auch Reicherts, M., Genoud, P. & Reicherts, L., 2015).

3 Statistische Verfahren in *Single-Case Expert*

Zu den zentralen Problemstellungen der Auswertungen mit *SCE* gehört die Ermittlung von Trends, insbesondere von Trends in der *zentralen Tendenz* und der *Streuung von Variablen im Zeitverlauf*. Speziell für (quasi-)experimentelle Untersuchungen stehen die Analyse von *Niveau-Unterschieden* und die Ermittlung von *Zusammenhängen* zwischen zwei Variablen im Vordergrund. Auch die Untersuchung *serieller Abhängigkeiten*, insbesondere *zyklischer Schwankungen* kann von Interesse sein.

3.1 Unterstützte Skalenniveaus

In *SCE* werden non-parametrische Methoden insbesondere für binäre und ordinalskalierte Merkmale zur Verfügung gestellt. Diese Ausrichtung rührt daher, dass im sozialwissenschaftlichen Kontext viele Daten auf (persönlichen) Einschätzungen beruhen, die kaum Intervall- oder gar Verhältnisskalen erreichen können (abgesehen von speziellen Methoden der Psychometrie, siehe z. B. Schmid und Reicherts, 2015, S. 131 ff.). Entsprechend sind (echte) Intervall- und Verhältnisskalen in den Sozialwissenschaften eher selten. Die Methoden zur Analyse ordinalskalierter Merkmale können jedoch grundsätzlich auch auf Daten mit höherem Skalenniveau (Intervall-/Verhältnisskala) angewendet werden (siehe dazu Fallbeispiel 5, Abschnitt 5.5).

3.2 Deskriptiv-statistische Methoden

Für die deskriptiv-statistische Analyse werden in der Applikation statistische Maßzahlen sowie grafische Darstellungen zur Verfügung gestellt, die an das Skalenniveau der jeweiligen Daten angepasst sind.

3.2.1 Statistische Maßzahlen

Die dargestellten statistischen Maßzahlen (*Summary Statistics*) sollen die Werteverteilung in ihrer Gesamtheit kompakt repräsentieren (Beispieldaten von zwei Variablen mit je 30 Messungen):

	<i>n</i>	<i>median</i>	<i>mean</i>	<i>sd</i>	<i>min</i>	<i>max</i>	<i>range</i>
<i>x</i>	30.000	5.000	5.200	2.188	1.000	9.000	8.000
<i>y</i>	30.000	4.000	4.567	1.888	1.000	8.000	7.000
<i>D</i>	0.000	-20.000	-12.173	-13.711	0.000	-11.111	-12.500

- | | |
|--|------------------------|
| (1) Anzahl der Messungen | (4) Standardabweichung |
| (2) Median* (insbes. bei Ordinalskala zu verwenden) | (5) Minimaler Wert |
| (3) arithmetisches Mittel
(insbes. bei Intervall-/Verhältnisskala zu verw.) | (6) Maximaler Wert |
| | (7) Spannweite/Umfang |

Werden *zwei* Variablen (wie im obigen Beispiel) oder *zwei* Zeitreihen-Segmente einer Variablen analysiert, gibt die Anwendung jeweils auch die *prozentualen Differenzen (D)* zwischen den Werten aus, womit die Beurteilung der Unterschiede oder auch die Hypothesenbildung erleichtert werden soll.

3.2.2 Grafische Anzeige

Weiter werden die Messwerte der jeweiligen Variablen in einem Graphen ausgegeben, der die entsprechenden Messwerte (Punkte) mit einer Geraden verbindet und eine Trendlinie berechnet. Die Trendlinie wird mit der – einfachen – *Split-Middle Methode* nach White (1974) bestimmt: Die Zeitreihen-Messungen der Variable werden in zwei gleich lange Abschnitte geteilt (bei ungerader Anzahl Messungen (*N*), wird der mittlere Wert in keinen der Abschnitte aufgenommen). Für diese werden die Mediane berechnet, welche die y-Koordinaten der beiden Punkte der Trendlinie bilden; als x-Koordinate wird die Mitte des jeweiligen Zeitabschnittes verwendet.

Abbildung 3.1: Grafische Darstellung einer Zeitreihe mit zwei ordinalskalierten Variablen

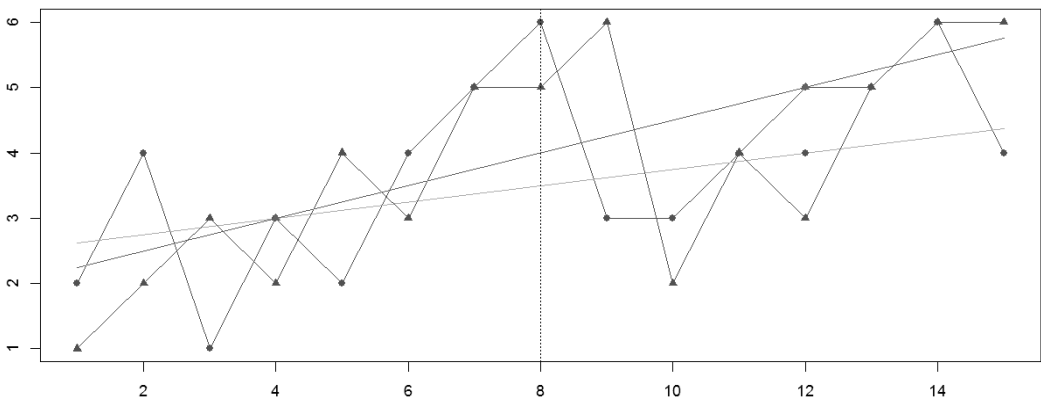
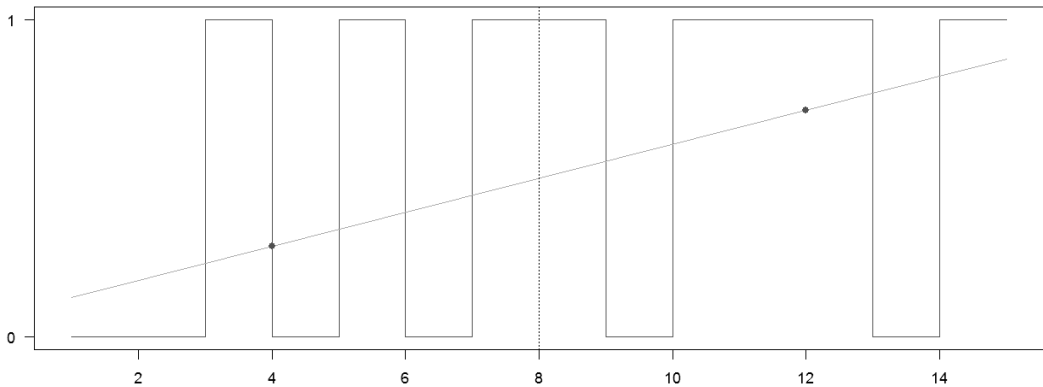


Abbildung 3.2: Grafische Darstellung einer Zeitreihe mit einer binären Variable



Bei allen in diesem Text dargestellten Graphen entspricht die x-Achse der Zeit (t_i) und die y-Achse den Messungen der Variable(n) x_i (oder y_i). Die Achsenbeschriftungen werden deshalb jeweils weggelassen.

3.3 Inferenzstatistische Methoden

3.3.1 Überblick

Können die Voraussetzungen parametrischer Verfahren erfüllt werden – sind diese aufgrund ihrer höheren Effizienz – gegenüber den non-parametrischen Methoden allgemein zu bevorzugen (siehe Abschnitt 2.3). Da bei Zeitreihen diesbezüglich meist eine gewisse Unsicherheit besteht und Einzelfallanalysen oft mit einem sehr kleinen N auskommen müssen, sind die Voraussetzungen der parametrischen Methoden häufig nicht erfüllt. Es ist dann methodisch meist sinnvoller, die voraussetzungsärmeren non-parametrischen Verfahren zu wählen, wie sie in *SCE* implementiert sind.

