Theorie Psychometrischer Verfahren

Prof. Dr. Dirk Ostwald

15. Juni 2025

Inhaltsverzeichnis

1 Psychologische Tests

Tests und Fragebögen sind in der alltäglichen Erfahrung unterschiedlicher Natur. Mit dem Begriff *Test* verbindet man vermutlich am ehesten kleinere Prüfungen, bei denen es korrekte Antworten gibt und eine Leistungsfähigkeit überprüft wird. Mit dem Begriff des *Fragebogens* verbindet man vermutlich mehr oder weniger interessante Sammlungen von Fragen, bei denen man eventuell zu einem wahrheitsgemäßen Antwortverhalten aufgefordert wird, bei denen es aber auch klar ist, dass es keine richtigen oder falschen Antworten gibt. Obwohl mit Tests und Fragebögen in der alltäglichen Erfahrung also durchaus unterschiedliche Verfahren verbunden sind, unterscheidet die psychologische Testtheorie zwischen diesen nicht grundlegend. Historisch ist die Analyse von Testdaten sicherlich zunächst durch die Analyse von Leistungstestdaten geprägt, allerdings werden heutzutage die dort entwickelten Konzepte der Testtheorie auf die gleiche Art und Weise zur Analyse von klinischen Fragebogendaten genutzt. In diesem Abschnitt wollen wir zunächst einen kurzen Überblick zum Begriff des *Psychologischen Tests* geben, die sogenannten *Testgütekriterien* auflisten, einige gängige *Normierungen* von Testrohdaten vorstellen und schließlich kurz die drei wesentlichen *Testtheorien* skizzieren.

1.1 Psychologische Tests

Um uns dem Testbegriff zu nähern, betrachten wir zunächst die Definitionen eines Tests nach Moosbrugger & Kelava (2012), Bühner (2010) und Krauth (1995).

Nach Moosbrugger & Kelava (2012) ist ein psychologischer Test "ein wissenschaftliches Routineverfahren zur Erfassung eines oder mehrerer empirisch abgrenzbarer psychologischer Merkmale mit dem Ziel einer möglichst genauen quantitativen Aussage über den Grad der individuellen Merkmalsausprägung".

Nach Bühner (2010) ist ein "psychometrischer Test ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale (vgl. Lienert & Raatz (1998), S.1). Das Ziel eines psychometrischen Tests besteht darin, die absolute oder relative Ausprägung einer Eigenschaft, einer Fähigkeit oder eines Zustands bei einer oder mehreren Personen zu messen oder eine qualitative Aussage zu treffen, welcher Personenklasse Personen zugeordnet werden können (vgl. Rost (2004)). Psychometrische Tests sind nach der Klassischen oder Probabilistischen Testtheorie entwickelt, sind theoretisch fundiert und genügen genau definierten Gütekriterien (Haupt- und Nebengütekriterien)."

Nach Krauth (1995) schließlich "besteht ein psychologischer Test aus einer Menge von Reizen mit den zugehörigen zugelassenen Reaktionen, also aus manifesten Variablen. Eine Skala ordnet den Reaktionsmustern der manifesten Variablen die Ausprägungen einer oder mehrerer latenter Variablen zu. Somit fungiert ein Test als Messinstrument zur Erfassung nicht direkt beobachtbarer (latenter) Variablen, deren Existenz für Personen und gelegentlich auch für Tiere postuliert wird."

Nach diesen Definitionen ist psychologischen Tests also gemein, dass sie Verfahren zur Messung psychologischer Eigenschaften darstellen und sich dadurch auszeichnen, dass sie einer wissenschaftlichen Qualitätssicherung unterliegen.

Generell lassen sich psychologische Tests in verschiedene Kategorien unterteilen:

- *Leistungstests.* Dazu gehören Entwicklungstests, Intelligenztests, allgemeine Leistungstests, Schultests sowie spezielle Funktionsprüfungs- und Eignungstests.
- *Psychometrische Persönlichkeitstests.* Hierunter fallen klinische Tests, Persönlichkeitsstrukturtests, Einstellungstests und Interessentests.
- *Persönlichkeitsentfaltungs-Verfahren.* Diese umfassen Formdeuteverfahren, verbalthematische Verfahren sowie zeichnerische und gestalterische Verfahren.

Im Kontext der Klinischen Psychologie und Psychotherapie sind sicherlich klinische Tests von vorherrschendem Interesse. Beispiele für klinische Tests sind etwa

- *BDI-II* (Beck Depressions-Inventar II, Beck et al. (1996), Hautzinger et al. (2006)), ein klinischer Test zur Messung der Schwere einer Depression. Er basiert auf den diagnostischen Kriterien einer Major Depression nach dem DSM-IV und fragt typische depressive Symptome ab.
- *BSI* (Brief Symptom Inventory, Derogatis & Melisaratos (1983)), ein klinischer Test, der psychische Belastungen und Symptome erfasst. Der BSI deckt neun Symptomdimensionen ab: Somatisierung, Zwanghaftigkeit, Unsicherheit im Sozialkontakt, Depressivität, Ängstlichkeit, Aggressivität/Feindseligkeit, phobische Angst, paranoides Denken und Psychotizismus.
- *BSL-23* (Borderline-Symptomliste 23, Bohus et al. (2009)), ein klinischer Test zur Erfassung der Schwere von Borderline-spezifischen Symptomen. Er misst typische emotionale, kognitive und Verhaltensmuster bei Borderline-Persönlichkeitsstörung.

Psychologische Test bestehen dabei in der Regel aus *Items*, die als Grundbausteine des Tests fungieren. Items stellen allgemein Reize dar, auf die eine Reaktion erwartet und registriert wird. Das Standardbeispiel für ein Item ist eine Frage mit entsprechenden Antwortmöglichkeiten. Beispielsweise umfasst der BDI-II 21 Items zur Bewertung des Schweregrads einer Depression auf einer Skala von 0 bis 3. Das Item zum Thema *Traurigkeit* dabei lautet:

Suchen Sie eine Aussage heraus, die am besten beschreibt, wie Sie sich in den letzten zwei Wochen, einschließelich heute, gefühlt haben:

(0) Ich bin nicht traurig.

(1) Ich bin oft traurig.

(2) Ich bin ständig traurig.

(3) Ich bin so traurig oder unglücklich, dass ich es nicht aushalte.

1.2 Testgütekriterien

Im wissenschaftlichen Diskurs haben sich eine Reihe von Qualitätsanforderungen an psychologische Tests herausgebildet, die ein Test, um wissenschaftlich anerkannt zu sein, idealerweise erfüllt (vgl. Moosbrugger & Kelava (2012)). Grundlegend sind die Kriterien der Objektivität, Reliabilität und Validität

1. *Objektivität.* Ein Test ist objektiv, wenn seine Durchführung, Auswertung und Interpretation unabhängig von den Testleitenden ist. Die Objektivität eines Testverfahrens wird meist durch eine entsprechende Manualisierung sichergestellt.
2. *Reliabilität.* Ein Test ist reliabel, wenn er das zu messende Merkmal exakt und ohne Messfehler erfasst. Wichtige Unterarten der Reliabilität sind Retest-Reliabilität, Paralleltest-Reliabilität, Testhalbierungs-Reliabilität und interne Konsistenz. Reliabilitäten werden in der Regel durch Korrelationen erfasst. Die Klassische Testtheorie stellt einen theoretischen Überbau zur Einordnung und quantitativen Definition von Retest-, Paralleltest, und Testhalbierungsreliabilität dar und entwickelt beispielsweise mit Cronbach's α (Cronbach (1951)) Maße für die interne Konsistenz von psychologischen Tests.
3. *Validität.* Ein Test ist valide, wenn er tatsächlich das misst, was er messen soll. Unterschieden werden Inhaltsvalidität, Augenscheinvalidität, Kriteriumsvalidität, Konstruktvalidität und, im Kontext der Testentwicklung faktorielle Validität. Inhalts- und Augenscheinvalidität werden in der Regel durch Expert:inneneinschätzungen belegt, Kriteriumsvalidität durch Korrelation mit einem geeigneten externen Kriterium, z.B. einer klinischen Diagnose. Die Etablierung der Konstruktvalidität eines Tests ist eine weitreichende Forderung, die vermutlich selten erfüllt wird (vgl. Borsboom et al. (2004)). Belege für die die faktorielle Validät schließlich tauchen häufig in Textmanualen auf und beziehen sich auf die Ergebnisse von Faktorenanalysen.

Neben den obigen Gütekriterien erfüllt ein wissenschaftlich anerkanntes Testverfahren idealerweise auch noch folgende sogenannten *Nebengütekriterien*:

4. *Skalierung.* Die Testwerte sollten die qualitativen Merkmalsrelationen im Sinne der Repräsentationstheorie des Messen adäquat abbilden.
5. *Normierung.* Das Bezugssystem ermöglicht die Vergleichbarkeit der Testwerte mit einer Eichstichprobe.
6. *Testökonomie.* Ein Test sollte mit möglichst geringem Zeit- und Ressourcenaufwand auswertbar sein. Die Digitalisierung der Antworterhebung und automatisierte Auswertung mithilfe von Computersoftware wären ein erster Schritt in dieser Richtung.
7. *Nützlichkeit.* Der Test sollte praktische Relevanz besitzen und einen positiven Nutzen bieten.
8. *Zumutbarkeit.* Die Belastung der Testperson sollte in einem angemessenen Verhältnis zum Nutzen stehen.
9. *Unverfälschbarkeit.* Das Testergebnis sollte nicht durch bewusstes Verhalten verfälscht werden können. Dies ist bei den meisten klinischen Verfahren, die auf einer Selbsteinschätzung beruhen, sicherlich nicht der Fall.
10. *Fairness.* Der Test darf keine systematische Benachteiligung bestimmter Gruppen verursachen. Dieses Kriterium ist im klinischen Kontext vermutlich eher von untergeordnerter Relevanz.

1.3 Normwerte

Primäres quantitatives Ergebnis eines psychologischen Tests oder auch eines einzelnen Items eines Tests ist zunächst einmal ein *Rohwert*. Beispielsweise sind die möglichen Rohwerte bei Beantwortung des oben aufgeführten Traurigkeitsitems des BDI-II die Zahlen 0,1,2, und 3. Der Rohwert eines Testergebnisses über mehrere Items ergibt sich häufig als die Summe der Rohwerte der einzelnen Items als sogenannter *Summenscore*. So variieren die Summenscores des BDI-II beispielsweise zwischen 0 und 63, wobei Werte zwischen 0 und 13 einer *minimalen*, Werte zwischen 14 und 19 einer *leichten*, Werte zwischen 20 und 28 einer *mäßigen* und Werte zwischen 29 und 63 einer *schweren Depressionssymptomatik* entsprechen. Die Rohwerte des BSI und des BSL-23 ergen sich durch Summation der einzelnen Itemwerte, mit den möglichen Ausprägungen 0,1,2,3,4 und Division durch die beantwortete Itemanzahl, wodurch sich dann ein *Mittelwertscore* im Bereich 0 bis 4 ergibt.

Um ein gewisses Maß an Vergleichbarkeit zwischen den Wertausprägungen verschiedener klinischer Testverfahren zu gewährleisten und ein einzelnes Testergebnis vor dem Hintergrund der üblichen Verteilungen von Testergebnissen einzuordnen, werden Testergebnisse oft *normiert* (vgl. De Beurs et al. (2022)). Dazu ist es zunächst einmal nötig, Testergebnisse von einer hinreichend großen Stichprobe von gesunden oder klinisch auffälligen Proband:innen zu erheben. Diese Testergebnisse seien im Folgenden als Testrohwerte $y^{(i)}$ für $i = 1, ..., n$ bezeichnet, wobei n der Stichprobenumfang ist. Die Resultate solcher groß angelegter Studien sind in der Regel in den entsprechenden Testmanualen dokumentiert. Legt man dann die Annahme zugrunde, dass die Testergebnisse in der entsprechenden Stichprobe einer Normalverteilung folgen, deren Erwartungswertparameter man durch das entsprechende Stichprobenmittel und deren Varianzparameter man durch die entsprechende Stichprobenvarianz schätzen kann, so ergeben sich, basierend auf dem Theorem zur linearen Transformation normalverteilter Zufallsvariablen eine Reihe von Möglichkeiten, die Testrohwerte zu standardisieren. Üblicherweise trifft man in diesem Zusammenhang auf *Z Scores* und *T Scores*, seltener auf *Stanines* oder *Stens*.

Um die Bedeutung dieser standardisierten Testwerte zu verstehen, erinnern wir zunächst an die Verteilung der Wahrscheinlichkeitsmasse einer normalverteilten Zufallsvariable. Bekanntlich gelten (vgl. Abbildung 1.1, nach Seashore (1955) und De Beurs et al. (2022))

$$\begin{aligned}
\int_{-1\sigma}^{1\sigma} N(x;\mu,\sigma^2)\,dx &= \Phi(1\sigma;\mu,\sigma^2) - \Phi(-1\sigma;\mu,\sigma^2) \approx 0.68 \\
\int_{-2\sigma}^{2\sigma} N(x;\mu,\sigma^2)\,dx &= \Phi(2\sigma;\mu,\sigma^2) - \Phi(-2\sigma;\mu,\sigma^2) \approx 0.95 \\
\int_{-3\sigma}^{3\sigma} N(x;\mu,\sigma^2)\,dx &= \Phi(3\sigma;\mu,\sigma^2) - \Phi(-3\sigma;\mu,\sigma^2) \approx 0.99
\end{aligned} \tag{1.1}$$

Innerhalb einer Standardabweichung um den Erwartungswertparameter liegen bei einer normalverteilten Zufallsvariable also 68% der Wahrscheinlichkeitsmasse, innerhalb von zwei Standardabweichungen 95% und innerhalb von drei Standardabweichungen mit 99% fast die gesamte Wahrscheinlichkeitsmasse. Diese Verteilung der Wahrscheinlichkeitsmasse gilt offensichtlich unabhängig von den konkreten Werten von μ und σ^2 und somit für alle normalverteilten Zufallsvariablen. Weiterhin erinnern wir daran, dass das Theorem zur

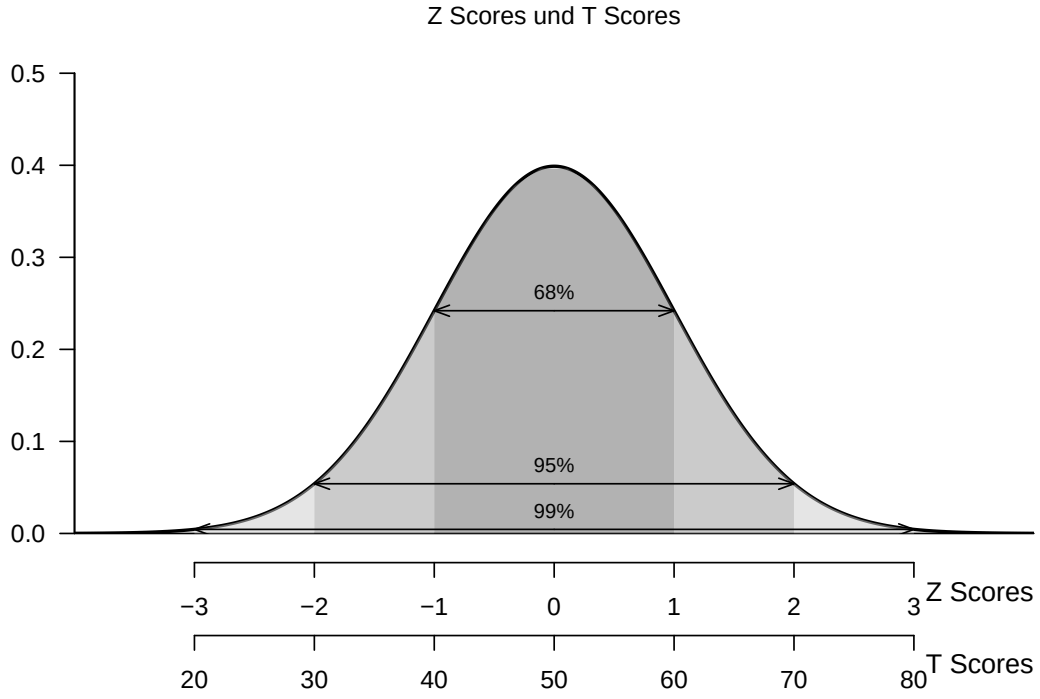

Abbildung 1.1 Z Scores und T Scores.

linearen Transformation einer normalverteilten Zufallsvariable besagt, dass

$$y \sim N(\mu, \sigma^2) \text{ und } z := ay + b \Rightarrow z \sim N(a\mu + b, a^2\sigma^2). \tag{1.2}$$

Seien also m und s^2 die Stichprobenschätzer für den Erwartungswertparameter μ und den Varianzparameter σ^2 normalverteilter Testrohwerte. Dann gilt für

$$z := \frac{y - m}{s} = \frac{1}{s}y - \frac{1}{s}m \approx \frac{1}{\sigma}y - \frac{1}{\sigma}\mu \tag{1.3}$$

näherungsweise

$$z \sim N\left(\frac{1}{\sigma}\mu - \frac{1}{\sigma}\mu, \frac{1}{s^2}s^2\right) = N(0, 1). \tag{1.4}$$

Berechnet man zu jedem Testrohwert $x^{(i)}$ also den sogenannten *Z Score*

$$z^{(i)} := \frac{x^{(i)} - m}{s}, \tag{1.5}$$

so sind die so generierten Werte, unabhängig von den Stichproben-spezifischen Werten von m und s immer näherungsweise standardnormalverteilt und damit über verschiedene Rohwertebereiche verschiedener Tests vergleichbar (vgl. Abbildung 1.2).

Nun hat eine standardnormalverteilte Zufallsvariable die Eigenschaft, mit höchster Wahrscheinlichkeit einen Wert zwischen -3 und 3 anzunehmen. Dies mag manchmal unpraktisch erscheinen, da in diesem Fall ein einzelnes Testergebnis, wenn es unter dem Stichprobenmittelwert liegt, negativ ausfallen kann. Weiterhin erfordert die Angabe eines Ergebnisses

mit hinreichender Genauigkeit die Angabe von Nachkommastellen. McCall (1922) hat deshalb vorgeschlagen, eine weitere lineare Transformation der berechneten Z Scores durchzuführen, die die Z Scores in einen zugänglicheren Wertebereich transformiert. Die so transfomierten Z Scores bezeichnet McCall (1922) zu Ehren von Edward Lee Thorndike, Lewis Terman und Louis Leon Thurstone als *T Scores*. Die Bezeichung *T Scores* geht also keinesfalls auf die T-Statistiken der Frequentisischen Inferenz und ihre entsprechenden Verteilungen zurück. Die von McCall (1922) vorgeschlagene Transformation ist durch

$$t^{(i)} := 10z^{(i)} + 50 \tag{1.6}$$

gegeben. Mit der näherungsweisen Standardnormalverteilung von z ergibt sich entsprechend

$$t \sim N\left(0 + 50, 10^2 \cdot 1\right) = N(50, 100). \tag{1.7}$$

T Scores haben näherungsweise also immer einen Erwartunsgwert von 50, eine Varianz von 100 und eine Standardabweichung von 10. Damit liegen dann im *T Score* Wertebereich von 40 bis 60 etwa 68% der Stichprobenergebnisse, im Wertebereich von 30 bis 70 etwa 95% der Stichprobenergebnisse und im Wertebereich von 20 bis 80 etwa 99% der Stichprobenergebnisse (vgl. Abbildung 1.1).

Die seltener anzutreffenden *Stanine* (standard nine) und *Sten* (standard ten) Werte folgen in ihrer Berechnung der gleichen Logik wie die T Scores, zielen aber auf den Wertebereich 1 bis 9 bzw. 1 bis 10. Sie ergeben sich aus Z Scores durch die Transformation

$$s^{(i)} := 2z^{(i)} + 5 \tag{1.8}$$

und anschließende Rundung auf Werte von 1 bis 9 bzw. 1 bis 10.

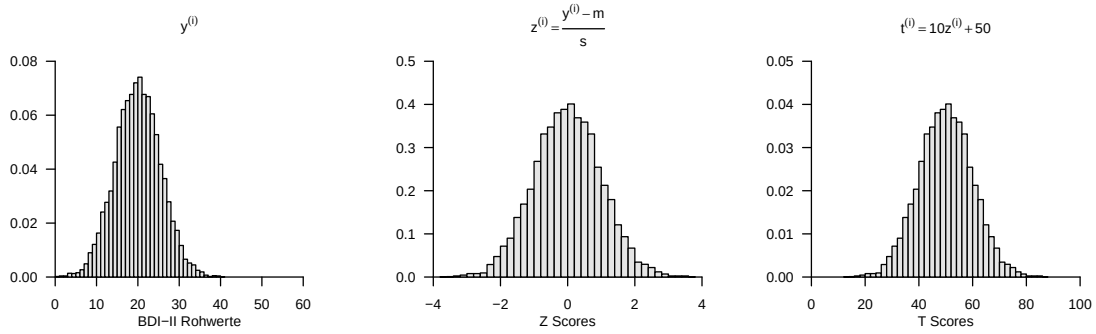

Abbildung 1.2 Simulierte Stichproben-BDI-II Rohwerte, Z Scores und T Scores

1.4 Testtheorien

Generell kann man mit der *Klassischen Testtheorie*, der *Faktorenanalyse* und der *Item-Response-Theorie* drei Theorien unterscheiden, die der Analyse von Testdaten zugrunde liegen. Allen drei Theorien ist gemein, dass es sich bei ihnen um probabilistische Modelle von beobachtbaren Item- und Testwerten handelt. Weiterhin ist ihnen gemein, dass sie neben der Modellierung von beobachtbaren Testwerten durch Zufallsvariablen auch latente (nicht direkt beobachtbare) Zufallsvariablen zur Erklärung beobachtbarer Daten

heranziehen. In der wissenschaftlichen Betrachtung klinisch-diagnostischer Verfahren spielen insbesondere die Klassische Testtheorie bezüglich der Reliabilität und der internen Konsistenz von Testverfahren sowie die Faktorenanalyse bezüglich der faktoriellen Validität von Testverfahren zentrale Rollen. Die Item-Response-Theorie spielt derzeit in Bezug auf klinisch-diagnostische Verfahren eher eine geringe Rolle (vgl. Reise & Waller (2009)). Wir wollen diese Hauptströmungen der Testtheorie im Folgenden kurz skizzieren.

Klassische Testtheorie

Die Klassische Testtheorie bildet die Grundlage für die Evaluation des Testgütekriteriums der Reliabilität. Die Klassische Testtheorie besteht im Kern aus einem probabilistischen Modell für Item- oder Testergebnisse eines oder mehrerer Individuen für einen oder mehrere Tests. Zentrale Aspekte der Klassische Testtheorie gehen auf die Arbeiten von Charles Spearman Anfang des 20. Jahrhunderts zurück und wurden unter anderem durch Gulliksen (1950) und Lord & Novick (1968) weiter etabliert. Grundlegendes Konzept der Klassische Testtheorie ist ein Messfehlermodell der Form

$$\text{Beobachteter Wert} = \text{Wahrer Wert} + \text{Messfehler} \tag{1.9}$$

wie in der Form "$y = \mu + \varepsilon$" aus der Frequentistischen Inferenztheorie bekannt. Allerdings treten dabei im Rahmen der Klassische Testtheorie einige Besonderheiten auf. Zum einen basieren die Modelle der Klassische Testtheorie in der Regel *nicht* auf Normalverteilungsannahmen, was dem diskreten Charakter der beobachteten Testergebnisse geschuldet sein mag, sondern beschränken sich auf die Modellierung durch allgemeine Zufallsvariablen und die Analyse ihrer Erwartungswerte, Varianzen und Kovarianzen. Zum anderen werden "wahre Werte" im Gegensatz zur Frequentistischen Inferenz nicht als feste Werte, sondern auch als Realisierungen von Zufallsvariablen modelliert. Schließlich tritt insbesondere in der Formulierung nach Lord & Novick (1968) noch der Versuch zutage, "wahre Werte" möglichst nicht naiv-realistisch bei der Analyse von Testergebnissen zu unterstellen, sondern ihnen eine frequentistisch-propensitäre Interpretation zu geben, um sie aus naturwissenschaftlicher Sicht weniger angreifbar zu machen. Die Klassische Testtheorie liefert unter anderem die Grundlage für die in der Anwendung häufig berichteten Korrelationen zur Retestreliabilität und Cronbach's α als Maß für die interne Konsistenz eines Tests. Wir orientieren uns in der folgenden Darstellung der Klassische Testtheorie überwiegend an Krauth (1995).

Faktorenanalyse

Grundlage der Faktorenanalyse ist ein ein probabilistisches Modell der Kovarianzeigenschaften von Testitems und bildet in der Anwendung die Grundlage für das Testgütekriterium der *faktoriellen Valididät*. Bedeutende klassische Beiträge zur Faktorenanalyse stammen von Spearman (1904), Hotelling (1933) und Lawley (1940). Wie die Klassische Testtheorie modelliert auch die Faktorenanalyse beobachtete Testergebnisse, hier meist auf der Ebene von Items, durch ein latentes Variablenmodell

$$\text{Beobachteter Wert} = \text{Faktorladungen} \cdot \text{Wahre Faktorenwerte} + \text{Messfehler}. \tag{1.10}$$

Die "wahren Werte" der Klassische Testtheorie und die "Wahren Faktorwerte" der Faktorenanalyse sind von ähnlichem Charakter, wobei in der Faktorenanalyse typischerweise mehr als ein einzelner wahrer Wert angenommen wird. Unter anderem aufgrund der Tatsache, dass mit dem bebachteten Wert in der Faktorenanalyse ein einzelnes Datum durch das Zusammenspiel dreier nicht beobachtbarer Werte (Faktorladung, Faktorwert, Messfehler) erklärt wird, ist die eindeutige Schätzung von Faktorladungen und Faktorwerten

nicht ohne weiteres möglich. Diese Tatsache hat klassischerweise dazu geführt, dass aus einer Vielzahl möglicher Lösungen eines Faktorenanalysemodells ein spezielles mithilfe heuristischer Nebenkriterien ausgewählt wird, ein Prozess, der sich in den *Rotationsverfahren* der *explorativen Faktorenanalyse* niederschlägt. Allerdings kann das Faktorenanalysemodell durch geeignete Annahmen, wie zum Beispiel die Normalverteilung von wahren Faktorwerten und Messfehlern und die Nicht-Effizienz ausgewählter Faktorenwerte zumindest in Ansätzen identifizierbarer gestaltet werden und zu einem validen Inferenzmodell ausgebaut werden. Dies ist das Thema der *konfirmatorischen Faktorenanalyse*. Mit den *Strukturgleichungsmodellen* (vgl. Bollen (1989)) schließlich wurde die Faktorenanalyse im Computerzeitalter weiter verallgemeinert und zu einem äußerst flexiblen Modellierungsansatz ausgebaut. Wir orientieren uns in der Darstellung der Faktorenanalyse überwiegend an Rencher & Christensen (2012).

Item-Response-Theorie

Die Item-Response-Theorie (IRT) modelliert die Wahrscheinlichkeit einer bestimmten Testantwort als Funktion der Fähigkeit oder des Zustands der Testperson. Wichtige Beiträge zur IRT stammen von Rasch (1960), Novick (1966) und Lord (1980). Die IRT verwendet im Ansatz logistische Regressionsmodelle und findet vor allem bei Leistungsstest Anwendung. Sie bezieht sich dabei explizit auf die Messtheorie nach Stevens (1946), hat jedoch im klinischen Kontext eine geringere bis keine Bedeutung (vgl. Reise & Waller (2009)).

2 Klassische Testtheorie

2.1 Modelle multipler Testmessungen

Die Klassische Testtheorie nimmt ihren Ausgang von Modellen für beobachtete Werte von mehreren Testmessungen von mehreren Personen. Um für die folgende Modellformulierung ein konkretes Datenbeispiel vor Augen zu haben, aufgrund dessen dann Parameter des Modells der Klassischen Testtheorie geschätzt werden können, gehen wir davon aus, dass wir einen Datensatz wie in untenstehender Tabelle gezeigt vorliegen haben. Dieser Datensatz soll die T-Scores von $n = 20$ Personen abbilden, für die jeweils $m = 10$ Tests zur Messung der Depressionssymptomatik durchgeführt wurden. Speziell stellen wir uns vor, dass für jede Person der T-Score des BDI-II, des PHQ-9, der HDRS, der CES-D, und der MADRS zu jeweils zwei Zeitpunkten (T0 und T1) bestimmt wurde. Im Folgenden werden wir Personen mit dem Index i bezeichnen, haben hier also $i = 1, ..., 20$ und Testmessungen mit dem Index j bezeichnen, haben hier also $j = 1, ..., 10$.

Tabelle 2.1 T-Score Beispieldatensatz für n = 20 Personen und m = 10 Testmessungen

Person	BDI-II T0	BDI-II T1	PHQ-9 T0	PHQ-9 T1	HDRS T0	HDRS T1	CES-D T0	CES-D T1	MADRS T0	MADRS T1
1	53.2	56.8	46.9	51.8	61.4	58.6	46.3	51.3	55.6	52.6
2	58.2	55.4	58.3	59.0	68.6	43.5	53.8	63.8	64.3	57.6
3	54.6	58.8	51.3	64.2	51.1	50.3	61.0	52.8	58.7	55.7
4	43.0	45.6	43.9	55.6	47.1	50.6	40.2	49.0	53.0	41.5
5	54.0	53.0	58.5	51.8	47.5	45.9	47.2	50.0	48.2	59.7
6	59.3	57.5	50.8	56.9	69.8	55.0	62.9	63.8	57.0	60.5
7	37.8	20.7	24.7	33.2	22.3	20.2	23.1	26.3	27.9	30.3
8	33.8	35.7	42.5	53.3	38.5	36.6	40.0	33.7	48.6	32.4
9	51.3	48.2	48.8	50.0	43.5	53.2	43.6	54.4	48.5	51.9
10	60.1	58.2	61.6	61.4	65.5	61.2	60.8	60.6	68.8	59.1
11	44.4	47.9	43.8	43.6	43.2	44.2	41.9	52.8	42.9	42.0
12	62.1	55.3	58.0	57.0	62.5	49.8	54.1	48.5	53.5	50.0
13	52.3	48.3	51.7	48.1	57.7	63.0	58.4	60.8	61.1	51.9
14	53.9	54.4	50.9	58.0	54.2	55.6	44.5	58.6	51.8	49.5
15	37.5	41.0	40.6	38.3	42.1	34.0	35.7	43.8	36.9	43.1
16	62.4	51.8	58.7	41.6	52.0	47.1	51.2	54.6	50.8	63.6
17	61.8	65.5	57.0	60.6	72.9	70.0	56.6	64.3	59.3	64.6
18	45.1	39.9	44.5	42.4	36.8	36.9	43.4	44.1	39.1	39.6
19	53.7	55.5	49.4	47.6	47.5	42.5	49.4	51.7	49.2	56.8
20	44.6	39.9	34.7	41.5	42.1	45.6	32.0	42.8	46.5	47.2

2.1.1 Das Modell multipler Testmessungen

Die Klassische Testtheorie in der Formalisierung nach Novick (1966) und Lord & Novick (1968) nimmt ihren Ausgang von der Definition des *wahren Werts* einer Testmessung einer Person, die die Definitionen des *beobachteten Werts* und des *Messfehlers* impliziert. Die Definition nutzt das Konzept eines *bedingten Erwartungswerts* (vgl. Kapitel 4) und hat folgende Form.

Definition 2.1 (Wahrer Wert, beobachteter Wert und Messfehler)**.** Für $i = 1, ..., n$ und $j = 1, ..., m$ ist der *wahre Wert (true score)* t_{ij} einer Person i für eine Testmessung j

definiert als der bedingte Erwartungswert des *beobachteten Werts (observed score)* der Person für diese Testmessung

$$t_{ij} := \mathbb{E}(y_{ij}|\tau_{ij} = t_{ij}). \tag{2.1}$$

Der *Messfehler (error score)* einer Person ist definiert als die Zufallsvariable

$$\varepsilon_{ij} := y_{ij} - t_{ij}. \tag{2.2}$$

•

Mit dem Begriff einer *Testmessung* ist hier etwas unspezifisch und je nach Anwendung entweder eine Messung mithilfe eines einzelnen Items oder mithilfe der Summe mehrer Items gemeint. An späterer Stelle werden wir die Unterscheidung von *Itemscores* und *Testsummenscores* explizit machen. Letztere induzieren im Rahmen der Klassischen Testtheorie den Begriff der *m-Komponententestmodelle* (vgl. Kapitel 2.3). Man beachte, dass in Definition 2.1 der Definition bedingter Erwartungswerte gemäß τ_{ij}, y_{ij} und ε_{ij} Zufallsvariablen sind und $t_{ij} \in \mathbb{R}$ eine Konstante ist.