Nachfolgend sind die in dieser Version von *SCE* unterstützten, überwiegend verteilungsfreien, Inferenzstatistiken aufgelistet:

- (1) Eine ordinale Variable:
 - (a) Trendanalyse: Rangkorrelationskoeffizienten Rho & Tau (x_i gegen t_i)
 - (b) Trendanalyse: S_2 -Test (Niveau-Unterschied Vergleich)
 - (c) Analyse serieller Abhängigkeiten: *Dufour-Test*
 - (d) Analyse der Varianzveränderung: *Rekordbrechertest*
 - (e) Trendanalyse: *lineare Regression* (x_i gegen t_i) [parametrisches Verfahren]

- (2) Zwei ordinale Variablen:
 - (a) Zusammenhangsanalyse: Rangkorrelationskoeffizienten Rho & Tau (x_i gegen y_i)
 - (b) Zusammenhangsanalyse: *lineare Regression* (x_i gegen y_i) [parametrisches Verfahren]

- (3) Eine binäre Variable:
 - (a) Trendanalyse: *Mann-Whitney U-Test* / *Wilcoxon-Rangsummentest*

- (4) Zwei binäre Variablen:
 - (a) Zusammenhangsanalyse: *Phi-Koeffizient*

(5) Eine ordinale und eine binäre Variable:

(a) Zusammenhangsanalyse: *Mann-Whitney U-Test* / *Wilcoxon-Rangsummentest*

Im Folgenden soll ein Überblick über diese Methoden gegeben werden, wobei auch – gerade bei den weniger verbreiteten Methoden – kurz auf deren Stärken und Schwächen insbesondere im Kontext der Einzelfallanalyse (z. B. die Anfälligkeit der Zusammenhangsanalyse auf Trends, siehe unten) eingegangen wird. Da es hier nicht um die Erklärung bestehender Verfahren gehen soll (hierfür ist das Referenzkapitel oder Bortz et al. (2008) zu empfehlen), wird nur dort auf die Berechnungsschritte eingegangen, wo dies zur Veranschaulichung und zur Begründung der Implementierung des jeweiligen Tests in *SCE* notwendig ist. Neben ihrer – relativen – Robustheit war das wichtigste Kriterium für die Wahl der Methoden die Einfachheit ihrer *Anwendung* und *Interpretation*.

Bei den meisten in *SCE* implementierten Tests wird zweiseitig getestet. Eine Ausnahme bilden der *U-Test* und der *S₂-Test*: Hier werden beide gerichteten Hypothesen jeweils einseitig getestet (Aufwärts- oder Abwärtstrend, bzw. positiver oder negativer Zusammenhang). Im Allgemeinen wird empfohlen, das Tool nur mit $N > 10$, besser noch $N > 15$ zu verwenden. Bei Zeitreihen die unter diesem Niveau liegen, können bestimmte Inferenzstatistiken nicht zu einem signifikanten Ergebnis kommen. Problematisch sind derart kurze Zeitreihen auch aus anderen methodischen Erwägungen (neben des geringeren Informationsgehalts z. B. aufgrund des Risikos zufälliger Verzerrungen oder Extremwerte).

3.3.2 Inferenzstatistische Methoden für ordinalskalierte Variablen

3.3.2.1 Rangkorrelationskoeffizienten: Spearmans *Rho* & Kendalls *Tau*

Die Rangkorrelationskoeffizienten Spearmans *Rho* (ρ) und Kendalls *Tau* (τ) sind in *SCE* von großer Bedeutung. Zum einen in der häufigsten Verwendung für die *Analyse des Zusammenhangs* zwischen zwei Variablen (siehe Übersicht oben: *2a*), aber auch für die *Trendanalyse* (einer Variablen über die Zeit) (*1a*).

Die beiden Korrelationskoeffizienten beschreiben die Stärke eines *monotonen Zusammenhangs* zwischen zwei Rangreihen und nehmen Werte zwischen -1 und 1 an. Weichen die Werte signifikant von Null ab, spricht dies für einen negativen bzw. positiven Zusammenhang.

Die beiden Verwendungsformen der Rangkorrelationskoeffizienten beinhalten einen kleinen Unterschied in den Berechnungsschritten. Für die *Korrelationsanalyse* zweier Variablen (vgl. Bortz et al., 2008, S. 414 ff.) wird in der Regel folgendermaßen vorgegangen:

- (1) Bildung der Rangreihen für beide Variablen (sofern diese noch nicht ordinalskaliert sind).
- (2) Sortierung der Wertepaare $(x_1, y_1; x_2, y_2; \dots)$ nach Größe der Ankerreihe $R(x_i)$.
- (3) Berechnung mittels der Formeln für *Rho* und *Tau*.

Bei der Trendanalyse *einer* Variablen erübrigt sich der zweite Schritt, da die „Variable“ Zeit (t_i) mit den Messzeitpunkten bereits die sortierte Ankerreihe $R(t_i)$ bildet (vgl. Bortz et al., 2008, S. 580 ff.).

Eine Schwäche der beiden Teststatistiken besteht darin, dass sie im Fall von (*2a*) den Zusammenhang zweier Variablen *über-* oder *unterschätzen* können, wenn den Wertverläufen *gleichgerichtete* oder *gegenläufige* Trends zugrunde liegen, wie in den Beispielen in Reicherts und Genoud (2015, S. 199; vgl. auch Köhler, 2008) ersichtlich: Die innere trendbedingte Variation der Variablen hat u. U. einen so erheblichen Einfluss auf die Teststatistik, dass diese zu einer Bestätigung des Zusammenhangs kommen kann, obwohl gar keine „*residuale*“ Kovarianz der Variablen vorliegt (*Fehler 1. Art*). Im Gegenzug kann es bei gegenläufigen Trends zur Unterschätzung des Zusammenhangs kommen (*Fehler 2. Art*).

In der vorliegenden Version des Programms wurden der Vollständigkeit halber beide Korrelationskoeffizienten implementiert, obwohl sich *Kendalls Tau* für ordinale Skalen oft besser eignet als *Spearman's Rho*, welches in diesem Fall die Zusammenhänge tendenziell überschätzt (Bortz et al., 2008, S. 443 ff.).

3.3.2.2 Cox-Stuart Trendtest / S_2 -Test

Der *Cox-Stuart Trendtest* bzw. S_2 -Test nach der Methode aus Bortz et al. (2008, S. 585 f.) untersucht ebenfalls einen *monotonen Trend* in einer Messreihe (1b), jedoch nach einem anderen Ansatz: Es wird die erste Hälfte der Messreihe mit der zweiten verglichen. Der S_2 -Test eignet sich dadurch gerade gut für Messreihen mit einer *Prä-* und *Post-* bzw. *Baseline-* und *Interventionsphase* (siehe auch Auswertung von Fallbeispiel 5, Abschnitt 5.5.7).

Eine weitere Form bildet der sog. S_3 -Test (vgl. Bortz et al., 2008, S. 586), welcher das erste und letzte Drittel einer Zeitreihe vergleicht, was bei Untersuchungsdesigns interessant ist, die neben *Prä-* und *Post-* auch eine mittlere Phase haben (z. B. zur Vorbereitung der Intervention). Im Grunde können beliebige gleich lange Abschnitte mit derselben Testlogik auf einen Trend geprüft werden, sofern vor der Durchführung der Analyse deren Länge festgelegt wurde.

Die Nullhypothese bezeichnet das *Nicht-Vorhandensein* eines Trends. Beim Testen *einer* Trendrichtung wird *einseitig*, andernfalls *zweiseitig* getestet. In *SCE* kann einseitig auf einen positiven oder negativen Trend getestet werden. Bei der rechnerischen Durchführung wird jeweils der erste Wert der ersten Hälfte mit dem ersten Wert der zweiten Hälfte verglichen, anschließend der zweite Wert der ersten Hälfte mit dem zweiten Wert der zweiten Hälfte, usw. Die Anzahl der dabei festgestellten positiven oder negativen Differenzen wird anhand der Binomialverteilung zufallskritisch bewertet.

Aus diesem Vorgehen wird deutlich, dass der Test die beiden Abschnitte nicht nur als „Ganzes“ betrachtet, sondern auch deren Binnensequenz; dies kann gerade dann von Vorteil sein, wenn den beiden Hälften eine *sequentielle Schwankung* (durch eine Drittvariable) zugrunde liegt (z. B., wenn das Erleben einer Person über zwei Wochen betrachtet wird – was eine eher kurze Zeitreihe wäre – vergleicht man zwei *Wochen-Rhythmen*).

Bei der Anwendung des Tests sind folgende Eigenheiten und Einschränkungen zu berücksichtigen:

- (1) Der Test prüft grundsätzlich das Vorhandensein eines Trends; da die jeweilige Sequenz aber miteinbezogen wird, kann eine sehr hohe Übereinstimmung der (relativen) Sequenzen eine Falsifikation der Nullhypothese zur Folge haben, obwohl eigentlich kein Trend vorhanden ist.
Z. B.: $x_i = \{2, 1, 3, 5, 4, 7, 6, 3, 2, 4, 6, 4, 8, 7\}$ ergibt beim Testen auf einen Aufwärtstrend $p = 0.0156$, woraus wir auf einen signifikanten zunehmenden Trend schließen, obwohl beide Hälften den gleichen Median haben (siehe Abbildung 3.3).
- (2) Im Gegenzug ist es möglich, dass die Nullhypothese verworfen wird, obwohl die Sequenzen sogar gegenläufig sind (Cox & Stuart, 1955, S. 16).
Z. B.: $x_i = \{1, 2, 4, 4, 6, 8, 11, 19, 18, 18, 16, 13, 12, 12\}$ ergibt beim Testen auf einen Aufwärtstrend $p = 0.0078$, woraus wir bei zweiseitigem Test ebenfalls auf einen signifikanten zunehmenden Trend schließen (siehe Abbildung 3.4).
- (3) Da der Test auf der Binomialverteilung beruht, spielt die Wahrscheinlichkeitsverteilung bei der Berechnung des p -Wertes eine wichtige Rolle. Aufgrund der theoretischen Verteilung kann der p -Wert erst ab $N \geq 14$ unter das Niveau von 5% fallen, wodurch sich dieser Test im Grunde nur für Messreihen mit $N \gg 14$ eignet (u.a. deshalb wird in Abschnitt 3.3.1 $N > 15$ empfohlen).

Abbildung 3.3: Fast übereinstimmende Sequenzen der ersten und zweiten Hälfte (Beispiel 1)

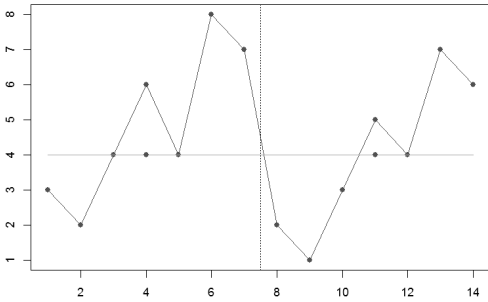
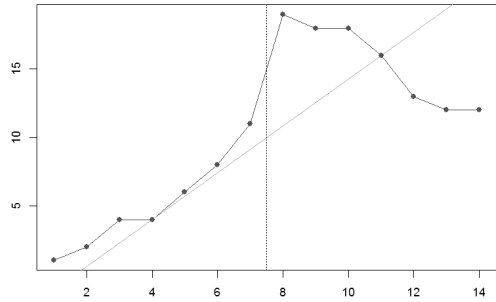


Abbildung 3.4: Gegenläufige Sequenzen/Trends der ersten und zweiten Hälfte (Beispiel 2)



3.3.2.3 Dufour-Test

Ob eine Zeitreihe überzufällige *serielle Abhängigkeiten* enthält, wie zum Beispiel *zyklische oder saisonale Schwankungen*, kann mit dem *Dufour-Test* (Dufour, 1981) geprüft werden (1c). Dieser konnte, wie der nachfolgende *Rekordbrechertest*, im Gegensatz zu den anderen hier verwendeten Verfahren nicht als vorprogrammierte Funktion gefunden werden (siehe hierzu 4. Kapitel, insbesondere Abschnitt 4.1.7). Die Tests wurden ausgehend von der Vorgehensweise nach Morley und Adams (1989) in *SCE* programmiert und implementiert.

Zunächst wird von den einzelnen Werten der Zeitreihe – wenn die Zeitreihe nicht bereits einen *Nullmedian* hat – der *Median* subtrahiert und daraus eine sog. *reduzierte Serie* gebildet. Anschließend wird ein plausibler *Lag-Wert* festgelegt, welcher in *SCE* direkt eingegeben werden kann. Der Wert x_i wird jeweils mit dem Wert x_{i+Lag} der *reduzierten Serie* multipliziert. Den *Beträgen* dieser neuen *Lag-Reihe* werden Ränge zugeordnet. Anschließend werden für die Ränge mit jeweils positiven oder negativen zugrundeliegenden *Lag-Werten* Summen gebildet und mit dem *Mann-Whitney U-Test* (siehe Abschnitt 3.3.3.1) zufallskritisch überprüft.

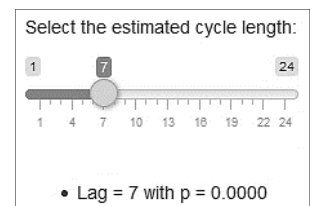
Zu den Vorteilen des Verfahrens gehört neben der Möglichkeit, mehrere *Lags* gegeneinander zu testen unter anderem auch die Robustheit gegen unterschiedliche Streuung innerhalb der Daten (sog. *Heteroskedastizität*). Zu den „Nachteilen“ gehört die Annahme einer symmetrischen Verteilung (vgl. Dufour, 1981). Bei der Anwendung des Tests an verschiedenen Zeitreihen, deren Verteilungen gezielt in der Symmetrie verändert wurden, variierte der *p-Wert* jedoch meist nur um sehr wenige Prozentpunkte.

Der Test ist wie *Rho* und *Tau* ebenfalls anfällig auf Trends; ist ein solcher vorhanden, werden die Abhängigkeiten zwischen den Werten tendenziell *überschätzt*, was zu einem *Fehler 1. Art* führen kann.

Während mit *Lag = 1* lediglich die *unmittelbare serielle Abhängigkeit* analysiert werden kann, erlaubt die Prüfung eines *Lag ≥ 2* das Vorhandensein einer *zyklischen Schwankung* (ein *Zyklus* umfasst mindestens 2 einander folgende Messpunkte).

Die rechnergestützte Durchführung hat gegenüber der manuellen Durchführung neben der generellen Zeiteinsparung zudem den Vorteil, dass der Benutzer mehrere Werte durch Verschieben eines „*Sliders*“ (siehe Abbildung 3.5) *in Echtzeit* testen kann. Dieser erlaubt das Prüfen von *Lags* zwischen 1 und 24 (es kommt selten vor, dass *Lags > 10* analysiert werden, dennoch wurde 24 als Maximum gewählt, sodass bei Bedarf auch z. B. stündlich *circadiane Zyklizitäten* getestet werden können).