Die Definition des wahren Werts t_{ij} in Definition 2.1 ist etwas speziell (um nicht zu sagen zirkulär bis tautologisch), da t_{ij} mithilfe von t_{ij} definiert wird. Die zentrale Motivation von Novick (1966) und Lord & Novick (1968) den wahren Wert t_{ij} als bedingten Erwartungswert, anstelle von zum Beispiel einfach einer Realisierung der Zufallvariable τ_{ij} zu definieren, war es, sich einer Diskussion zur metaphysischen Bedeutung eines wahren Werts zu entziehen. Hätten Novick (1966) und Lord & Novick (1968) beispielsweise den wahren Wert als Realisierung einer latenten Variable definiert, die einer Person für eine bestimmte Testmessung eigen sein soll, so hätte dies in der zeitgenössischen Diskussion eindeutig den Charakter einer angreifbaren metaphysischen Aussage. Stattdessen versuchen Novick (1966) und Lord & Novick (1968) in ihrer Definition möglichst operationalistisch vorzugehen und den wahren Wert einer Person für eine Testmessung als "Durchschnitt der beobachteten Werte" einer Person über "wiederholte Testmessungen unter identischen Bedingungen" darzustellen. Zur Bedeutung des wahren Werts zitieren Lord & Novick (1968), S. 29 - 30 Lazarsfeld (1959):

"Angenommen, wir fragen eine Person, Herrn Brown, wiederholt, ob er die Vereinten Nationen befürwortet; nehmen wir weiter an, dass wir ihm nach jeder Frage „das Gehirn waschen" und ihm dann dieselbe Frage erneut stellen. Da Herr Brown unsicher ist, wie er zu den Vereinten Nationen steht, wird er manchmal eine befürwortende und manchmal eine ablehnende Antwort geben. Nachdem wir dieses Verfahren viele Male durchgeführt haben, berechnen wir anschließend den Anteil der Male, in denen Herr Brown die Vereinten Nationen befürwortet hat."

Diese Proportion wollen Novick (1966) und Lord & Novick (1968) dann als wahren Wert verstehen (vgl. auch Borsboom et al. (2004)). Natürlich ist dieser Versuch einer antimetaphysischen Grundhaltung nicht durchhaltbar: zum einen handelt es sich bei der "wiederholten Testmessung unter identischen Bedingugen" um ein idealisiertes, in der Realität nicht durchführbares, Gedankenexperiment. Zum anderen definieren Novick (1966) und Lord & Novick (1968) den wahren Wert auch gerade nicht als (endlichen) Mittelwert einer Messreihe, sondern als idealisierten Erwartungswert einer Zufallsvariable. Als wirklich operationalistische Definition überzeugt Definition 2.1 also nicht, verkompliziert die Entwicklung einer Standardmessfehlertheorie, wie sie beispielsweise die klassische Formalisierung des Allgemeinen Linearen Modells darstellt, für die Analyse von Tests und Fragebögen

aber beträchtlich. Dies mag einer der Gründe sein, warum die Klassische Testtheorie bis heute lediglich in der Psychologie und wenig darüberhinaus von Bedeutung ist.

Ähnlich gelagert ist die Bezeichnung der bedingten Verteilung des beobachteten Werts $\mathbb{P}(y_{ij}|\tau_{ij} = t_{ij})$ als *Propensitätsverteilung* durch Lord & Novick (1968), welche die intraindividuelle Variabilität des beobachteten Werts bei festem wahren Wert modelliert. Dabei klingt an, dass Lord & Novick (1968) eine Propensitätsinterpretation von Wahrscheinlichkeiten als kausal bedingte "Verwirklichungstendenzen", die, im Gegensatz zur Frequentistischen Interpretation auch im Einzelfall Sinn ergeben, implizieren. Allerdings unterstellen Propensitätsverteilungen kausale Prozesse, die in der Regel nicht spezifiziert und damit auch nicht beobachtbar sind, und führen damit letztlich auch wieder auf metaphysische Aussagen (vgl. auch Borsboom et al. (2004) und Borsboom (2009)).

In unserer Darstellung wollen wir dem modell-basierten realistischen Ansatz folgen und wählen mit Definition 2.2 deshalb eine Formulierung des Modells multipler Testmessungen der Klassischen Testtheorie, die mit Definition 2.1 und damit natürlich auch den theoretischen Ergebnissen der Klassischen Testtheorie kongruent ist, aber nicht versucht, ihren Modellcharakter zu verschleiern. Insbesondere betont unserer Ansatz dabei das Gesamtziel der Modellierung einer Menge von m Testmessungen von n Personen als eines Datensatzes von nm Datenpunkten. Wir definieren das *Modell multipler Testmessungen* daher wie folgt.

Definition 2.2 (Modell multipler Testmessungen)**.** Für $i = 1, ..., n$ und $j = 1, ..., m$ seien τ_{ij} eine Zufallsvariable, die den *wahren Wert* der $i$ten Person in der $j$ten Testmessung modelliert und y_{ij} eine Zufallsvariable, die den *beobachteten Wert* der $i$ten Person in der $j$ten Testmessung modelliert. Dann nennen wir die gemeinsame Verteilung der τ_{ij} und y_{ij} mit der Faktorisierungseigenschaft

$$\mathbb{P}\left(\tau_{11}, y_{11}, ..., \tau_{nm}, y_{nm}\right) := \prod_{i=1}^{n} \mathbb{P}(\tau_{i1}, ..., \tau_{im}) \prod_{j=1}^{m} \mathbb{P}(y_{ij}|\tau_{ij}) \tag{2.3}$$

das *Modell multipler Testmessungen*, wenn gilt, dass

$$\mathbb{P}(\tau_{11}, ..., \tau_{1m}) \prod_{j=1}^{m} \mathbb{P}(y_{1j}|\tau_{1j}) = \cdots = \mathbb{P}(\tau_{n1}, ..., \tau_{nm}) \prod_{j=1}^{m} \mathbb{P}(y_{nj}|\tau_{nj}). \tag{2.4}$$

•

Die Definition des Modells multipler Testmessungen bildet einige grundlegende Annahmen zur Unabhängigkeit und Identität von Verteilungen in der Klassischen Testtheorie ab. Zunächst einmal wird angenommen, dass die gemeinsame Verteilung der τ_{ij} und y_{ij} über $i = 1, ..., n$ faktorisiert, dass die Verteilungen der wahren Werte und beobachteten Werte also über Personen unabhängig sind. Wissen um wahre oder beobachtete Werte einer Person ändert die angenommenen Verteilungen der wahren und beobachteten Werte anderer Personen also nicht. Im Gegensatz dazu faktorisiert für jedes $i = 1, ..., n$ die gemeinsame Verteilung der $\tau_{i1}, ..., \tau_{im}$ über Testmessungen $j = 1, ..., m$ nicht notwendigerweise. Die wahren Werte einer Person können also abhängig sein und damit Wissen um die Ausprägung einer Testmessung bei einer Person die Verteilung des wahren Werts bei anderen Testmessungen derselben Person informieren. In der Klassischen Testtheorie werden verschiedene Arten dieser Form von Abhängigkeiten unterschieden und, wie wir später sehen

werden, beispielsweise als *Parallelität*, τ-*Äquivalenz* oder *essentielle* τ-*Äquivalenz* bezeichnet. Weiterhin wird für jede Person $i = 1, ..., n$ angenommen, dass die beobachteten Werte y_{ij} für $j = 1, ..., m$ gegeben τ_{ij} bedingt unabhängig sind. Dies impliziert einerseits, dass für eine Person der wahre Wert in Testmessung $k \neq j$ den beobachteten Wert in Testmessung j nicht beinflusst und andererseits, dass der beobachtete Wert in Testmessung $k \neq j$ den beobachteten Wert in Testmessung j nicht beeinflusst.

Schließlich wird angenommen, dass die Marginalverteilungen

$$\mathbb{P}(\tau_{i1}, y_{i1}, ..., \tau_{im}, y_{im}) = \mathbb{P}(\tau_{i1}, ..., \tau_{im}) \prod_{j=1}^{m} \mathbb{P}(y_{ij}|\tau_{ij}) \tag{2.5}$$

über Personen $i = 1, ..., n$ identisch sind. Man mag sich die Realisierungen von wahren Werten und beobachteten Werten also als unabhängige und identische Realisierungen aus einer "Populationsverteilung"

$$\mathbb{P}(\tau_{\bullet 1}, y_{\bullet 1}, ..., \tau_{\bullet m}, y_{\bullet m}) = \mathbb{P}(\tau_{\bullet 1}, ..., \tau_{\bullet m}) \prod_{j=1}^{m} \mathbb{P}(y_{\bullet j}|\tau_{\bullet j}) \tag{2.6}$$

vorstellen, wobei das Subskript $\bullet$ die Unspezifität dieser Verteilung bezüglich einer Person symbolisieren soll.

Betrachtet man den Spezialfall einer einzelnen Testmessung bei n Personen, so ergibt sich eine vereinfachte Form von Definition 2.2, auf die man häufig in der psychologischen Literatur trifft. Nach Definition 2.2 gilt für eine Testmessung, also $m = 1$,

$$\mathbb{P}\left(\tau_{11}, y_{11}, ..., \tau_{n1}, y_{n1}\right) := \prod_{i=1}^{n} \mathbb{P}(\tau_{i1})\mathbb{P}(y_{i1}|\tau_{i1}), \tag{2.7}$$

wobei nach Annahme von Definition 2.2 die gemeinsamen Marginalverteilungen $\mathbb{P}(\tau_{i1}, y_{i1})$ über Personen $i = 1, ..., n$ identisch sind. Wie oben kann man sich die wahren und beobachteten Werte für eine Testmessung also als unabhängige Realisierungen der "Populationsverteilung"

$$\mathbb{P}(\tau_{\bullet 1})\mathbb{P}(y_{\bullet 1}|\tau_{\bullet 1}) \tag{2.8}$$

vorstellen. Verzichtet man nun noch auf die Subskripte $\bullet 1$, gelangt man zu folgender vereinfachter Definition des Modells der Klassischen Testtheorie.

Definition 2.3 (Vereinfachtes Modell der Klassischen Testtheorie)**.** τ sei eine Zufallsvariable, die die Verteilung der *wahren Werte* einer Testmessung in einer Population von Individuen beschreibt und y sei eine Zufallsvariable, die die Verteilung der *beobachteten Werte* dieser Testmessung beschreibt. Dann heißt die gemeinsame Verteilung von τ und y

$$\mathbb{P}\left(\tau, y\right) = \mathbb{P}(\tau)\mathbb{P}(y|\tau) \tag{2.9}$$

das *Vereinfachte Modell der Klassischen Testtheorie*.

•

Definition 2.3 hat gegenüber Definition 2.2 den Vorteil, dass weniger Zufallsvariablen und Indizes auftreten und auf die Redundanz der unterschiedlichen Bezeichnung vieler gleicher

Verteilungen verzichtet werden kann. Im Sinne Frequentistischer Produktmodelle mag man bezüglich der Daten von n Personen hier auch

$$(\tau_1, y_1), ..., (\tau_n, y_n) \sim \mathbb{P}(\tau, y) \tag{2.10}$$

schreiben, wobei nur die $y_1, ..., y_n$ beobachtete, die $\tau_1, ..., \tau_n$ dagegen latente Zufallsvariablen sind. Weiterhin gilt, dass man viele wichtige Eigenschaften des Modells der Klassischen Testtheorie schon basierend auf den Eigenschaften von $\mathbb{P}(\tau, y)$ begründen kann, wie wir unten sehen werden. Generell ist das vereinfachte Modell der Klassischen Testtheorie einfacher zu handhaben als das Modell multipler Testmessungen. Problematisch wird es allerdings, sobald mehrere Testmessungen ins Spiel kommen, beispielsweise bei Abhängigkeitsbetrachtungen zwischen zwei Tests oder zwei Items eines Tests. Dies ist allerdings bei den meisten Aussagen der Klassischen Testtheorie der Fall. Außerdem gilt auch, dass Schätzer der Modellparameter immer auf allen Werten der beobachteten Zufallsvariablen $y_{1j}, ...y_{nj}$ beruhen, diese in Definition 2.3 aber überhaupt nicht auftreten. Definition 2.3 ist für theoretische Betrachtungen also oft einfacher zu handhaben als Definition 2.2, der Anwendung der Klassischen Testtheorie in der Testdatenanalyse liegt im Allgemeinen aber Definition 2.2 zugrunde. Wir werden je nach Bedarf zwischen beiden Modellformulierungen hin und her wechseln, halten aber fest, dass die Modellformulierung im Sinne von Definition 2.2 unser Standardfall ist.

Eigenschaften des Modells multipler Testmessungen

Eigenschaften bezüglich einer Testmessung

Das Modell multipler Testmessungen nach Definition 2.2 hat zunächst eine Reihe von Eigenschaften bezüglich einer und damit jeder Testmessung, die für die Anwendung der Klassischen Testtheorie grundlegend sind. Wir fassen fünf dieser Eigenschaften in folgendem Theorem zusammen.

Theorem 2.1 (Eigenschaften bezüglich einer Testmessung)**.** *Gegeben sei das Modell multipler Testmessungen. Dann gelten für alle $i = 1, ..., n$ und alle $j = 1, ..., m$*

(1) $\mathbb{E}(\varepsilon_{ij}|\tau_{ij} = t_{ij}) = 0$
(2) $\mathbb{E}(\varepsilon_{ij}) = 0$
(3) $\mathbb{C}(\tau_{ij}, \varepsilon_{ij}) = 0$
(4) $\mathbb{V}(y_{ij}) = \mathbb{V}(\tau_{ij}) + \mathbb{V}(\varepsilon_{ij})$
(5) $\mathbb{C}(y_{ij}, \tau_{ij}) = \mathbb{V}(\tau_{ij})$

∘

Beweis. Zur Vereinfachung der Notation im Sinne des einfachen Modells der klassischen Testtheorie nach Definition 2.3 setzen wir zunächst für alle $i = 1, ..., n$ und $j = 1, ...m$

$$y := y_{ij} \text{ und } \tau := \tau_{ij} \text{ mit Ergebnisräumen } Y := Y_{ij} \text{ und } T := T_{ij}. \tag{2.11}$$

Weiterhin bezeichnen wir die Werte von y mit y und die Werte von τ mit t. Schließlich betrachten wir nur den diskreten Fall, setzen also die Existenz einer WMF $p : Y \times T \to [0, 1]$ der Form

$$p(t, \mathrm{y}) = p(\mathrm{y}|t)p(t) \tag{2.12}$$

voraus. Der kontinuierliche Fall oder gemischt diskret-kontinuierliche Fall folgt dann jeweils analog.