Abbildung 3.5: Der „*Slider*“ für den *Dufour-Test*:



3.3.2.4 Rekordbrechertest

Soll geprüft werden, ob sich die Streuung einer Zeitreihe über die Zeit verändert – ob also ein *Trend in der Varianz* vorliegt – kann der *Rekordbrechertest* von Foster und Stuart (1954) verwendet werden (1d). Auch hier wurde von der Beschreibung von Morley und Adams (1989) ausgegangen.

Eine solche Zeitreihe hat normalerweise die Eigenschaft, dass sich ihre Werte im Zeitverlauf immer weiter – relativ gleichmäßig – über- und unterbieten. Der *Rekordbrechertest* ermittelt dementsprechend, wie oft im Verlauf der Zeitreihe ein neuer *Höhen-* oder *Tiefenrekord* aufgestellt wird und summiert die beiden Werte für deren anschließende Überprüfung (mit *konservativer Kontinuitätskorrektur*) mit der Standardnormalverteilung, da sich die Summe der Rekorde rasch einer Normalverteilung nähert (vgl. Bortz et al., 2008, S. 584).

Im Folgenden soll ein kurzer Überblick über die Stärken und Schwächen des *Rekordbrechertests* gegeben werden, wofür nachfolgende Zeitreihe (siehe Abbildung 3.6) betrachtet wird.

Abbildung 3.6: Zeitreihe mit abnehmendem Trend in der Varianz

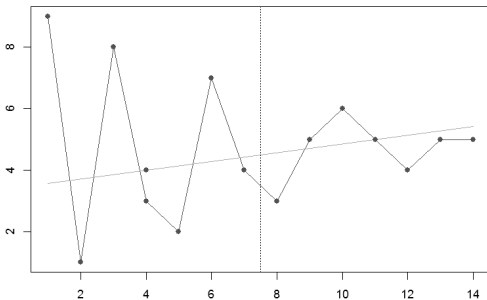
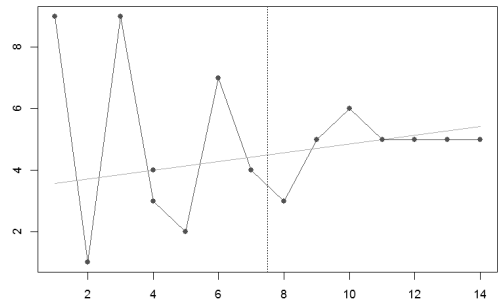


Abbildung 3.7: Zeitreihe mit abnehmendem Trend in der Varianz (t_3 & t_{12} modifiziert)



Betrachtet man Abbildung 3.6, fällt sofort auf, dass in der linken Hälfte die Werte weiter auseinanderliegen als in der rechten, womit sich die Vermutung einer *Veränderung der Varianz* – genauer einer *abnehmenden* Varianz im Zeitverlauf – aufdrängt. Erwartungsgemäß ist die Teststatistik beim „Rückwärtszählen“ mit einem $p=0.0216$ signifikant, was die Hypothese eines *abnehmenden* Trends in der Varianz stützt.

Zur Veranschaulichung der eher geringen Robustheit dieses Tests, werden nachfolgend lediglich zwei Werte aus der Zeitreihe um eine Einheit erhöht: t_3 und t_{12} (siehe Abbildung 3.7).

Dadurch wird, wenn man Abbildung 3.7 betrachtet, der Unterschied in der Varianz zwischen dem Anfang und Ende der Reihe im Vergleich zu Abbildung 3.6 *noch deutlicher*: Mit der Erhöhung von t_3 auf 9, liegt t_3 weiter vom *Median* (= 5) der Reihe weg, wodurch sich die Dispersion zu Beginn der Zeitreihe *verstärkt*; mit der Erhöhung von t_{12} auf 5 entspricht dieser neue Wert genau dem *Median*, wodurch sich die Dispersion am Ende der Zeitreihe *vermindert*. Durch die noch deutlichere Abnahme der Dispersion im Verlauf der Zeitreihe, müsste die Teststatistik eigentlich *noch „signifikanter“* ausfallen. Es ist jedoch das *Gegenteil* der Fall: die Teststatistik ist mit $p=0.2507$ nicht mehr signifikant. Der Grund dafür ist, dass sich durch die Anpassung der beiden Werte die Summe der Rekorde um 2 *verringert* (!) und damit die Teststatistik *abgeschwächt* wird.

Neben dieser Problematik – der sich wiederholenden Werte – weist der Test noch weitere nachteilige „*Sensibilitäten*“ auf. Beispielsweise kann eine *kleine Skalenbreite* (kleiner Wertebereich) dazu führen, dass sich die Werte häufiger wiederholen und somit weniger Rekorde entstehen. Weil mit jedem weiteren

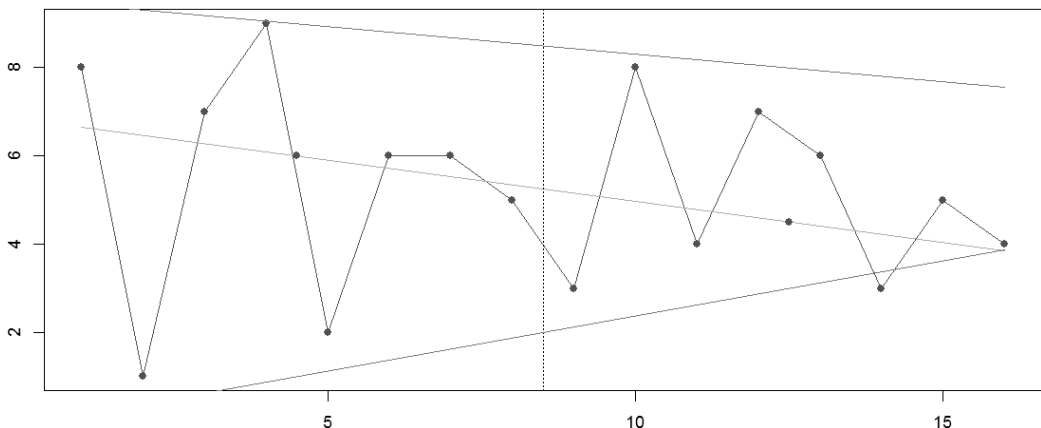
Wert die Wahrscheinlichkeit eines neuen Rekords abnimmt, verliert der Test bei *zunehmender Länge der Zeitreihe* ebenfalls an Stärke – Morley und Adams (1989) nennen hier einen oberen Grenzwert $N=15$, in Bortz et al. (2008, S. 584) ist kein fixer Wert zu finden. Aufgrund der nicht durchweg eindeutigen Darstellung wird in *SCE* diese Obergrenze nicht so strikt gehandhabt, zumal auch die *Skalenbreite* darauf einen Einfluss hat (siehe oben).

Zudem kann es bei einem *Auf-* oder *Abwärtstrend* im Werteverlauf *ohne* Vorhandensein einer Dispersion zu einem *Fehler 1. Art* kommen: Die Zeitreihe *über-* oder *untertrifft* sich selbst kontinuierlich *nur* oder zumindest *überwiegend* in einer Richtung, was trotz der *Einseitigkeit* der Rekorde die Rekordsumme erhöht. Wenn ein Trend in der Zeitreihe (z. B. mittels *Rho/Tau*) nachgewiesen werden konnte, ist der Test daher *nicht* anzuwenden, was einen Nachteil dieses Verfahrens darstellt, denn ein *Dispersionstrend* tritt oftmals unabhängig von einem *Trend in der zentralen Tendenz* auf. Nach der Durchführung des Tests an verschiedenen Zeitreihendaten konnte beobachtet werden, dass der Test in dieser Version von *SCE* mit Zeitreihen mit 10 bis 30 Messungen und einer Skalenbreite > 10 relativ gut funktioniert.

Der *Rekordbrechertest* ist in dieser Grundform, wie aus der Darstellung ersichtlich, nur begrenzt einsetzbar. Es wird deshalb beabsichtigt, in einer künftigen Version von *SCE* eine verbesserte Version zu implementieren. Zunächst soll eine *Trendbereinigung* in die Berechnungsprozedur integriert werden, sodass auch Zeitreihen mit einem *Trend in der zentralen Tendenz* auf Varianzveränderungen geprüft werden können. Des Weiteren wird die Implementierung der *Gewichtungsmethode* nach Morell und Fried (2009) beabsichtigt, welche den „vorderen“ Rekorden tiefere Gewichte gibt als den „hinteren“, wodurch die Stärke des Tests bei längeren Zeitreihen nicht abnimmt.

Da der Test nicht sehr robust ist, wurde zur erleichterten Beurteilung zudem eine grafische Methode aus Kapitel 6.4 des Referenzwerkes (Genoud, 2015) integriert: In die grafische Darstellung der Zeitreihe werden zwei Geraden integriert. Die *x-Koordinaten* der Punkte, durch welche die Geraden verlaufen, bilden bei der oberen Geraden jeweils die *Maximalwerte* der ersten und zweiten Hälfte der Zeitreihe, bei der unteren die *Minimalwerte*. Die *y-Koordinaten* der Punkte entsprechen jeweils dem mittleren Wert der beiden *y-Achsenabschnitte/Segmente*.

Abbildung 3.8: Beispiel einer Zeitreihe mit abnehmender Varianz im Zeitverlauf, grafisch dargestellt mit der Methode aus dem Referenzwerk, Kapitel 6.4 (Genoud, 2015, S. 149)



3.3.2.5 Einfache Lineare Regression

Zur Analyse der Stärke von Zusammenhängen zwischen zwei oder mehr Variablen sind *Regressionsanalysen* die Methode der Wahl. In *SCE* wird die *bivariate lineare Regression* einerseits zur *Trendanalyse* einer Variablen über die Zeit (1e) und andererseits zur *Zusammenhangsanalyse* zweier Variablen (2b) verwendet. Das *einfache lineare Regressionsmodell* ist das einzige parametrische Verfahren in *SCE*.

Das Modell trifft mehrere Annahmen: die Linearität des Zusammenhangs, eine zufällige Stichprobe, die Stichprobenvariation der unabhängigen Variablen, einen bedingten Erwartungswert und eine nicht signifikant abweichende Streuung (*Homoskedastizität*) (vgl. Wooldridge, 2012, S. 45 ff.).

Da bei Zeitreihen, die im Zeitverlauf gemessenen Werte meist voneinander abhängig sind (Autokorrelation), kann die Annahme der Zufälligkeit der Reihe nur selten erfüllt werden. Zur Beurteilung dieser Problematik in *SCE* kann die Durchführung des *Dufour-Tests* (siehe Abschnitt 3.3.2.3) speziell mit *Lag* = 1 sinnvoll sein.

Auch die Annahme der Homoskedastizität, als Voraussetzung effizienter Schätzer und konsistenter Standardfehler der Koeffizienten, wird bei Zeitreihen häufig verletzt. Dies ist jedoch weniger problematisch, da für diesen Fall *heteroskedastie-robuste Standardfehler* verwendet werden können. Dennoch ist es sinnvoll, vor einer *Regressionsanalyse* die Zeitreihe mit den grafischen Hilfsmitteln in *SCE* sowie ggf. dem *Rekordbrechertest* (siehe oben) auf eine konstante Varianz zu prüfen.

Eine weitere Einschränkung dieser Methode ist, dass die Koeffizienten nur interpretiert werden dürfen, wenn die *Skalenstufen* der unabhängigen und abhängigen Variablen äquidistant sind, was jedoch eine Kardinalskala voraussetzt. Somit ist auch bei der Regression einer Variablen gegen die Zeit darauf zu achten, dass die Messungen in gleichen Zeitabständen erfolgt sind. Bei den in diesem Kontext üblichen Ordinalskalen dürfen nur die Vorzeichen der Koeffizienten analysiert werden.

Aufgrund dieser oft nicht erfüllbaren Voraussetzungen der methodisch korrekten Anwendung ist die einfache lineare Regression als ergänzende Funktionalität von *SCE* zu verstehen, die in diesem Kontext nicht der Absicherung von Evidenz, sondern vielmehr der *Exploration* dienen soll und auch entsprechend – nicht zu prominent – in der grafischen Benutzeroberfläche (*GUI*) platziert ist. So kann sich der Benutzer – wenn nötig – eine *Idee* von der Stärke des Zusammenhangs zwischen zwei *intervall- oder verhältnisskalierten* Variablen machen.

3.3.2.6 Weitere Möglichkeiten

Die Implementierung der *Multivariaten Linearen Regression* mit *Dummykodierung* zur Analyse von Zusammenhängen zwischen *zwei* Variablen wurde ebenfalls in Erwägung gezogen. Mit der Methode ließe sich die Problematik der fehlenden Äquidistanz ordinaler Variablen umgehen, indem für jeden Rang eine *Dummyvariable* definiert wird. Dadurch wären die Koeffizienten auch bei der Regression ordinaler Variablen interpretierbar. Gegenüber der Regression *ohne* Dummykodierung hat diese Methode jedoch den Nachteil, dass für die „*eigentliche*“ *abhängige Variable* nicht mehr nur *einer* sondern *mehrere* Koeffizienten (für jeden Rang einer) interpretiert werden müssen, was die Interpretation für Nutzer mit weniger fundiertem statistischem Grundwissen schwieriger macht. Dies war letztendlich der Grund, weshalb nur die *einfache lineare Regression* implementiert wurde, auch wenn sich diese eigentlich nicht für die in diesem Anwendungskontext vorherrschenden ordinalen Daten eignet.

3.3.3 Inferenzstatistische Methoden für binärskalierte Variablen

3.3.3.1 Mann-Whitney U-Test / Wilcoxon-Rangsummentest

Der *Mann-Whitney U-Test* (auch *Wilcoxon-Rangsummentest* genannt) – als *non-parametrisches* Äquivalent zum *t-Test* – analysiert, ob *eine* von *zwei* Reihen dazu tendiert, höhere Werte als die andere anzunehmen. In der hier angepassten Form – nach der Methode von Meyer-Bahlburg (1969) – kann der Test auch prüfen, ob sich der gemessene Parameter einer *Merkmalsalternative* im Verlauf der Zeit einem bestimmten *monotonen Trend* folgend verändert (3a): Es werden für beide *Variablenausprägungen* (0/1) die jeweiligen Summen der zugehörigen Werte der Zeitachse (t_i) gebildet, welche anschließend mit dem *U-Test* geprüft werden (vgl. Bortz et al., 2008, S. 561).

Zudem kommt der *U-Test* im ebenfalls in *SCE* implementierten *Dufour-Test* (1c) zur Anwendung, wo er dafür verwendet wird, anhand der Positionen im Zeitverlauf (t_i) der positiven bzw. negativen *Lag-Werte* deren Verteilung über die Zeitachse zufallskritisch zu bewerten (siehe Abschnitt 3.3.2.3).

Weiter wird der Test in angepasster Form zur *Zusammenhangsanalyse* für die Kombination einer ordinalen und einer binären Variablen verwendet (5a) (vgl. auch Bortz et al., 2008, S. 200 ff.). Hier werden nicht – wie in der Beschreibung oben – die zugehörigen Werte der Zeitachse, sondern jene der ordinalen Variablen zur Bildung der Summen verwendet. Damit lässt sich anschließend prüfen, ob das Niveau der Werte der Ordinalskala mit der jeweiligen Ausprägung der Binärvariable zusammenhängt.