(1) Es gilt

$$\begin{aligned}\mathbb{E}(\varepsilon|\tau = t) &:= \mathbb{E}(y - \tau|\tau = t)\\ &= \sum_{\mathrm{y}\in Y} (\mathrm{y} - t)\, p(\mathrm{y}|t)\\ &= \sum_{\mathrm{y}\in Y} \mathrm{y}p(\mathrm{y}|t) - \sum_{\mathrm{y}\in Y} tp(\mathrm{y}|t)\\ &= \mathbb{E}(y|\tau = t) - t\sum_{\mathrm{y}\in Y} p(\mathrm{y}|t)\\ &= t - t\cdot 1\\ &= 0.\end{aligned} \tag{2.13}$$

(2) Es gilt

$$\begin{aligned}\mathbb{E}(\varepsilon) &:= \mathbb{E}(y - \tau)\\ &= \sum_{t\in T}\sum_{\mathrm{y}\in Y} (\mathrm{y} - t)\, p(t, \mathrm{y})\\ &= \sum_{t\in T}\sum_{\mathrm{y}\in Y} (\mathrm{y} - t)\, p(\mathrm{y}|t)p(t)\\ &= \sum_{t\in T}\sum_{\mathrm{y}\in Y} \mathrm{y}p(\mathrm{y}|t)p(t) - tp(\mathrm{y}|t)p(t)\\ &= \sum_{t\in T}\sum_{\mathrm{y}\in Y} p(t)\,(\mathrm{y}p(\mathrm{y}|t) - tp(\mathrm{y}|t))\\ &= \sum_{t\in T} p(t)\left(\sum_{\mathrm{y}\in Y} \mathrm{y}p(\mathrm{y}|t) - t\sum_{\mathrm{y}\in Y} p(\mathrm{y}|t)\right)\\ &= \sum_{t\in T} p(t)\,(t - t\cdot 1)\\ &= \sum_{t\in T} p(t)\cdot 0\\ &= 0.\end{aligned} \tag{2.14}$$

(3) Es gilt

$$\begin{aligned}\mathbb{C}(\tau, \varepsilon) &= \mathbb{E}\left((\tau - \mathbb{E}(\tau))(\varepsilon - \mathbb{E}(\varepsilon))\right)\\ &= \mathbb{E}\left((\tau - \mathbb{E}(\tau))\varepsilon\right)\\ &= \mathbb{E}\left((\tau - \mathbb{E}(\tau))(y - \tau)\right)\\ &= \sum_{t\in T}\sum_{\mathrm{y}\in Y} \left((t - \mathbb{E}(\tau))(\mathrm{y} - t)\right) p(t, \mathrm{y})\\ &= \sum_{t\in T}\sum_{\mathrm{y}\in Y} (t - \mathbb{E}(\tau))(\mathrm{y} - t)p(\mathrm{y}|t)p(t)\\ &= \sum_{t\in T} p(t)(t - \mathbb{E}(\tau))\sum_{\mathrm{y}\in Y} (\mathrm{y} - t)\, p(\mathrm{y}|t)\\ &= \sum_{t\in T} p(t)(t - \mathbb{E}(\tau))\sum_{\mathrm{y}\in Y} (\mathrm{y}p(\mathrm{y}|t) - tp(\mathrm{y}|t))\\ &= \sum_{t\in T} p(t)(t - \mathbb{E}(\tau))\left(\sum_{\mathrm{y}\in Y} \mathrm{y}p(\mathrm{y}|t) - t\sum_{\mathrm{y}\in Y} p(\mathrm{y}|t)\right)\\ &= \sum_{t\in T} p(t)(t - \mathbb{E}(\tau))\,(t - t\cdot 1)\\ &= \sum_{t\in T} p(t)(t - \mathbb{E}(\tau))\cdot 0\\ &= 0.\end{aligned} \tag{2.15}$$

(4) Mit dem Theorem zu Varianzen von Summen und Differenzen von Zufallsvariablen (Theorem 4.21) sowie Aussage (3) des vorliegenden Theorems gilt

$$\mathbb{V}(y) = \mathbb{V}(\tau + \varepsilon) = \mathbb{V}(\tau) + \mathbb{V}(\varepsilon) + 2\mathbb{C}(\tau, \varepsilon) = \mathbb{V}(\tau) + \mathbb{V}(\varepsilon) + 2\cdot 0 = \mathbb{V}(\tau) + \mathbb{V}(\varepsilon) \tag{2.16}$$

(5) Mit dem Kovarianzverschiebungssatz (Theorem 4.16), der Linearkombinationseigenschaft des Erwartungswerts (Theorem 4.5), Aussage (2) des vorliegenden Theorems und der Tatsache, dass mit Aussage (4) des vorliegenden Theorems außerdem folgt, dass

$$\mathbb{E}(\varepsilon\tau) = \mathbb{E}(\tau\varepsilon) = \mathbb{C}(\tau, \varepsilon) + \mathbb{E}(\tau)\mathbb{E}(\varepsilon) = 0 + \mathbb{E}(\tau) \cdot 0 = 0, \tag{2.17}$$

gilt

$$\begin{aligned}
\mathbb{C}(y, \tau) &= \mathbb{E}(y\tau) - \mathbb{E}(y)\mathbb{E}(\tau) \\
&= \mathbb{E}((\tau + \varepsilon)\tau) - \mathbb{E}(\tau + \varepsilon)\mathbb{E}(\tau) \\
&= \mathbb{E}(\tau^2 + \varepsilon\tau) - (\mathbb{E}(\tau) + \mathbb{E}(\varepsilon))\,\mathbb{E}(\tau) \\
&= \mathbb{E}(\tau^2) + \mathbb{E}(\varepsilon\tau) - \mathbb{E}(\tau)^2 - \mathbb{E}(\varepsilon)\mathbb{E}(\tau) \\
&= \mathbb{E}(\tau^2) + \mathbb{E}(\varepsilon\tau) - \mathbb{E}(\tau)^2 - 0 \cdot \mathbb{E}(\tau) \\
&= \mathbb{E}(\tau^2) + 0 - \mathbb{E}(\tau)^2 - 0 \cdot \mathbb{E}(\tau) \\
&= \mathbb{E}(\tau^2) - \mathbb{E}(\tau)^2 \\
&= \mathbb{V}(\tau).
\end{aligned} \tag{2.18}$$

□

Wie die Subskripte in Theorem 2.1 verdeutlichen, beziehen sich die Aussagen von Theorem 2.1 auf die Zufallsvariablen zur Modellierung der Daten einer Person i und einer Testmessung j und gelten gleichermaßen für alle $i = 1, ..., n$ und $j = 1, ..., m$. Aussage (1) von Theorem 2.1 betrifft den Erwartungswert des Messfehlers bedingt auf einem festen Wert des wahren Werts, in diesem Fall ist der wahre Wert also *keine* Zufallsvariable und die Verteilung von Interesse ist $\mathbb{P}(\varepsilon_{ij}|\tau_{ij} = t_{ij})$. Aussagen (2) bis (5) beziehen sich auf Eigenschaften der (gemeinsamen) Marginalverteilungen von y_{ij}, τ_{ij} und ε_{ij}. Im Sinne des vereinfachten Modells der Klassischen Testtheorie (Definition 2.3) werden obige Eigenschaften oft auch als

(1) $\mathbb{E}(\varepsilon|\tau = t) = 0$
(2) $\mathbb{E}(\varepsilon) = 0$
(3) $\mathbb{C}(\tau, \varepsilon) = 0$
(4) $\mathbb{V}(y) = \mathbb{V}(\tau) + \mathbb{V}(\varepsilon)$
(5) $\mathbb{C}(y, \tau) = \mathbb{V}(\tau)$

geschrieben.

Spezieller Weise *folgt* nach Aussage (1) von Theorem 2.1 für das Modell multipler Messungen, dass der bedingte Erwartungswert des Messfehlers $\mathbb{E}(\varepsilon|\tau = t)$ gleich Null ist aus der Definition des wahren Werts. Dies steht im direkten Gegenentwurf zu typischen Annahmen über Messfehler, die üblicherweise einen Messfehler mit einem (bedingten) Erwartungswert von Null *definieren*, da sie von Null verschiedene Beiträge zu einem Datenpunkt als Teil der durch sie repräsentierten Theorie konzeptualisieren. Weil im Modell multipler Testmessungen der bedingte Erwartungswert des Messfehlers für *jeden* wahren Wert t gleich Null, folgt dann nach (2) von Theorem 2.1, dass auch der marginale Erwarungswert des Messfehlers $\mathbb{E}(\varepsilon)$ gleich Null ist.

Aussage (3) von Theorem 2.1, dass im Modell multipler Testmessungen also gilt, dass die Kovarianz der wahren Werte und der Messfehler gleich Null ist, besagt bekanntlich, dass hohe oder niedrige wahre Werte nicht systematisch mit hohen oder niedrigen Messfehlern assoziiert sind. Die daraus folgende Aussagen (4) und (5) von Theorem 2.1, dass also die Varianz der beobachteten Werte additiv in Varianzbeiträge der wahren Wert und der Messfehler zerlegt werden kann und dass die Kovarianz der beobachteten Werte und der wahre Werte einer Testmessung der Varianz der wahren Werte enstpricht, werden an späterer Stelle essentiell für die Eigenschaften der Reliabilität von Testmessungen sein.

Beispiel

Im Folgenden wollen wir Theorem 2.1 noch an einem konkreten ersten Beispiel nach Lord & Novick (1968), Exercise 2.17 für ein Modell multipler Testmessungen veranschaulichen, wobei wir hierbei natürlich auf eine einzige Testmessung fokussieren.

Theorem 2.2 (Normalverteilungsbeispiel)**.** *Es seien* $i = 1, ..., n$ *und* $m := 1$ *mit*

$$\mathbb{P}(\tau_{i1}) := N(\mu, 1) \text{ und } \mathbb{P}(y_{i1}|\tau_{i1}) := N(\tau_{i1}, 1) \tag{2.19}$$

Dann gelten

(1) $\mathbb{P}(y_{i1}) = N(\mu, 2)$
(2) $\mathbb{P}(\varepsilon_{i1}|\tau_{i1}) = N(0, 1)$
(3) $\mathbb{P}(\varepsilon_{i1}) = N(0, 1)$
(4) $\mathbb{C}(\tau_{i1}, \varepsilon_{i1}) = 0$

∘

Beweis. Zur Vereinfachung der Notation setzen wir $\tau := \tau_{i1}, y := y_{i1}, \varepsilon := \varepsilon_{i1}$.

(1) Wir betrachten die durch

$$\mathbb{P}(\tau) = N(\mu, 1) \text{ und } \mathbb{P}(y|\tau) = N(\tau, 1) \tag{2.20}$$

induzierte gemeinsame Verteilung von τ und y, wobei offenbar

$$\mathbb{P}(y|\tau) = N(a \cdot \mu + b, 1) \text{ mit } a := 1 \text{ und } b := 0 \tag{2.21}$$

gilt. Aus dem Theorem zu gemeinsamen Normalverteilungen (vgl. Theorem 4.27) ergibt sich dann zunächst, dass

$$\begin{pmatrix}\tau\\ y\end{pmatrix} \sim N\left(\begin{pmatrix}\mu\\ 1\cdot\mu+0\end{pmatrix}, \begin{pmatrix}1 & 1\cdot 1\\ 1\cdot 1 & 1+1\cdot 1\cdot 1\end{pmatrix}\right) = N\left(\begin{pmatrix}\mu\\ \mu\end{pmatrix}, \begin{pmatrix}1 & 1\\ 1 & 2\end{pmatrix}\right) \tag{2.22}$$

Aus dem Theorem zu marginalen Normalverteilungen (vgl. Theorem 4.26) ergibt sich dann durch Ablesen $y \sim N(\mu, 2)$.

(2) Wir betrachten $\mathbb{P}(\varepsilon|\tau = t)$ für einen beliebigen Wert $t \in \mathbb{R}$. Dann gilt

$$\varepsilon := y - t \text{ mit } y \sim N(t, 1) \tag{2.23}$$

Mit dem Theorem zu linear-affinen Transformation normalverteilter Zufallsvariablen (vgl. Theorem 4.30) gilt dann

$$\varepsilon \sim N\left(t - t, 1^2 \cdot 1\right) = N(0, 1) \tag{2.24}$$

Die Tatsache, dass dies für alle möglichen Werte von τ gilt, ist gerade Aussage (2).

(3) und (4) Wir betrachten die durch

$$\mathbb{P}(\tau) = N(\mu, 1) \text{ und } \mathbb{P}(\varepsilon|\tau) = N(0, 1) \tag{2.25}$$

induzierte gemeinsame Verteilung von τ und ε, wobei offenbar

$$\mathbb{P}(\varepsilon|\tau) = N(a \cdot \mu + b, 1) \text{ mit } a := 0 \text{ und } b := 0 \tag{2.26}$$

gilt. Aus dem Theorem zu gemeinsamen Normalverteilungen (vgl. Theorem 4.27) ergibt sich dann zunächst, dass

$$\begin{pmatrix}\tau\\ \varepsilon\end{pmatrix} \sim N\left(\begin{pmatrix}\mu\\ 0\cdot\mu+0\end{pmatrix}, \begin{pmatrix}1 & 1\cdot 0\\ 0\cdot 1 & 1+0\cdot 1\cdot 0\end{pmatrix}\right) = N\left(\begin{pmatrix}\mu\\ 0\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right) \tag{2.27}$$

Aus dem Theorem zu marginalen Normalverteilungen (vgl. Theorem 4.26) ergeben sich dann durch Ablesen

$$\varepsilon \sim N(0, 1) \text{ und } \mathbb{C}(\tau, \varepsilon) = 0. \tag{2.28}$$

□

Das Beispiel zeigt insbesondere, wie die Spezifika des Modells multipler Testmessungen der Klassischen Testtheorie äquivalent durch Definition eines zeitgemäßen probabilistischen Modells erzeugt werden können. Hier induzieren die Definition der marginalen Verteilung des wahren Werts und die Definition der bedingten Verteilung des beobachteten Werts zunächst die gemeinsame Normalverteilung von τ_{i1} und y_{i1}. Die Definition des Messfehlers als Differenz zwischen beobachtetem Wert und dem bedingten Erwartungswert des wahren Werts ergibt dann die bedingte Verteilung des Messfehlers. Diese und die Definitionen des Beispiels induzieren dann eine gemeinsame Normalverteilung von τ_{i1} und ε_{i1}. Die weiteren Eigenschaften im Sinne von Theorem 2.1 ergeben sich in diesem Beispiel dann mit den Eigenschaften gemeinsamer Normalverteilungen.

Mithilfe folgender Simulation wollen wir das hier diskutierte Beispiel noch weiter verdeutlichen. Wir setzen dafür $\mu := 1$ und generieren 10^4 Realisierungen der marginalen Verteilung von τ_{i1} und der bedingten Verteilung von y_{i1}.

```
n           = 1e4                                  # Personenanzahl
m           = 1                                    # Testmessungsanzahl
mu          = 1                                    # Erwartungswertparameter Wahrer Werte
T           = matrix(rep(NaN, n*m), nrow = n)      # Wahrer Wert Array
Y           = matrix(rep(NaN, n*m), nrow = n)      # Beobachteteter Wert Array
E           = matrix(rep(NaN, n*m), nrow = n)      # Messfehler Array
for(i in 1:n){                                     # Iteration über Personen
  for(j in 1:m){                                   # Iteration über Testmessungen
    T[i,j]  = rnorm(1,mu,1)                        # Wahrer Wert Realisierung
    Y[i,j]  = rnorm(1,T[i,j],1)                    # Beobachteter Wert Realisierung
    E[i,j]  = Y[i,j] - T[i,j]}}                    # Messfehler Realisierung
e_hat_es    = mean(E[,1])                          # Erwartungswertschätzung Messfehler
c_hat_ts_es = cov(T[,1],E[,1])                     # Kovarianzschätzung Wahren Wert, Messfehler
v_hat_os    = var(Y[,1])                           # Varianzschätzung Beobachteter Wert
v_hat_ts    = var(T[,1])                           # Varianzschätzung Wahrer Wert
v_hat_es    = var(E[,1])                           # Varianzschätzung Messfehler
c_hat_os_ts = cov(Y[,1],E[,1])                     # Kovarianzschätzung Beobachteter Wert, Wahrer Wert
```

Abbildung 2.1 A und B zeigen die resultierenden marginale Verteilungen von y_{i1} und ε_{i1} mit ihren theoretischen Entsprechungen laut Theorem 2.2. Abbildung 2.1 C zeigt die gemeinsame Verteilung von τ_{i1} und ε_{i1} mit ihrer theoretischen Entsprechung. Abbildung 2.1 D schließlich zeigt die auf Grundlage der Simulation gewonnenen Schätzer relevanter Erwartungswerte, Varianzen und Kovarianzen in diesem Modell, die die theoretischen Einsichten nach Theorem 2.2 bestätigen.