Der vielseitig einsetzbare *U-Test* wird somit gleich drei Mal in *SCE* verwendet: zur *Trendanalyse* (3a), zur *Zusammenhangsanalyse* (5a) oder als Teil des *Dufour-Tests* (1c).

3.3.3.2 Phi-Koeffizient

Der *Phi-Koeffizient* ist das Äquivalent des *Pearson Korrelationskoeffizienten* für dichotome Variablen und somit für die *Analyse eines Zusammenhangs* zwischen zwei Variablen geeignet (4a). Der Hauptunterschied besteht darin, dass bei der zufallskritischen Bewertung des *Phi-Koeffizienten* ein *Chi-Quadrat-Test* mit einer *Kontinuitätskorrektur* nach Yates (Yates, 1934) angewendet und somit vorsichtiger getestet wird. Bei kürzeren Zeitreihen fällt diese zum Teil „zu“ vorsichtig aus (vgl. Sheskin, 2003, S. 523) und kann zu erheblich größeren *p-Werten* führen. Da in *SCE* die vorsichtige Bewertung allgemein im Vordergrund steht, ist die relativ starke Anpassung im vorliegenden Anwendungskontext eher von Vorteil und das erhöhte Risiko eines *Fehlers 2. Art* wird in Kauf genommen.

Auch hier besteht wie bei *Rho* und *Tau* die Problematik, dass bei einem Trend in der interessierenden Variablen der Zusammenhang *über-* oder *unterschätzt* werden kann.

3.3.3.3 Weitere Möglichkeiten

Zur Analyse der *Zusammenhangsstärke* zwischen einer binären und einer intervall- oder verhältnis-skalierten Variablen oder zwischen zwei binären Variablen wurden auch *Logit-Modelle* in Erwägung gezogen. Auch für eine *Trendanalyse* einer binären Variablen würde sich dieses Verfahren eignen, wobei man diese gegen t_i regrediert (vorausgesetzt die Abstände zwischen den t_i sind äquidistant). Bei ihrer Anwendung auf zwei binäre Variablen, bestünde der Vorteil, dass keine Verteilungsannahmen zur Anwendung notwendig wären. Damit hätte man neben dem *Wilcoxon Rangsummentest* und dem *Phi-Koeffizienten*, die bereits für diese Zusammenhangsanalyse verwendet werden, zudem ein Maß für die Stärke des Zusammenhangs. Aufgrund der Nichtlinearität des Zusammenhangs ist die Interpretation der Koeffizienten jedoch weniger leicht zugänglich als beim *linearen Modell*. Die Methode wurde deshalb aus Gründen der Benutzerfreundlichkeit (noch) nicht integriert.

4 Das Tool *Single-Case Expert*

4.1 Programmierung von *Single-Case Expert*

4.1.1 Ziel der Entwicklung

Mit der Entwicklung dieser Webapplikation soll für die Einzelfallanalyse ein neuer Zugang geschaffen werden – vor allem in der Praxis und der angewandten Forschung. Die Applikation stellt verschiedene deskriptive und inferenzstatistische verteilungsfreie Methoden zusammen und erlaubt den Benutzern, die von ihnen gewonnenen Einzelfalldaten „geführt“ auszuwerten. Die Verwendung non-parametrischer Methoden ist zwar nicht in allen Fällen unproblematisch (siehe Abschnitt 2.3), in diesem Kontext aber am besten geeignet, da sie weniger Anforderungen an die Daten stellen und deshalb einfacher in einer Applikation verwendbar sind, die „universell“ und einfach einsetzbar sein möchte.

Die aus Einzelfallanalysen generierte Evidenz, die erhobenen Daten, haben auch für die Wissenschaft ein großes Potential. Diese Daten zu sammeln ist jedoch entsprechend aufwändig und erklärt zumindest teilweise das bisher begrenzte Interesse in der Wissenschaft. Unter Verwendung entsprechender informationstechnologischer Hilfsmittel könnten diese Kosten erheblich reduziert werden, gerade mit den Entwicklungen der letzten Jahre, die es vereinfachen, übers Internet – auch komplexe – Applikationen zur Verfügung zu stellen. Längerfristig hat *SCE* neben der erleichterten Auswertung somit auch zum Ziel, einen neuartigen Ansatz von bidirektionalem Wissensaustausch zu ermöglichen: Experten machen den Praktikern die methodischen Werkzeuge mit den nötigen Hilfestellungen zugänglich und die Praktiker stellen im Gegenzug Daten – insbesondere von Einzelfällen, wenn möglich mit zusätzlichen Fallinformationen – zur Verfügung. Diese Funktionalität ist in *SCE* wegen ihres Entwicklungsaufwands noch nicht integriert worden und würde erst bei entsprechender Nachfrage hinzugefügt.

4.1.2 Aktueller Stand der Entwicklung

Die mit diesem Text erschienene *Release Version* (v1.0) von *SCE* mit gut 1400 Zeilen Code (wobei gut ein Drittel davon Textinhalte der *GUI* sind) stellt eine erste Version dar. Sie vereint eine Auswahl von einfach interpretierbaren deskriptiv-statistischen (siehe Abschnitt 3.2) und inferenzstatistischen (siehe Abschnitt 3.3) verteilungsfreien Methoden, die sich speziell für die Auswertung von Zeitreihen aus Einzelfallanalysen eignen. Nach einschlägigen Vortests liegt sie in einer für die Nutzung hinreichend zuverlässigen Version vor.

Version 1.0 baut auf einer im Vorfeld entwickelten Beta-Version der Anwendung auf, in welcher zur Prüfung der technischen Machbarkeit das Grundgerüst der Applikation und der *GUI* erstellt wurde. Diese enthielt erst einige deskriptiv-statistische Methoden, den S_2 -Test, sowie die Rangkorrelationskoeffizienten *Rho* und *Tau*. Die aktuelle Version bietet grafische *Outputs* (Ausgaben) der Daten, mehrere deskriptiv-statistische Maßzahlen und acht Inferenzstatistiken, von denen sich zwei für jeweils zwei Fragestellungen eignen (siehe Abschnitt 3.3.1). Jede Funktionalität verfügt über einige Erläuterungen und Hilfestellungen. Unter der *Hilfe*-Seite finden sich die allgemeinen Hinweise zur Nutzung von *SCE*.

Die Verfahren eignen sich für verschiedene Designs der Einzelfallanalyse und erlauben die Untersuchung der gewonnenen Zeitreihe auf ihre Zufälligkeit, auf zyklische Variation und auf Trends in der zentralen Tendenz und der Dispersion, sowie auf Zusammenhänge zwischen zwei Zeitreihen. Bei der Auswahl wurde darauf geachtet, dass die Verfahren genügend robust und gleichzeitig einfach zu interpretieren sind. Konnte in dieser Hinsicht nicht der erwünschte Level erreicht werden, wurden die Verfahren entweder nicht integriert oder es wurden Erläuterungen zu den Anwendungsschwierigkeiten

oder den problematischen Aspekten eingefügt. In der aktuellen Version ist mit dem *Rekordbrechertest* ein Verfahren integriert, welches hinsichtlich der Robustheit deutlich unter dem Niveau der anderen Methoden liegt. Der Test wurde dennoch integriert, damit die Varianzveränderung zumindest *explorativ* untersucht werden kann

Die jetzige funktionierende Version ist als Ausgangspunkt für Weiterentwicklungen, Modifikationen und ggf. Adaptierungen an spezifische Anwendungsbereiche zu verstehen (siehe Abschnitt 4.1.9 zur weiterführenden Entwicklung).

4.1.3 Kriterien der Software-Entwicklung

Zentral für die Entwicklung war die Frage, ob es möglich sein würde, ein Werkzeug für die Einzelfallanalyse zu schaffen, das die Kriterien der Verfügbarkeit, Bedienbarkeit, der einfachen und angemessenen Interpretierbarkeit und Vernetzung ausreichend erfüllt – vor allem hinsichtlich der Ausrichtung auf die Praxis. Dafür waren folgende Grundsatzfragen zu klären:

- Welche Programmiersprache ist geeignet?
- Wie und mit welchen Mitteln soll die *GUI* realisiert werden?
- Ist es möglich eine reaktive bzw. hinreichend klar strukturierte *GUI* zu bauen?
- Können die Daten übersichtlich in Grafiken und Tabellen dargestellt werden?
- Welche Funktionalitäten, wie beispielsweise die verwendeten Tests, sind schon (in sog. *Packages*) vorhanden?
- Kann die Software auf einfache Weise für die Nutzung zugänglich gemacht werden?

Abgesehen von den ersten beiden Fragen geht es um Optimierungsprobleme: Welche Lösungen sind im Hinblick auf die Entwicklung und die Verwendung am *effizientesten*?

4.1.4 Gründe für die Entwicklung mit *R* und *Shiny*

R ist eine weit verbreitete Programmiersprache im Bereich der Statistik, womit sie sich für die Entwicklung von *SCE* anbot. Ob sie auch die Anforderungen an die *GUI* erfüllen könnte, war zu Beginn unklar – Recherchen führten schließlich zum Ergebnis, dass das relativ neue *R-Package* „*shiny*“ die meisten Kriterien für die *GUI* erfüllen kann. *Shiny* ermöglicht zum Beispiel die Anordnung und Formattierung der Outputs, die Definition sog. *reaktiver (GUI-)Elemente*, die Verarbeitung von Dateneingaben (*Inputs*) in Echtzeit, die gleichzeitige Datenausgabe (wie z. B. die Teststatistiken) in der *GUI* und die Veröffentlichung als Webapplikation.

4.1.5 Prüfen der technischen Möglichkeiten von *R* und *Shiny*

Zunächst wurde recherchiert, welche Tests bereits in *R-Packages* implementiert sind und direkt in das Programm integriert werden können. Es zeigte sich, dass einige der benötigten Verfahren (siehe Abschnitt 4.1.7) – erwartungsgemäß – bereits vorhanden sind und daher nur wenige Tests (z. B. der *Rekordbrechertest*) eigens für *SCE* zu programmieren waren.

Anschließend war zu klären, wie die Zeitreihendaten einfach eingegeben werden können, da in den Standard-Funktionalitäten von *shiny* keine Möglichkeit gefunden wurde, die das Einfügen per „Copy-Paste“ ermöglicht. Letztendlich stellte sich der *Ace Editor*, der eigentlich ein Text-Editor für die

Integration auf Webseiten ist (und meist für die Darstellung von Programmcode verwendet wird), aber auch in *shiny* funktioniert, als gute Lösung heraus: Die aus einem Tabellenprogramm eingefügten Daten können ausreichend übersichtlich dargestellt werden und sind vom Programmcode einfach abrufbar.

Wichtig war weiter die Prüfung der Möglichkeiten, die *shiny* für die *GUI* bietet. Essentiell war das einfache Anzeigen und Ausblenden bestimmter Informationen (Testinformationen, Hilfestellungen) durch die Interaktion mit der *GUI*. „*Collapses*“ (aus dem Package *shinyBS*), auch „*Accordions*“ genannt, erwiesen sich als beste Methode, da der Benutzer so immer sieht, an welcher Stelle zusätzliche Informationen vorhanden sind, diese aber nicht anzeigen muss. Zudem war in Bezug auf die *GUI* zu klären, ob es möglich ist, eine dynamische, von der Dateneingabe abhängige Ausgabe zu definieren, sodass jeweils nur die zu den Daten passenden Verfahren dargestellt werden. Es zeigte sich, dass in den Standardfunktionen vom Package *shiny* mit den sog. „*ConditionalPanels*“ bereits eine gute Lösung vorhanden ist.

Mit der Klärung dieser technischen „Machbarkeitsfragen“ waren die wichtigsten Voraussetzungen für den Beginn der darauffolgenden *konkreten* Entwicklung von *SCE* erfüllt.

4.1.6 Ablauf der Entwicklung

Die Entwicklung lässt sich grob in die nachfolgende Sequenz von Arbeitsschritten aufteilen, wobei einige Aktivitäten, wie die *globale Definition von mehrfach verwendeten Funktionen* (Punkt 11), im Grunde während des ganzen Entwicklungsprozesses eine Rolle spielten:

- (1) Definieren des *GUI*-Grundgerüsts („Datenanalyse“, „Hilfe“, „Über“)
- (2) Integration des *Ace Editors*
- (3) Definieren einer Methode zum Auslesen und Auf-/Vorbereiten der Daten für die jeweiligen Analysefunktionen und Testverfahren
- (4) Implementieren der grafischen Ausgaben (Plots mit Trendlinien)
- (5) Integrieren erster Testverfahren zur Funktionsprüfung der Dateneingabe/-vorbereitung
- (6) Erste Tests mit verschiedenen Beispieldaten*
- (7) Implementieren der Funktion zur Ermittlung der Anzahl eingegebener Variablen und ihrem Skalenniveau
- (8) Vorbereiten und Anordnen der *GUI*-Elemente (hauptsächlich *Collapses*, siehe oben) für die noch zu integrierenden Funktionalitäten und Verfahren
- (9) Definieren der reaktiven Anzeige auf Basis des Datentyps (*ConditionalPanels*, siehe oben).
- (10) Implementieren der restlichen Funktionen und Verfahren
- (11) Globale Definition von mehrfach verwendeten Funktionen
- (12) Einfügen der Hilfestellungen und Zusatzinformationen
- (13) Erstellen der *Hilfe*-Seite zur Nutzung des Programms
- (14) Ausführliche Tests mit verschiedenen Beispieldaten*
- (15) Veröffentlichung der Version 1.0

* Die Verfahren wurden in *R* mit verschiedenen Beispieldaten – von verschiedenen Usern – getestet, um sicherzustellen, dass das Einlesen und Verarbeiten der Daten sowie das Erstellen der Ausgabe bzw. Berechnen der Ergebnisse korrekt funktionierten.

4.1.7 Verwendete *R-Packages*

Für die Entwicklung wurden verschiedene Funktionen aus diversen *R-Packages* genutzt. Wie weiter oben bereits erwähnt war das Package *shiny* zum Erstellen der *GUI* essentiell. Als Ergänzung dazu wurde *shinyAce* verwendet, welches den *Ace Editor* zur Verfügung stellt (worüber in *SCE* jeweils die Daten eingegeben werden) und *shinyBS*, welches verschiedene Elemente für die *GUI* wie Buttons und Warnungen, sowie die in *SCE* genutzten „*Collapses*“ enthält. Weiter werden die *Plot-Funktionen* zur grafischen Darstellung aus *ggplot2* eingesetzt.

Die Tests, die nicht programmiert werden mussten, stammten aus den Packages *stats*, *psych* und *randtests*. Ersteres enthält den *U-Test* (*wilcox.test*), *Rho* und *Tau* (*cor.test*), sowie den *Chi-Quadrat-Test* (*chisq.test*), welcher für die zufallskritische Überprüfung des *Phi-Koeffizienten* benötigt wird. Aus *psych* wurden *describe* zur einfachen Auflistung der statistischen Maßzahlen und *phi* zur Berechnung des *Phi-Koeffizienten* integriert. Aus *randtests* wurde der *S₂-Test* verwendet (*cox.stuart.test*).