Eigenschaften bezüglich zweier Testmessungen

Bisher haben wir fünf Eigenschaften des Modells multipler Testmessungen kennengelernt, die für den wahren Wert τ_{ij}, den beobachteten Wert y_{ij} und den Messfehler ε_{ij} einer (und damit jeder) Person i und einer (und damit jeder) Testmessung j gelten. Im Folgenden beschäftigen wir uns mit Eigenschaften des Modells multipler Testmessungen, die für die wahren Werte τ_{ij} und τ_{ik}, beobachteten Werte y_{ij} und y_{ik} und Messfehler ε_{ij} und ε_{ik} einer (und damit jeder) Person i hinsichtlich *zweier* Testmessungen j und k gelten. Wir fassen diese Eigenschaften, die manchmal als *lokale Unkorreliertheit* des Modells multipler Testmessungen bezeichnet werden (vgl. Krauth (1995)), in folgende Theorem zusammen.

Theorem 2.3 (Eigenschaften bezüglich zweier Testmessungen)**.** *Gegeben sei das Modell multipler Testmessungen. Dann gelten für alle* $i = 1, ..., n$ *und alle* j *und* k *mit* $1 \leq j, k \leq m$ *und* $j \neq k$*, dass*

(1) $\mathbb{C}(y_{ij}, y_{ik} | \tau_{ij} = t_{ij}, \tau_{ik} = t_{ik}) = 0$
(2) $\mathbb{C}(\varepsilon_{ij}, \varepsilon_{ik} | \tau_{ij} = t_{ij}, \tau_{ik} = t_{ik}) = 0$

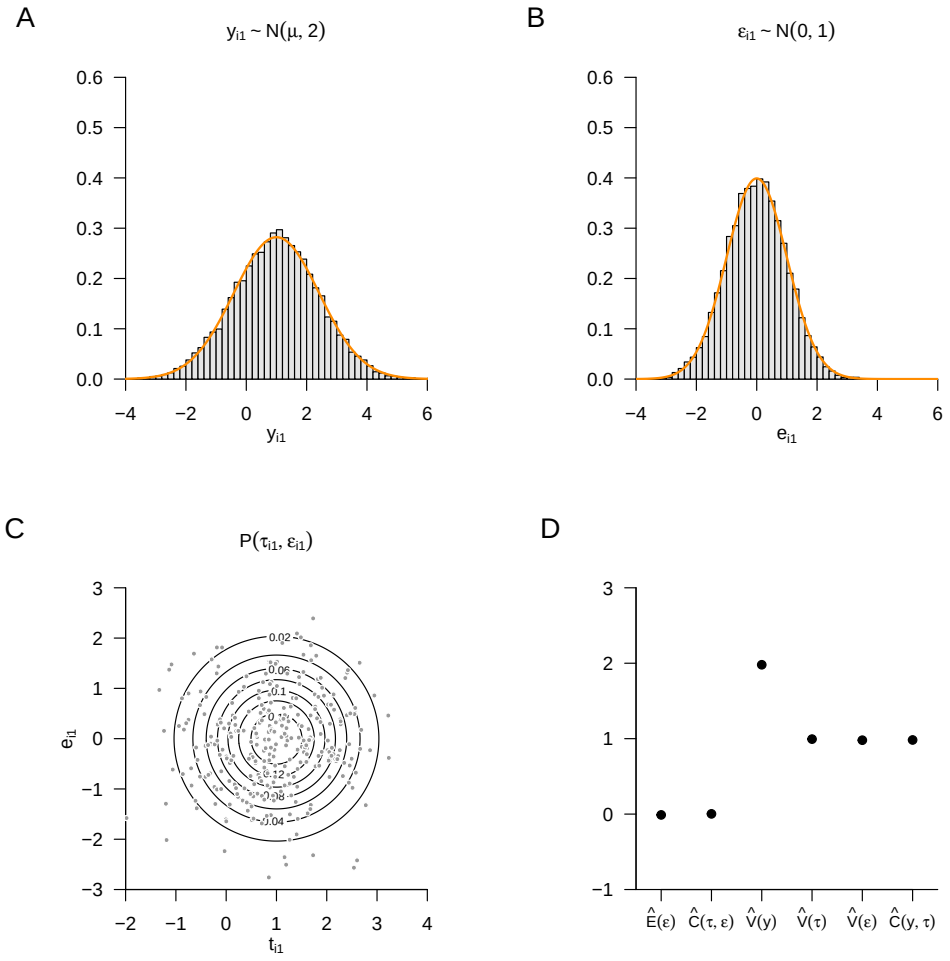

Abbildung 2.1 Normalverteilungsbeispiel bezüglich einer Testmessung.

(3) $\mathbb{C}(\varepsilon_{ij}, \varepsilon_{ik}) = 0$
(4) $\mathbb{C}(\tau_{ij}, \varepsilon_{ik}) = 0$
(5) $\mathbb{C}(y_{ij}, y_{ik}) = \mathbb{C}(\tau_{ij}, \tau_{ik})$

∘

Beweis. Zur Vereinfachung der Notation verzichten wir in den Beweisen auf das i Subskript. Wir betrachten weiterhin nur den diskreten Fall und setzen die Existenz der marginalen WMF

$$p(t_j, \mathrm{y}_j, t_k, \mathrm{y}_k) = p(t_j, t_k)p(\mathrm{y}_j|t_j)p(\mathrm{y}_k|t_k) \tag{2.29}$$

und folglich auch der bedingten Wahrscheinlichkeitsmassefunktion

$$p(\mathrm{y}_j, \mathrm{y}_k|t_j, t_k) = \frac{p(t_j, \mathrm{y}_j, t_k, \mathrm{y}_k)}{p(t_j, t_k)} = \frac{p(t_j, t_k)p(\mathrm{y}_j|t_j)p(\mathrm{y}_k|t_k)}{p(t_j, t_k)} = p(\mathrm{y}_j|t_j)p(\mathrm{y}_k|t_k) \tag{2.30}$$

voraus. Der kontinuierliche Fall folgt dann wieder analog.

(1) Es gilt

$$\begin{aligned}
\mathbb{C}(y_j, y_k|\tau_j = t_j, \tau_k = t_k) &= \sum_{\mathrm{y}_j \in Y_j} \sum_{\mathrm{y}_k \in Y_k} (\mathrm{y}_j - \mathbb{E}(y_j|\tau_j = t_j))(\mathrm{y}_k - \mathbb{E}(y_k|\tau_k = t_k))p(\mathrm{y}_j|t_j)p(\mathrm{y}_k|t_k) \\
&= \sum_{\mathrm{y}_j \in Y_j} (\mathrm{y}_j - t_j)p(\mathrm{y}_j|t_j) \sum_{\mathrm{y}_k \in Y_k} (\mathrm{y}_k - t_k)p(\mathrm{y}_k|t_k) \\
&= \sum_{\mathrm{y}_j \in Y_j} (\mathrm{y}_j - t_j)p(\mathrm{y}_j|t_j) \left(\sum_{\mathrm{y}_k \in Y_k} \mathrm{y}_k p(\mathrm{y}_k|t_k) - t_k \sum_{\mathrm{y}_k \in Y_k} p(\mathrm{y}_k|t_k) \right) \\
&= \sum_{\mathrm{y}_j \in Y_j} (\mathrm{y}_j - t_j)p(\mathrm{y}_j|t_j) \left(t_k - t_k \cdot 1 \right) \\
&= \sum_{\mathrm{y}_j \in Y_j} (\mathrm{y}_j - t_j)p(\mathrm{y}_j|t_j) \cdot 0 \\
&= 0.
\end{aligned} \tag{2.31}$$

(2) Wir bestimmen zunächst $\mathbb{E}(\varepsilon_j \varepsilon_k|\tau_j = t_j, \tau_k = t_k)$. Es gilt

$$\begin{aligned}
\mathbb{E}(\varepsilon_j \varepsilon_k|\tau_j = t_j, \tau_k = t_k) &= \mathbb{E}((y_j - \tau_j)(y_k - \tau_k)|\tau_j = t_j, \tau_k = t_k) \\
&= \sum_{\mathrm{y}_j \in Y_j} \sum_{\mathrm{y}_k \in Y_k} (\mathrm{y}_j - t_j)(\mathrm{y}_k - t_k)p(\mathrm{y}_j|t_j)p(\mathrm{y}_k|t_k) \\
&= \sum_{\mathrm{y}_j \in Y_j} (\mathrm{y}_j - t_j)p(\mathrm{y}_j|t_j) \sum_{\mathrm{y}_k \in Y_k} (\mathrm{y}_k - t_k)p(\mathrm{y}_k|t_k) \\
&= \sum_{\mathrm{y}_j \in Y_j} (\mathrm{y}_j - t_j)p(\mathrm{y}_j|t_j) \left(\sum_{\mathrm{y}_k \in Y_k} \mathrm{y}_k p(\mathrm{y}_k|t_k) - t_k \sum_{\mathrm{y}_k \in Y_k} p(\mathrm{y}_k|t_k) \right) \\
&= \sum_{\mathrm{y}_j \in Y_j} (\mathrm{y}_j - t_j)p(\mathrm{y}_j|t_j) \left(t_k - t_k \cdot 1 \right) \\
&= \sum_{\mathrm{y}_j \in Y_j} (\mathrm{y}_j - t_j)p(\mathrm{y}_j|t_j) \cdot 0 \\
&= 0.
\end{aligned} \tag{2.32}$$

Mit dem Verschiebungssatz der bedingten Kovarianz (Theorem 4.22) und Aussage (1) des Theorems zu den Eigenschaften bezüglich einer Testmessung (Theorem 2.1) folgt dann

$$\mathbb{C}(\varepsilon_j, \varepsilon_k|\tau_j = t_j, \tau_k = t_k) = \mathbb{E}(\varepsilon_j \varepsilon_k|\tau_j = t_j, \tau_k = t_k) - \mathbb{E}(\varepsilon_j|\tau_j = t_j)\mathbb{E}(\varepsilon_k|\tau_k = t_k) = 0 - 0 \cdot 0 = 0. \tag{2.33}$$

(3) Wir bestimmen zunächst $\mathbb{E}(\varepsilon_j\varepsilon_k)$. Mit dem Beweis von Aussage (2) des vorliegenden Theorems ergibt sich

$$\begin{aligned}
\mathbb{E}(\varepsilon_j\varepsilon_k) &= \mathbb{E}((y_j - \tau_j)(y_k - \tau_k)) \\
&= \sum_{t_j \in T_j} \sum_{t_k \in T_k} \sum_{\mathrm{y}_j \in Y_j} \sum_{\mathrm{y}_k \in Y_k} (\mathrm{y}_j - t_j)(\mathrm{y}_k - t_k) p(\mathrm{y}_j|t_j) p(\mathrm{y}_k|t_k) p(t_j, t_k) \\
&= \sum_{t_j \in T_j} \sum_{t_k \in T_k} \sum_{\mathrm{y}_j \in Y_j} \sum_{\mathrm{y}_k \in Y_k} (\mathrm{y}_j - \mathbb{E}(y_j|\tau_j = t_j))(\mathrm{y}_k - \mathbb{E}(y_k|\tau_k = t_k)) p(\mathrm{y}_j|t_j) p(\mathrm{y}_k|t_k) p(t_j, t_k) \\
&= \sum_{t_j \in T_j} \sum_{t_k \in T_k} \mathbb{E}\left((y_j - \tau_j)(y_k - \tau_k)|\tau_j = t_j, \tau_j = t_j\right) p(t_j, t_k) \\
&= \sum_{t_j \in T_j} \sum_{t_k \in T_k} \mathbb{E}\left(\varepsilon_j\varepsilon_k|\tau_j = t_j, \tau_j = t_j\right) p(t_j, t_k) \\
&= \sum_{t_j \in T_j} \sum_{t_k \in T_k} 0 \cdot p(t_j, t_k) \\
&= 0.
\end{aligned} \tag{2.34}$$

Mit dem Kovarianzverschiebungssatz (Theorem 4.16) und Aussage (2) des Theorems zu den Eigenschaften bezüglich einer Testmessung (Theorem 2.1) ergibt sich dann

$$\mathbb{C}(\varepsilon_j, \varepsilon_k) = \mathbb{E}(\varepsilon_j\varepsilon_k) - \mathbb{E}(\varepsilon_j)\mathbb{E}(\varepsilon_k) = 0 - 0 \cdot 0 = 0. \tag{2.35}$$

(4) Mit dem Kovarianzverschiebungssatz (Theorem 4.16) und Aussage (2) des Theorems zu den Eigenschaften bezüglich einer Testmessung (Theorem 2.1) gilt

$$\begin{aligned}
\mathbb{C}(\tau_j, \varepsilon_k) &= \mathbb{E}(\tau_j\varepsilon_k) - \mathbb{E}(\tau_j)\mathbb{E}(\varepsilon_k) \\
&= \mathbb{E}(\tau_j\varepsilon_k) - \mathbb{E}(\tau_j) \cdot 0 \\
&= \mathbb{E}(\tau_j(\mathrm{y}_k - \tau_k)) \\
&= \sum_{t_j \in T_j} \sum_{t_k \in T_k} \sum_{\mathrm{y}_k \in Y_k} t_j(y_k - t_k) p(\mathrm{y}_k|t_k) p(t_j, t_k) \\
&= \sum_{t_j \in T_j} t_j \sum_{t_k \in T_k} p(t_j, t_k) \sum_{\mathrm{y}_k \in Y_k} (y_k - t_k) p(\mathrm{y}_k|t_k) \\
&= \sum_{t_j \in T_j} t_j \sum_{t_k \in T_k} p(t_j, t_k) \left(\sum_{\mathrm{y}_k \in Y_k} y_k p(\mathrm{y}_k|t_k) - t_k \sum_{\mathrm{y}_k \in Y_k} p(\mathrm{y}_k|t_k) \right) \\
&= \sum_{t_j \in T_j} t_j \sum_{t_k \in T_k} p(t_j, t_k) \left(t_k - t_k \cdot 1\right) \\
&= \sum_{t_j \in T_j} t_j \sum_{t_k \in T_k} p(t_j, t_k) \cdot 0 \\
&= 0.
\end{aligned} \tag{2.36}$$

(5) Wir halten zunächst fest, dass mit Aussage (4) durch Vertauschen der Indizes und der Symmetrie der Kovarianz auch

$$\mathbb{C}(\tau_k, \varepsilon_j) = \mathbb{C}(\varepsilon_j, \tau_k) = 0 \tag{2.37}$$

gilt. Mit dem Theorem zur paarweisen Addition von Zufallsvariablen (Theorem 4.19) und Aussage (3) des vorliegenden Theorems gilt dann

$$\begin{aligned}
\mathbb{C}(y_j, y_k) &= \mathbb{C}(\tau_j + \varepsilon_j, \tau_k + \varepsilon_k) \\
&= \mathbb{C}(\tau_j, \tau_k) + \mathbb{C}(\tau_j, \varepsilon_k) + \mathbb{C}(\varepsilon_j, \tau_k) + \mathbb{C}(\varepsilon_j, \varepsilon_k) \\
&= \mathbb{C}(\tau_j, \tau_k) + 0 + 0 + 0 \\
&= \mathbb{C}(\tau_j, \tau_k).
\end{aligned} \tag{2.38}$$

□

Im Sinne des vereinfachten Modells der Klassischen Testtheorie nach Definition 2.3 werden die Aussagen von Theorem 2.3 oft auch als

(1) $\mathbb{C}(y_j, y_k|\tau_j = t_j, \tau_k = t_k) = 0$

(2) $\mathbb{C}(\varepsilon_j, \varepsilon_k | \tau_j = t_j, \tau_k = t_k) = 0$
(3) $\mathbb{C}(\varepsilon_j, \varepsilon_k) = 0$
(4) $\mathbb{C}(\tau_j, \varepsilon_k) = 0$
(5) $\mathbb{C}(y_j, y_k) = \mathbb{C}(\tau_j, \tau_k)$

geschrieben. Die ersten beiden Aussagen von Theorem 2.3 besagen, dass, bedingt auf den jeweiligen wahren Werten, die Kovarianzen von beobachteten Werten sowie die Kovarianzen der Messfehlert einer Person zwischen zwei Testmessungen gleich Null sind. Für die Messfehler gilt dies nach Aussage (3) von Theorem 2.3 auch im Sinne der unbedingten Marginalverteilung. Dies entspricht damit der paarweisen Unabhängigkeit von Messfehlren über Testmessungen im Modell multipler Testmessungen. Weiterhin ist nach Aussage (4) die Kovarianz des wahren Werts bei einer Testmessung mit dem Messfehler einer anderen Testmessung gleich Null. Schließlich gilt nach Aussage (5), dass die Kovarianz der beobachteten Werte zweier Testmessungen gleich der Kovarianz der wahren Werte ist.

Beispiel

Als erstes Beispiel für ein Modell multipler Testmessungen mit $m > 1$ setzen wir das Beispiel aus Theorem 2.2 fort und betrachten den Fall zweier Testmessungen $j = 1, 2$. Für $i = 1, ..., n$ seien entsprechend

$$\mathbb{P}(\tau_{i1}, \tau_{i2}) = \mathbb{P}(\tau_{i2} | \tau_{i1}) \mathbb{P}(\tau_{i1}) \tag{2.39}$$

mit

$$\mathbb{P}(\tau_{i1}) := N(1, 1) \text{ und } \mathbb{P}(\tau_{i2} | \tau_{i1}) := N(\tau_{i1} + 1, 1) \tag{2.40}$$

Die Verteilung des True-Score von Person i in Testmessung $j = 2$ hängt in diesem Beispiel also explizit von der Verteilung des wahren Werts von Person i in Testmessung $j = 1$ ab. Weiterhin seien

$$\mathbb{P}(y_{i1} | \tau_{i1}) := N(\tau_{i1}, 1) \text{ und } \mathbb{P}(y_{i2} | \tau_{i2}) := N(\tau_{i2}, 2) \tag{2.41}$$

Die Propensitätsverteilungen von Person i in Testmessung $j = 1$ unterscheide sich also von der von Person i in Testmessung $j = 2$. Folgender **R** Code generiert 10^5 Realisierungen dieses Modells und schätzt die in Theorem 2.3 betrachteten marginalen Kovarianzen $\mathbb{C}(\varepsilon_{i1}, \varepsilon_{i2})$, $\mathbb{C}(\tau_{i1}, \varepsilon_{i2})$, $\mathbb{C}(y_{i1}, y_{i2})$ und $\mathbb{C}(\tau_{i1}, \tau_{i2})$.

```
n          = 1e5                                  # Personenanzahl
m          = 2                                    # Testmessungsanzahl
mu         = 1                                    # Wahrer Werte Erwartungswertparameter
T          = matrix(rep(NaN, n*m), nrow = n)      # Wahrer Wert Array
Y          = matrix(rep(NaN, n*m), nrow = n)      # Beobachteter Wert Array
E          = matrix(rep(NaN, n*m), nrow = n)      # Messfehler Array
for(i in 1:n){                                    # Iteration über Personen
    T[i,1]  = rnorm(1,1,1)                        # Wahrer Wert Realisierung     für j = 1
    Y[i,1]  = rnorm(1,T[i,1],1)                   # Beobachteter Wert Realisierung für j = 1
    E[i,1]  = Y[i,1] - T[i,1]                     # Messfehler Realisierung    für j = 1
    T[i,2]  = rnorm(1,T[i,1] + 1,.5)              # Wahrer Wert Realisierung     für j = 2
    Y[i,2]  = rnorm(1,T[i,2],.5)                  # Beobachteter Wert Realisierung für j = 2
    E[i,2]  = Y[i,2] - T[i,2]}                    # Messfehler Realisierung    für j = 2
c_hat_e1_e2 = cov(E[,1],E[,2])                    # Kovarianzschätzung Messfehler  1, Messfehler 2
c_hat_t1_e2 = cov(T[,1],E[,2])                    # Kovarianzschätzung Wahrer Wert 1, Messfehler 2
c_hat_y1_y2 = cov(Y[,1],Y[,2])                    # Kovarianzschätzung Beobachteter Wert 1, Beobachteter Wert 2
c_hat_t1_t2 = cov(T[,1],T[,2])                    # Kovarianzschätzung Wahrer Wert 1, Wahrer Wert 2
```

Abbildung 2.2 visualisiert 500 der so generierten Zufallsvariablenrealisierungen und dokumentiert die resultierenden Kovarianzschätzer. Abbildung 2.2 A zeigt die marginale Unkorrelierheit der Messfehler bezüglich zweier Testmessungen (Aussage (2) von von Theorem 2.3), Abbildung 2.2 B zeigt die marginale Unkorrelierheit des wahren Wertes und des

Messfehlers bezüglich zweier Testmessungen (Aussage (3) von von Theorem 2.3) und Abbildung 2.2 C und Abbildung 2.2 D zeigen die Gleichheit der marginalen Kovarianzen von beobachteten und wahren Werten bezüglich zweier Testmessungen.

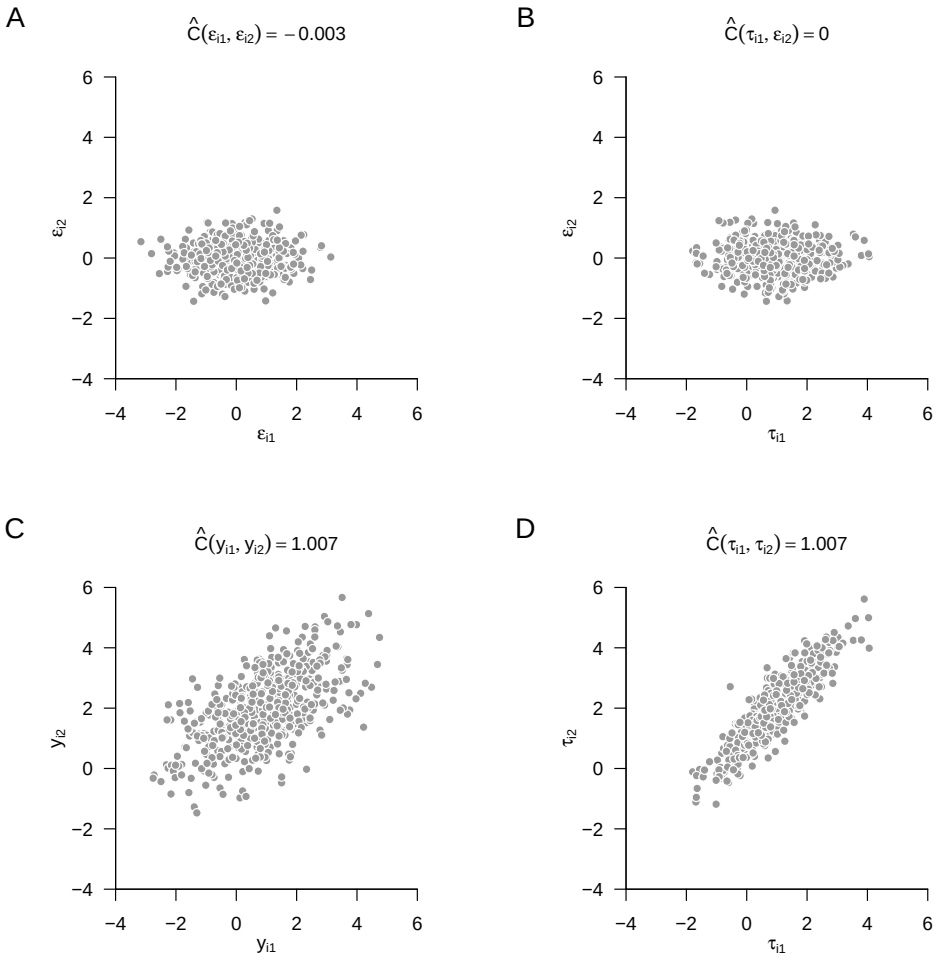

Abbildung 2.2 Normalverteilungsbeispiel bezüglich zweier Testmessungen.

2.1.2 Das Modell paralleler Testmessungen

Bisher haben wir im Modell der multiplen Testmessungen keine Aussage zu den Verhältnissen der wahren Werte über verschiedene Testmessungen hinweg gemacht. Wir haben einerseits angenommen, dass für die Marginalverteilung der Testmessungen bei einer Person i keine Unabhängigkeit gelten muss, dass also im Allgemeinen gilt, dass für $i = 1, ..., n$

$$\mathbb{P}(\tau_{i1}, ..., \tau_{im}) \neq \prod_{j=1}^{m} \mathbb{P}(\tau_{ij}). \tag{2.42}$$

Andererseits haben wir die Form möglicher Abhängigkeiten zwischen den wahren Werten $\tau_{i1}, ..., \tau_{im}$ bislang nicht genauer spezifiziert. Die Klassische Testtheorie betrachtet in dieser Hinsicht einige Spezialfälle, die sich allgemein durch funktionale Abhängigkeiten zwischen

τ_{i1} und $\tau_{i2}, ..., \tau_{im}$ der Form

$$\tau_{ij} = f(\tau_{i1}) \text{ für } f : \mathbb{R} \to \mathbb{R} \text{ und } j = 2, ..., m. \tag{2.43}$$

ausdrücken lassen. Wir betrachten im Folgenden den Fall, dass $f := \mathrm{id}_{\mathbb{R}}$, dass also insbesondere für Realisierungen t_{ij} von τ_{ij} gilt, dass

$$t_{ij} = \mathrm{id}_{\mathbb{R}}(t_{i1}) = t_{i1} \text{ für } j = 2, ..., m, \tag{2.44}$$

dass also die Werte der wahren Werte einer Person über Testmessungen identisch sind. Die Klassische Testtheorie bezeichnet solche Testmessungen als *parallele Testmessungen*. Eine weitere Form der funktionalen Abhängigkeit, die wir hier nicht weiter vertiefen wollen, ist der Fall, dass es sich bei f um eine linear-affine Funktion handelt, dass also

$$\tau_{ij} = f(\tau_{i1}) = a\tau_{i1} + b \text{ für } a, b \in \mathbb{R} \text{ und } j = 2, ..., m. \tag{2.45}$$

Die Klassische Testtheorie bezeichnet solche Testmessungen als *wesentlich τ-äquivalente Testmessungen*.

Definition 2.4 (Modell paralleler Testmessungen)**.** Für $i = 1, ..., n$ und $j = 1, ..., m$ seien τ_i eine Zufallsvariable, die den wahren Wert der $i$ten Person in jeder Testmessung $j = 1, ..., m$ modelliere, y_{ij} eine Zufallsvariable, die den beobachteten Wert der $i$ten Person in der $j$ten Testmessung modelliere und $\varepsilon_{ij} := y_{ij} - \tau_i$ die Zufallsvariable, die den Messfehlerder $i$ten Person in der $j$ten Testmessung modelliere. Dann heißt die gemeinsame Verteilung der τ_i und y_{ij} mit den Faktorisierungseigenschaften

$$\mathbb{P}(\tau_1, y_{11}, ...y_{1m}, ..., \tau_n, y_{n1}, ..., y_{nm}) := \prod_{i=1}^{n} \mathbb{P}(\tau_i) \prod_{j=1}^{m} \mathbb{P}(y_{ij}|\tau_i) \tag{2.46}$$

das *Modell paralleler Testmessungen*, wenn gilt, dass

(1) $\mathbb{P}(\tau_1) = \cdots = \mathbb{P}(\tau_n)$
(2) $\mathbb{P}(y_{1j}|\tau_1) = \cdots = \mathbb{P}(y_{nj}|\tau_n)$ für alle $1 \leq j \leq m$
(3) $\mathbb{E}(y_{ij}|\tau_i = t_i) = \mathbb{E}(y_{ik}|\tau_i = t_i) := t_i$ für alle $1 \leq i \leq n, 1 \leq j, k \leq m$
(4) $\mathbb{V}(y_{ij}|\tau_i = t_i) = \mathbb{V}(y_{ik}|\tau_i = t_i)$ für alle $1 \leq i \leq n, 1 \leq j, k \leq m$

•

Insbesondere werden also im Modell paralleler Testmessungen die Werte des wahren Werts einer Person werden über Testmessungen hinweg als identisch angenommen. Damit geht dann die Varianz der beobachteten Werte einer Person zwischen zwei Testmessungen allein auf die Propensitätsverteilung zurück. Aus generativer Sichtweise entstehen Werte der beobachteten Werte also wie folgt: Zunächst wird für die $i$te Person
und Testmessungen $j = 1, ..., m$ ein True-Score t_i von gemäß $\mathbb{P}(\tau_i)$ realisiert. Dann wird für die $i$te und die Testmessung $j = 1, ..., m$ ein Observed-Score y_{ij} anhand von $\mathbb{P}(y_{ij}|\tau_i = t_i)$ realisiert.

Eigenschaften des Modells paralleler Testmessungen

Betrachtet man nur eine einzige Testmessung, so hat das Modell paralleler Testmessungen natürlich die gleiche Form wie das Modell multipler Testmessungen. Damit gilt dann aber auch Theorem 2.1 analog für das Modell paralleler Testmessungen. Betrachtet man allerdings mehr als eine Testmessung, so ergeben sich für das Modell paralleler Testmessungen speziellere Eigenschaften, die wir in folgendem Theorem festhalten

Theorem 2.4 (Eigenschaften des Modells paralleler Testmessungen bezüglich zweier Testmessungen)**.** *Gegeben sei das Modell paralleler Testmessungen. Dann gelten für alle $i = 1, ..., n$ und alle j, k mit $1 \leq j, k \leq m$ und $j \neq k$, dass*

(1) $\mathbb{E}(y_{ij}) = \mathbb{E}(y_{ik})$
(2) $\mathbb{V}(y_{ij}) = \mathbb{V}(y_{ik})$
(3) $\mathbb{C}(\varepsilon_{ij}, \varepsilon_{ik}) = 0$
(4) $\mathbb{C}(\tau_i, \varepsilon_{ik}) = 0$
(5) $\mathbb{C}(y_{ij}, y_{ik}) = \mathbb{V}(\tau_i)$

∘

Beweis. Zum Beweis setzen zur Vereinfachung der Notation zunächst

$$y_j := y_{ij}, y_k := y_{ik}, \tau := \tau_i, \mathrm{y}_j := \mathrm{y}_{ij}, \mathrm{y}_k := \mathrm{y}_{ik}, t := t_i, Y_j := Y_{ij}, Y_k := Y_{ik} \text{ und } T := T_i \tag{2.47}$$

für alle $i = 1, ..., n$. Wir betrachten weiterhin nur den diskreten Fall, setzen also die Existenz einer Wahrscheinlichkeitsmassefunktion der Form

$$p(t, \mathrm{y}_j, \mathrm{y}_k) = p(\mathrm{y}_j|t)p(\mathrm{y}_k|t)p(t) \tag{2.48}$$

voraus. Der kontinuierliche Fall folgt dann analog.