Wie in Abschnitt 3.3.2.3 erwähnt, konnten der *Dufour-Test* und der *Rekordbrechertest* (noch) in keinem offiziellen *R-Package* gefunden werden und wurden deshalb eigens für *SCE* programmiert.

4.1.8 Herausforderungen der Entwicklung

Da im Sinne der *Wiederverwendbarkeit* die meisten wichtigen „Funktionen“ (Programmteil mit spezifischer Aufgabe, in der Regel mit einem *Input* und einem *Output*) *global* definiert wurden, mussten diese so programmiert werden, dass sie mit mehreren – meist mit allen fünf – Dateneingaben (siehe Abschnitt 4.2.1) umgehen können (z. B. die *Plot-Outputs*), was vor allem zu Beginn der Entwicklung eine Schwierigkeit darstellte. Auch wenn die globale Definition der Funktionen zum Teil zu erhöhter Komplexität führt, hat sie einerseits den Vorteil, dass der Code erheblich kürzer wird, und erleichtert andererseits die Modifizierbarkeit, da eine bestimmte Anpassung jeweils nur an einer Stelle vorgenommen werden muss.

Eine weitere Schwierigkeit bestand darin, dass viele neue Funktionen nicht ohne Weiteres direkt in der *shiny*-Applikation getestet werden konnten, da in diesem Fall keine detaillierten Fehler-Informationen ausgegeben werden. Die Funktionen mussten deshalb meist in angepasster Form – ohne die Bezüge zu den anderen Programmteilen und ohne Verknüpfung mit der *GUI* – in einem separaten *R Script* getestet werden, wodurch das Testen und die Fehlersuche im Vergleich zu anderen Programmiersprachen (wie z. B. *Java*) zum Teil relativ aufwändig war.

Da jede Teststatistik ihre eigenen Charakteristiken und Ausgaben (z. B. die jeweiligen Koeffizienten) hat, musste immer darauf geachtet werden, dass die Oberfläche möglichst einheitlich und verständlich bleibt. Z. B. musste der *Dufour-Test*, der das Testen auf unterschiedliche zyklische Abhängigkeiten erlaubt, über eine einfach bedienbare Eingabe-Möglichkeit zum Auswählen der gewünschten *Lags* verfügen. Hierfür stellte sich ein sog. „*Slider-Input*“ als sinnvolle Variante heraus, da die Auswahl mit der Maus sehr einfach, und zugleich eine Begrenzung der zulässigen Werte möglich ist (das Prüfen sehr langer *Lags* ist meist nicht sinnvoll, siehe Abschnitt 3.3.2.3). Bei den Ausgaben war allgemein darauf zu achten, dass wirklich nur die Werte ausgegeben werden, die zwingend für die Interpretation benötigt werden, damit die *GUI* übersichtlich bleibt. Beispielsweise werden bei der *Regressionsanalyse* nur die beiden Koeffizienten und deren *p-Werte* angezeigt. Auf das Anzeigen weiterer Informationen zu den Standardfehlern, zur Anpassungsgüte (R^2) oder zur *F-Statistik* wurde verzichtet. Bei der Anwendung solcher Modelle sind dies sonst wichtige Informationen, da in *SCE* aber nur das *einfache lineare Modell* zur Verfügung steht, und der Benutzer ohnehin nicht mehrere Modelle anwenden kann, wurden diese nicht in der Ausgabe integriert.

Dieses Abwägen der *Einfachheit* gegen die *methodische Vollständigkeit und Korrektheit* „liegt in der Natur“ eines Tools wie *SCE*, das einen *angemessenen methodischen Zugriff* verbunden mit Benutzerfreundlichkeit gewährleisten möchte. Dabei müssen auch immer wieder Kompromisse eingegangen werden.

4.1.9 Weiterführende Entwicklung

In einer künftigen Version von *SCE* ist die Integration einer untergeordneten Seite vorgesehen, die dem Nutzer bei der Versuchsplanung helfen soll („Design-Guide“): Durch Auswählen verschiedener Eigenschaften (beobachtend vs. (quasi-)experimentell, Art und Zahl zu messender Merkmale, Zeitfenster, vorgesehene Replikation etc.) werden Vorschläge für mögliche Versuchsanordnungen gemacht. Dadurch würde der Nutzer nicht nur bei der Auswertung unterstützt, sondern bereits *zuvor*, bei der Wahl eines angemessenen Designs. Damit soll der ganze Prozess der Durchführung einer Einzelfallanalyse unterstützt werden, mit dem Ziel, nicht nur eine angemessene *Datenanalyse*, sondern bereits die methodisch korrekte *Erhebung* der Daten zu fördern. Mithilfe des „Design-Guides“ lassen sich so beide Teile schon in der Planungsphase aufeinander abstimmen.

Für die Weiterentwicklung ist auch das Feedback der Benutzer – sowohl von Wissenschaftlern als auch Praktikern – ausschlaggebend; gerade Anregungen zur Ergänzung und Anpassung von Hilfestellungen oder Ideen für weitere Testverfahren sind von Interesse.

Auch unabhängig vom Feedback der Benutzer sind bereits diverse Optimierungsideen vorhanden. Zunächst sollte der *Rekordbrechertest* durch die Integration einer Gewichtung der Rekorde und einer Trendbereinigung (siehe Abschnitt 3.2.2.4) verbessert werden. Die Option einer Trendbereinigung der Zeitreihen bei der Durchführung der Zusammenhangsanalyse mit *Rho* und *Tau* (siehe Abschnitt 3.3.2.1) wäre ebenfalls eine interessante Zusatzfunktion, wobei dies auch in *Excel* relativ einfach durchgeführt werden kann (siehe Abbildung 5.5). Des Weiteren sind zusätzliche deskriptive Methoden vorgesehen, z. B. *Histogramme* bei größeren *N*, damit visuell beurteilt werden kann, ob eine Verteilungsannahme sinnvoll ist.

Als Teil der *Hilfe-Seite* wäre zudem die Integration von Fallbeispielen interessant (siehe 5. Kapitel), damit sich der Benutzer noch leichter ein Bild davon machen kann, *wie* und *wofür* die Verfahren einsetzbar, und wie sie zu interpretieren sind.

Wie eingangs dieses Kapitels erwähnt, wird *SCE* wegen des Entwicklungsaufwands in dieser Version noch nicht über eine Datenbank verfügen, die notwendig wäre, um die generierten Daten entsprechend strukturiert für interessierte Benutzer (z. B. Wissenschaftler) abzuspeichern. Auf längere Sicht besteht jedoch auch diese Möglichkeit der Weiterentwicklung.

Des Weiteren haben nicht alle Benutzer der Applikation das gleiche statistische Grundwissen. Dementsprechend wäre es unter Umständen sinnvoll, zwei verschiedene Benutzungsmodi einzubauen, wovon jeweils beim Programmstart der passende ausgewählt würde. Entsprechend würden im „*Fortgeschrittenen-Modus*“ mehr Informationen zu den Tests (z. B. die *Anpassungsgüte* bei der *Regressionsanalyse*) sowie die komplizierteren Verfahren (z. B. das *Logit-Modell*) und gleichzeitig weniger – für den „Experten“ nicht notwendige – Hilfestellungen angezeigt als im „*normalen Modus*“.

Denkbar sind auch Adaptierungen der Software an spezifische Problemstellungen oder konkrete praktische Anwendungen.

Was sich von diesen Ideen – vor allem von letzteren – in Zukunft realisieren lässt, ist davon abhängig, *ob* und *wie* bzw. *wo* *SCE* eingesetzt werden kann. Zweifellos sind bereits zu diesem Zeitpunkt die Ideen zu möglichen Anwendungen und Perspektiven der Weiterentwicklung vielseitig.