(1) Mit der Gleichheit der bedingten Erwartungswerte im Falle paralleler Testmessungen gilt

$$\mathbb{E}(y_j) = \sum_{t \in T} \sum_{\mathrm{y}_j \in Y_j} \mathrm{y}_j p(\mathrm{y}_j|t)p(t) = \sum_{t \in T} \mathbb{E}(y_j|\tau = t)p(t) = \sum_{t \in T} \sum_{\mathrm{y}_k \in Y_k} \mathrm{y}_k p(\mathrm{y}_k|t)p(t) = \mathbb{E}(y_k). \tag{2.49}$$

(2) Mit dem Satz von der iterierten Varianz (Theorem 4.14) gilt

$$\mathbb{V}(y_j) = \mathbb{V}\left(\mathbb{E}(y_j|\tau)\right) + \mathbb{E}\left(\mathbb{V}(y_j|\tau)\right) = \mathbb{V}\left(\mathbb{E}(y_k|\tau)\right) + \mathbb{E}\left(\mathbb{V}(y_k|\tau)\right) = \mathbb{V}(y_k). \tag{2.50}$$

(3) Wir bestimmen zunächst $\mathbb{E}(\varepsilon_j \varepsilon_k)$. Mit Aussage (2) von Theorem 2.1 gilt dann

$$\begin{aligned}
\mathbb{E}(\varepsilon_j \varepsilon_k) &:= \mathbb{E}((y_j - \tau)(y_k - \tau)) \\
&= \sum_{t \in T} \sum_{\mathrm{y}_j \in Y_j} \sum_{\mathrm{y}_k \in Y_k} (\mathrm{y}_j - t)(\mathrm{y}_k - t)p(t)p(\mathrm{y}_j|t)p(\mathrm{y}_k|t) \\
&= \sum_{t \in T} p(t) \sum_{\mathrm{y}_j \in Y_j} \sum_{\mathrm{y}_k \in Y_k} (\mathrm{y}_j - t)(\mathrm{y}_k - t)p(\mathrm{y}_j|t)p(\mathrm{y}_k|t) \\
&= \sum_{t \in T} p(t) \sum_{\mathrm{y}_j \in Y_j} (\mathrm{y}_j - t)p(\mathrm{y}_j|t) \sum_{\mathrm{y}_k \in Y_k} (\mathrm{y}_k - t)p(\mathrm{y}_k|t) \\
&= \sum_{t \in T} p(t) \left(\sum_{\mathrm{y}_j \in Y_j} \mathrm{y}_j p(\mathrm{y}_j|t) - t \sum_{\mathrm{y}_j \in Y_j} p(\mathrm{y}_j|t) \right) \left(\sum_{\mathrm{y}_k \in Y_k} \mathrm{y}_k p(\mathrm{y}_k|t) - t \sum_{\mathrm{y}_k \in Y_k} p(\mathrm{y}_k|t) \right) \\
&= \sum_{t \in T} p(t) \left(t - t\right) \left(t - t\right) \\
&= \sum_{t \in T} p(t) \cdot 0 \cdot 0 \\
&= 0.
\end{aligned} \tag{2.51}$$

Mit dem Kovarianzverschiebungssatz (Theorem 4.16) und wiederrum mit Aussage (2) von Theorem 2.1 folgt dann

$$\mathbb{C}(\varepsilon_j, \varepsilon_k) = \mathbb{E}(\varepsilon_j \varepsilon_k) - \mathbb{E}(\varepsilon_j)\mathbb{E}(\varepsilon_k) = 0 - 0 \cdot 0 = 0 \tag{2.52}$$

(4) Mit dem Kovarianzverschiebungssatz (Theorem 4.16) und Aussage (2) von Theorem 2.1 gilt

$$\begin{aligned}
\mathbb{C}(\tau, \varepsilon_k) &= \mathbb{E}(\tau\varepsilon_k) - \mathbb{E}(\tau)\mathbb{E}(\varepsilon_k) \\
&= \mathbb{E}(\tau\varepsilon_k) - \mathbb{E}(\tau) \cdot 0 \\
&= \mathbb{E}(\tau(y_k - \tau)) \\
&= \sum_{t \in T} \sum_{\mathrm{y}_k \in Y_k} t(\mathrm{y}_k - t) p(t) p(\mathrm{y}_k | t) \\
&= \sum_{t \in T} tp(t) \sum_{\mathrm{y}_k \in Y_k} (\mathrm{y}_k - t) p(\mathrm{y}_k | t) \\
&= \sum_{t \in T} tp(t) \left(\sum_{\mathrm{y}_k \in Y_k} \mathrm{y}_k p(\mathrm{y}_k | t) - t \sum_{\mathrm{y}_k \in Y_k} p(\mathrm{y}_k | t) \right) \\
&= \sum_{t \in T} tp(t) \, (t - t) \\
&= \sum_{t \in T} tp(t) \cdot 0 \\
&= 0.
\end{aligned} \tag{2.53}$$

(5) Wir halten zunächst fest, dass mit Aussage (2) des Theorems neben $\mathbb{C}(\tau, \varepsilon_k) = 0$ durch Austausch des Index und der Symmetrie der Kovarianz auch

$$\mathbb{C}(\tau, \varepsilon_j) = \mathbb{C}(\varepsilon_j, \tau) = 0 \tag{2.54}$$

gilt. Mit dem Theorem zur Kovarianz bei paarweiser Addition von Zufallsvariablen (Theorem 4.19) und Aussagen (3) und (4) des vorliegenden Theorems gilt dann

$$\begin{aligned}
\mathbb{C}(y_j, y_k) &= \mathbb{C}(\tau + \varepsilon_j, \tau + \varepsilon_k) \\
&= \mathbb{C}(\tau, \tau) + \mathbb{C}(\tau, \varepsilon_k) + \mathbb{C}(\varepsilon_j, \tau) + \mathbb{C}(\varepsilon_j, \varepsilon_k) \\
&= \mathbb{C}(\tau, \tau) + 0 + 0 + 0 \\
&= \mathbb{C}(\tau, \tau) \\
&= \mathbb{V}(\tau).
\end{aligned} \tag{2.55}$$

□

Aussage (1) von Theorem 2.4 besagt, dass bei Paralleltestmessungen alle Erwartungswerte der beobachteten Werte identisch sind und Aussage (2) des Theorems besagt, dass bei Paralleltestmessungen alle Varianzen der beobachteten Werte identisch sind. Die Aussagen (3) und (4) von Theorem 2.4 sind analog zu den lokalen Unkorreliertheitseigenschaften des Modells multipler Testmessungen. Aussage (5) schließlich besagt insbesondere, dass $\mathbb{C}(y_{ij}, y_{ik})$ für beliebe j und k identisch zu $\mathbb{V}(\tau_i)$ sind. Zusammen mit Aussage (2) des Theorems sind im Modell paralleler Testmessungen also alle paarweisen Korrelationen verschiedener Testmessungen identisch.

Beispiel

Wir betrachten den Fall zweier Testmessungen $j = 1, 2$ im Modell paralleler Testmessungen. Für $i = 1, ..., n$ seien

$$\mathbb{P}(\tau_i) = N(1, 1) \text{ und } \mathbb{P}(y_{i1} | \tau_i) := \mathbb{P}(y_{i2} | \tau_i) := N(\tau_i, 1) \tag{2.56}$$

Für Person i gibt es also nur eine True-Score Zufallsvariable für alle Testmessungen und die Propensitätsverteilungen unterscheiden sich zwischen Testmessungen nicht.

```r
n           = 1e5                               # Personenanzahl
m           = 2                                 # Testmessungsanzahl
T           = matrix(rep(NaN, n)  , nrow = n)   # Wahrer Wert Array
Y           = matrix(rep(NaN, n*m), nrow = n)   # Beobachteter Wert Array
E           = matrix(rep(NaN, n*m), nrow = n)   # Messfehler Array
for(i in 1:n){                                  # Personeniterationen
    T[i]  = rnorm(1,1,1)                        # Wahrer Wert Realisierung für j = 1,2
    for(j in 1:m){                              # Testmessungsiterationen
        Y[i,j]  = rnorm(1,T[i],1)               # Beobachteter Wert Realisierung f
        E[i,j]  = Y[i,j] - T[i]}}               # Messfehler Realisierung
e_hat_o1_o2 = apply(Y, 2, mean)                 # Erwartungswertschätzung Beobachteter Wert 1, Beobachteter Wert 2
v_hat_o1_o2 = apply(Y, 2, var)                  # Varianzzschätzung Beobachteter Wert 1, Beobachteter Wert 2
c_hat_e1_e2 = cov(E[,1],E[,2])                  # Kovarianzzschätzung Messfehler 1, Messfehler 2
c_hat_t_e2  = cov(T    ,E[,2])                  # Kovarianzzschätzung Wahrer Wert 1, Messfehler 2
c_hat_y1_y2 = cov(Y[,1],Y[,2])                  # Kovarianzzschätzung Beobachteter Wert 1, Beobachteter Wert 2
v_hat_t     = var(T)                            # Varianzschätzung Wahrer Wert
```

Abbildung 2.3 visualisiert 500 der so generierten Zufallsvariablenrealisierungen und dokumentiert die resultierenden Kovarianzschätzer. Abbildung 2.3 A zeigt die Unkorrelierheit der Messfehler bezüglich zweier paralleler Testmessungen (Aussage (3) von von Theorem 2.4), Abbildung 2.3 B zeigt die Unkorrelierheit des wahren Wertes und des Messfehlers bezüglich zweier paralleler Testmessungen (Aussage (4) von Theorem 2.4) und Abbildung 2.3 C zeigt die durch die identischen wahren Werte induzierte Korrelation zwischen den marginalen Kovarianzen der beobachteten Werten zweier paralleler Testmessungen.

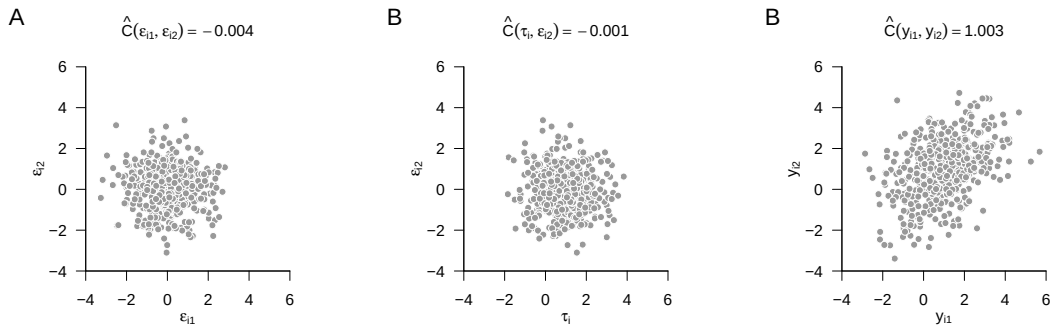

Abbildung 2.3 Normalverteilungsbeispiel bezüglich zweier paralleler Testmessungen.

2.2 Reliabilität

Die *Reliabilität* einesr Testmessung ist das zentrale Konzept der Klassischen Testtheorie. Wir folgen hier dem Ansatz nach Lord & Novick (1968), wonach die Reliabilität einer Testmessung als quadrierte Korrelation von beobachtetem und wahrem Wert definiert ist. Basierend auf den Eigenschaften des Modells multipler Testmessungen ergeben sich dann verschiedene Möglichkeiten, diese Korrelation darzustellen und damit verschiedene Möglichkeiten, die Reliabilität einer Testmessung zu interpretieren. Allerdings ergibt sich dabei keine Möglichkeit, die Reliabilität von Testmessungen empirisch zu messen. Zentral ist dann die Übertragung des Konzepts der Reliabilität in das Modell paralleler Testmessungen. Die so definierte *Paralleltestreliabilität* ist dann empirisch schätzbar.

2.2.1 Reliabilität einer Testmessung

Wir beginnen zunächst mit der Definition der Reliabilität einer Testmessung im Modell multipler Testmessungen.

Definition 2.5 (Reliabilität einer Testmessung)**.** Gegeben sei das Modell multipler Testmessungen für eine beliebige Testmessung j mit $1 \leq j \leq m$,

$$\mathbb{P}(\tau_{1j}, y_{1j} ..., \tau_{nj}, y_{nj}) := \prod_{i=1}^{n} \mathbb{P}(\tau_{ij}, y_{ij}) := \prod_{i=1}^{n} \mathbb{P}(y_{ij}|\tau_{ij})\mathbb{P}(\tau_{ij}) \tag{2.57}$$

wobei nach Definition des Modells gilt, dass

$$\mathbb{P}(\tau_{1j}, y_{1j}) = \cdots = \mathbb{P}(\tau_{nj}, y_{nj}). \tag{2.58}$$

Die *Reliabilität der Testmessung j* ist dann definiert als

$$\mathrm{R}_j := \rho(y_{ij}, \tau_{ij})^2 \text{ für ein beliebiges } 1 \leq i \leq n. \tag{2.59}$$

•

Vor dem Hintergrund des vereinfachen Modells der Klassischen Testtheorie Definition 2.3 schreibt man auch

$$\mathrm{R} = \rho(y, \tau)^2. \tag{2.60}$$

Per Definition ist die Reliabilität eine Testmessung also die quadrierte Korrelation von beobachtetem und wahrem Wert. Mit

$$-1 \leq \rho(y_{ij}, \tau_{ij}) \leq 1 \tag{2.61}$$

folgt direkt, dass für die Reliabilität einer Testmessung gilt, dass

$$0 \leq \mathrm{R}_j \leq 1 \tag{2.62}$$

Die Aussage $\mathrm{R}_j = 0$ impliziert dann $\rho(y_{ij}, \tau_{ij}) = 0$, also die linear-affine Unabhängigkeit von beobachtetem und wahrem Wert, und damit den Umstand, dass der beobachtete Wert hinsichtlich des wahren Werts nicht aussagekräftig ist. $\mathrm{R}_j = 1$ dagegen impliziert $\rho(y_{ij}, \tau_{ij}) = \pm 1$, also die deterministische linear-affine Abhängigkeit von beobachtetem und wahrem Wert und damit, dass der beobachtete Wert hinsichtlich des wahren Werts vollständig aussagekräftig ist.

Folgendes Theorem zeigt weitere Möglichkeiten auf, die Reliabilität einer Testmessung äquivalent zu formulieren und damit zu interpretieren.

Theorem 2.5 (Eigenschaften der Reliabilität einer Testmessung)**.** *Gegeben sei das Modell multipler Testmessungen für eine beliebige Testmessung j mit $1 \leq j \leq m$,*

$$\mathbb{P}(\tau_{1j}, y_{nj} ..., \tau_{nj}, y_{nj}) := \prod_{i=1}^{n} \mathbb{P}(y_{ij}|\tau_{ij})\mathbb{P}(\tau_{ij}). \tag{2.63}$$

Dann gelten für die Reliabilät R_j der Testmessung j, dass

(1) $R_j = \frac{\mathbb{V}(\tau_{ij})}{\mathbb{V}(y_{ij})}$
(2) $R_j = 1 - \frac{\mathbb{V}(\varepsilon_{ij})}{\mathbb{V}(y_{ij})}$

∘

Beweis. (1) Mit Aussage (5) von Theorem 2.1 gilt

$$\mathrm{R}_j = \rho(y_{ij}, \tau_{ij})^2 = \left(\frac{\mathbb{C}(y_{ij}, \tau_{ij})}{\mathbb{S}(y_{ij})\mathbb{S}(\tau_{ij})} \right)^2 = \frac{\mathbb{C}(y_{ij}, \tau_{ij})^2}{\mathbb{V}(y_{ij})\mathbb{V}(\tau_{ij})} = \frac{\mathbb{V}(\tau_{ij})^2}{\mathbb{V}(y_{ij})\mathbb{V}(\tau_{ij})} = \frac{\mathbb{V}(\tau_{ij})}{\mathbb{V}(y_{ij})}. \tag{2.64}$$

(2) Mit Aussage (4) von Theorem 2.1 gilt dann weiter

$$\mathrm{R}_j = \frac{\mathbb{V}(\tau_{ij})}{\mathbb{V}(y_{ij})} = \frac{\mathbb{V}(y_{ij}) - \mathbb{V}(\varepsilon_{ij})}{\mathbb{V}(y_{ij})} = \frac{\mathbb{V}(y_{ij})}{\mathbb{V}(y_{ij})} - \frac{\mathbb{V}(\varepsilon_{ij})}{\mathbb{V}(y_{ij})} = 1 - \frac{\mathbb{V}(\varepsilon_{ij})}{\mathbb{V}(y_{ij})}. \tag{2.65}$$

□

Da τ_{ij} und ε_{ij} weiterhin latent und damit nur indirekt beobachtbar sind, sind die Darstellungen nach Theorem 2.5 nur von theoretischem Interesse. Die erste Aussage besagt dabei insbesondere, dass die Reliabilität einer Testmessung der Anteil der Varianz der wahren Werte an der Varianz der beobachteten Werte ist,

$$\mathrm{R}_j = \frac{\mathbb{V}(\tau_{ij})}{\mathbb{V}(y_{ij})} = \frac{\mathbb{V}(\tau_{ij})}{\mathbb{V}(\tau_{ij}) + \mathbb{V}(\varepsilon_{ij})}. \tag{2.66}$$

Gilt dabei, dass $\mathbb{V}(\tau_{ij}) = 0$, so ist auch $\mathrm{R}_j = 0$, ist dagegen $\mathbb{V}(\varepsilon_{ij}) = 0$ so ist $\mathrm{R}_j = 1$. Eine Reliabilität $\mathrm{R}_j > 0$ impliziert also immer eine von Null verschiedene Varianz der wahren Werte.

2.2.2 Paralletestreliabilität

Um das Konzept der Reliabilität nun anhand empirischer Daten von beobachteten Werten schätzbar zu machen, bedarf es seiner Übertragung in das Modell paralleler Testmessungen. Wir definieren daher zunächst explizit die Reliabilität einer Paralleltestmessung.

Definition 2.6 (Reliabilität einer Paralleltestmessung)**.** Gegeben sei das Modell paralleler Testmessungen für eine beliebige Testmessung j mit $1 \leq j \leq m$,

$$\mathbb{P}(\tau_1, y_{1j} ..., \tau_n, y_{nj}) := \prod_{i=1}^{n} \mathbb{P}(\tau_i, y_{ij}) = \prod_{i=1}^{n} \mathbb{P}(\tau_i)\mathbb{P}(y_{ij}|\tau_i) \tag{2.67}$$

wobei nach Definition des Modells gilt, dass

$$\mathbb{P}(\tau_1, y_{1j}) = \cdots = \mathbb{P}(\tau_n, y_{nj}). \tag{2.68}$$

Die *Reliabilität der Paralleltestmessung* j ist dann definiert als

$$\mathrm{R}_j := \rho(y_{ij}, \tau_i)^2 \text{ für ein beliebiges } 1 \leq i \leq n. \tag{2.69}$$

•

Vor dem Hintergrund des vereinfachten Modells der Klassischen Testtheorie (Definition 2.3) schreibt man auch hier

$$\mathrm{R} = \rho(y, \tau)^2. \tag{2.70}$$

Praktisch bedeutsam ist nun folgendes Theorem.

Theorem 2.6 (Paralleltestreliabilität)**.** *Gegeben sei das Modell paralleler Testmessungen*

$$\mathbb{P}(\tau_1, y_{1j}..., \tau_n, y_{nj}) := \prod_{i=1}^{n} \mathbb{P}(\tau_i, y_{ij}) = \prod_{i=1}^{n} \mathbb{P}(\tau_i)\mathbb{P}(y_{ij}|\tau_i) \tag{2.71}$$

Dann gelten

(1) $R_j = \frac{\mathbb{V}(\tau_i)}{\mathbb{V}(y_{ij})}$ *für alle* $1 \leq j \leq m$.
(2) $R_j = 1 - \frac{\mathbb{V}(\varepsilon_{ij})}{\mathbb{V}(y_{ij})}$ *für alle* $1 \leq j \leq m$.
(3) $R_j = \rho(y_{ij}, y_{ik}) = R_k$ *für alle* $1 \leq j, k \leq m$.

∘

Beweis. (1) Mit Aussage (5) von Theorem 2.1 gilt

$$\mathrm{R}_j = \rho(y_{ij}, \tau_i)^2 = \left(\frac{\mathbb{C}(y_{ij}, \tau_i)}{\mathbb{S}(y_{ij})\mathbb{S}(\tau_i)}\right)^2 = \frac{\mathbb{C}(y_{ij}, \tau_i)^2}{\mathbb{V}(y_{ij})\mathbb{V}(\tau_i)} = \frac{\mathbb{V}(\tau_i)^2}{\mathbb{V}(y_{ij})\mathbb{V}(\tau_i)} = \frac{\mathbb{V}(\tau_i)}{\mathbb{V}(y_{ij})}. \tag{2.72}$$

(2) Mit Aussage (4) von Theorem 2.1 gilt dann weiter

$$\mathrm{R}_j = \frac{\mathbb{V}(\tau_i)}{\mathbb{V}(y_{ij})} = \frac{\mathbb{V}(y_{ij}) - \mathbb{V}(\varepsilon_{ij})}{\mathbb{V}(y_{ij})} = \frac{\mathbb{V}(y_{ij})}{\mathbb{V}(y_{ij})} - \frac{\mathbb{V}(\varepsilon_{ij})}{\mathbb{V}(y_{ij})} = 1 - \frac{\mathbb{V}(\varepsilon_{ij})}{\mathbb{V}(y_{ij})}. \tag{2.73}$$

(3) Mit Aussagen (2) und (5) von Theorem 2.3 gilt dann weiter

$$\mathrm{R}_j = \frac{\mathbb{V}(\tau_i)}{\mathbb{V}(y_{ij})} = \frac{\mathbb{C}(y_{ij}, y_{ik})}{\sqrt{\mathbb{V}(y_{ij})}\sqrt{\mathbb{V}(y_{ij})}} = \frac{\mathbb{C}(y_{ij}, y_{ik})}{\sqrt{\mathbb{V}(y_{ij})}\sqrt{\mathbb{V}(y_{ik})}} = \rho(y_{ij}, y_{ik}) \tag{2.74}$$

und dass ebenso

$$\mathrm{R}_k = \frac{\mathbb{V}(\tau_i)}{\mathbb{V}(y_{ik})} = \frac{\mathbb{C}(y_{ij}, y_{ik})}{\sqrt{\mathbb{V}(y_{ik})}\sqrt{\mathbb{V}(y_{ik})}} = \frac{\mathbb{C}(y_{ij}, y_{ik})}{\sqrt{\mathbb{V}(y_{ij})}\sqrt{\mathbb{V}(y_{ik})}} = \rho(y_{ij}, y_{ik}). \tag{2.75}$$

□

Aussagen (1) und (2) von Theorem 2.6 sind analog zu den Aussagen von Theorem 2.5. Aussage (3) von Theorem 2.6 begründet das praktische Vorgehen zur Reliabilitätsschätzung mithilfe von Parallel- oder Retestverfahren. Zur Bestimmung der Reliabilität eines Tests nutzt man entspreched eine Schätzung der Korrelation zweier paralleler Testmessungen und die Klassische Testtheorie begründet dieses Vorgehen vor der Annahme von von latenten wahren Werten und Messfehlern. Vereinfacht wird Aussage (3) von Theorem 2.6 oft dadurch ausgedrückt, dass man sagt, dass "die Korrelation paralleler Testmessungen gleich ihrer Reliabilität ist".

Beispiel

Wir betrachten den Fall zweier Testmessungen $j = 1, 2$ im Modell paralleler Testmessungen. Für alle $i = 1, ..., n$ seien, wobei wir der notationellen Einfachheit halber auf das i Subskript verzichten wollen,

$$p(t) = N(t; 0, \sigma_\tau^2) \text{ und } p(\mathrm{y}_1|t) := N(\mathrm{y}_1; t, \sigma_\varepsilon^2) \text{ und } p(\mathrm{y}_2|t) := N(\mathrm{y}_2; t, \sigma_\varepsilon^2) \tag{2.76}$$

Für Person i gibt es also nur eine Zufallsvariabl für den wahren Wert für alle Testmessungen und die Propensitätsverteilungen unterscheiden sich zwischen Testmessungen nicht. Dann

gilt zunächst mit dem Theorem zu gemeinsamen Normalverteilungen (Theorem 4.27) mit $A := 1$ und $b := 0$

$$p(t)p(\mathrm{y}_1|t) = p(t, \mathrm{y}_1) = N\left(\begin{pmatrix} t \\ \mathrm{y}_1 \end{pmatrix}; \begin{pmatrix} \mu_\tau \\ \mu_\tau \end{pmatrix}, \begin{pmatrix} \sigma_\tau^2 & \sigma_\tau^2 \\ \sigma_\tau^2 & \sigma_\tau^2 + \sigma_\varepsilon^2 \end{pmatrix}\right) \tag{2.77}$$

und weiterhin mit $A := \begin{pmatrix} 1 & 0 \end{pmatrix}$ und $b := 0$

$$p(t, \mathrm{y}_1)p(\mathrm{y}_2|t) = p(t, \mathrm{y}_1, \mathrm{y}_2) = N\left(\begin{pmatrix} t \\ \mathrm{y}_1 \\ \mathrm{y}_2 \end{pmatrix}; \begin{pmatrix} \mu_\tau \\ \mu_\tau \\ \mu_\tau \end{pmatrix}, \begin{pmatrix} \sigma_\tau^2 & \sigma_\tau^2 & \sigma_\tau^2 \\ \sigma_\tau^2 & \sigma_\tau^2 + \sigma_\varepsilon^2 & \sigma_\tau^2 \\ \sigma_\tau^2 & \sigma_\tau^2 & \sigma_\tau^2 + \sigma_\varepsilon^2 \end{pmatrix}\right) \tag{2.78}$$

Damit gilt dann durch Ablesen am Kovarianzmatrixparameter von $p(t, \mathrm{y}_1, \mathrm{y}_2)$, dass

$$\rho(y_1, y_2) = \frac{\mathbb{C}(y_1, y_2)}{\sqrt{\mathbb{V}(y_1)}\sqrt{\mathbb{V}(y_2)}} = \frac{\sigma_\tau^2}{\sqrt{\sigma_\tau^2 + \sigma_\varepsilon^2}\sqrt{\sigma_\tau^2 + \sigma_\varepsilon^2}} = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_\varepsilon^2}. \tag{2.79}$$

Insbesondere gilt also auch für $j = 1, 2$

$$\mathrm{R}_j = \rho(y_j, \tau)^2 = \left(\frac{\mathbb{C}(y_j, \tau)}{\mathbb{S}(y_j)\mathbb{S}(\tau)}\right)^2 = \frac{\mathbb{C}(y_j, \tau)^2}{\mathbb{V}(y_j)\mathbb{V}(\tau)} = \frac{\left(\sigma_\tau^2\right)^2}{(\sigma_\tau^2 + \sigma_\varepsilon^2)\sigma_\tau^2} = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_\varepsilon^2} = \rho(y_1, y_2). \tag{2.80}$$

Gilt also beispielsweise $\sigma_\tau^2 := 1.0$ und $\sigma_\varepsilon^2 := 0.2$, so ergibt sich für die Paralleltestreliabilität

$$\mathrm{R}_j = \rho(y_j, \tau)^2 = \rho(y_1, y_2) = \frac{1.0}{1.0 + 0.2} \approx 0.83. \tag{2.81}$$

2.2.3 Schätzung der Paralleltestreliabilität

Wie üblich wird die Korrelation zweier paralleler Testmessungen in der Anwendung durch eine Stichprobenkorrelation geschätzt. Wir formulieren dies für das vorliegende Szenario mithilfe folgender Definition.

Definition 2.7 (Paralleltestreliabilitätsschätzer)**.** Gegeben sei das Modell paralleler Testmessungen für n Personen und zwei Testmessungen $j = 1, 2$,

$$\mathbb{P}(\tau_1, y_{11}, y_{12}, ..., \tau_n, y_{n1}, y_{n2}) = \prod_{i=1}^{n} \mathbb{P}(\tau_i)\mathbb{P}(y_{i1}|\tau_i)\mathbb{P}(y_{i2}|\tau_i). \tag{2.82}$$

Dann wird der mit den Stichprobenmitteln

$$\bar{y}_1 := \frac{1}{n}\sum_{i=1}^{n} y_{i1} \text{ und } \bar{y}_2 := \frac{1}{n}\sum_{i=1}^{n} y_{i2} \tag{2.83}$$

definierte Stichprobenkorrelationskoeffizient

$$r_{12} := \frac{\frac{1}{n-1}\sum_{i=1}^{n}(y_{i1} - \bar{y}_1)(y_{i2} - \bar{y}_2)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_{i1} - \bar{y}_1)^2}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_{i2} - \bar{y}_2)^2}} \tag{2.84}$$

Paralleltestreliabilitätsschätzer genannt.

•

Beispiel

Bekanntlich bietet **R** mit der `cor()` Funktion eine Möglichkeit, Stichprobenkorrelationskoeffizienten zu berechnen. Wir demonstrieren dies an einem Simulationsbeispiel. Dazu seien

$$p(t_i) = N(t_i; 0, \sigma_\tau^2) \text{ und } p(y_{ij}|t_i) := N(y_{ij}; t_i, \sigma_\varepsilon^2) \tag{2.85}$$

für $j = 1, 2$ mit $\sigma_\tau^2 := 1.0$ und $\sigma_\varepsilon^2 := 0.2$ für $i = 1, ..., 30$. Die wahre, aber unbekannte, Paralleltestreliabilität ergibt sich hier zu

$$R_{12} = \frac{1.0}{1.0 + 0.2} \approx 0.833. \tag{2.86}$$

Anhand von $n = 30$ simulierten Datenpunkten wird diese in folgender Simulation als $r_{12} = 0.857$ geschätzt.

```
set.seed(1)
n           = 30                                                  # Personenanzahl
m           = 2                                                   # Testmessungsanzahl
sigsqr_tau  = 1                                                   # Wahrer Wert Varianz
sigsqr_eps  = .2                                                  # Beobachteter Wert Varianz
R_12        = sigsqr_tau/(sigsqr_tau+sigsqr_eps)                  # Paralletestreliabilität
T           = matrix(rep(NaN, n)  , nrow = n)                     # Wahrer Wert Array
Y           = matrix(rep(NaN, n*m), nrow = n)                     # Beobachteter Wert Array
E           = matrix(rep(NaN, n*m), nrow = n)                     # Messfehler Array
for(i in 1:n){                                                    # Personeniterationen
    T[i]  = rnorm(1,1,sqrt(sigsqr_tau))                           # Wahrer Wert Realisierung für j = 1,2
    for(j in 1:m){                                                # Testmessungsiterationen
        Y[i,j]  = rnorm(1,T[i],sqrt(sigsqr_eps))                  # Beobachteter Wert Realisierung
        E[i,j]  = Y[i,j] - T[i]}}                                 # Messfehler Realisierung
r_12      = cor(Y[,1],Y[,2])                                      # Paralleltestreliabilitätsschätzer
cat("Paralleltestreliabilität R_12           : ", round(R_12,4),  # Ausgabe w.a.u. Paralleltestreliabilität
    "\nParalleltestrelibabilitätsschätzer  r_12 : ",  round(r_12,4))  # Ausgabe geschätzte Paralleltestreliabilität
```

```
Paralleltestreliabilität R_12           :  0.8333
Paralleltestrelibabilitätsschätzer  r_12 :  0.8569
```

Um neben einem Punktschätzer auch ein Konfidenzintervall für die Paralleltestreliabilät anzugeben, ist eine Annahme zur empirischen Verteilung des Reliabilitätsschätzers erforderlich. Dazu nutzt man üblicherweise folgende Aussage zur asymptotischen Verteilung einer Stichprobenkorrelation.

Theorem 2.7 (Approximative Verteilung der Fishertransformation einer Stichprobenkorrelation)**.** *Gegeben sei eine Stichprobe* $(y_{11}, y_{12}), ..., (y_{n1}, y_{n2}) \sim \mathbb{P}(y_1, y_2)$ *von Beobachtungen zweier Zufallsvariablen* y_1 *und* y_2 *mit Korrelation* $\rho := \rho(y_1, y_2)$ *und Stichprobenkorrelation* r*. Weiterhin bezeichne*

$$\tilde{r} := \frac{1}{2} \ln \left(\frac{1 + r}{1 - r} \right) \tag{2.87}$$

die Fishertransformation *von* r*. Dann ist* $\tilde{r}$ *asymptotisch normalverteilt mit*

$$\tilde{r} \overset{a}{\sim} N \left(\frac{1}{2} \ln \left(\frac{1 + \rho}{1 - \rho} \right), (n - 3)^{-1} \right). \tag{2.88}$$

∘

Wir verzichten auf einen Beweis von Theorem 2.7 und verweisen für eine ausführliche Darstellung auf Kapitel 32 in Johnson et al. (1994). Theorem 2.7 bildet die Grundlage zur Konstruktion des in folgendem Theorem angegebenen approximativen Konfidenzintervalls für eine Korrelation.

Theorem 2.8 (Approximatives Konfidenzintervall einer Korrelation)**.** *Gegeben sei eine Stichprobe* $(y_{11}, y_{12}), ..., (y_{n1}, y_{n2}) \sim \mathbb{P}(y_1, y_2)$ *von Beobachtungen zweier Zufallsvariablen* y_1 *und* y_2 *mit Korrelation* ρ*, Stichprobenkorrelation* r *und Fishertransformation* $\tilde{r}$*. Ferner sei* $\delta \in]0, 1[$ *ein Konfidenzlevel und es sei*

$$z_\delta := \phi^{-1}\left(\frac{1+\delta}{2}\right) \tag{2.89}$$

mit der inversen KVF ϕ^{-1} *einer standardnormalverteilten Zufallsvariable. Schließlich seien*

$$\tilde{r}_u := \tilde{r} - z_\delta\left(\sqrt{n-3}\right)^{-1} \text{ und } \tilde{r}_o := \tilde{r} + z_\delta\left(\sqrt{n-3}\right)^{-1}. \tag{2.90}$$

Dann gilt für $n \to \infty$ *für das Intervall*

$$\kappa(r) := \left[\frac{\exp(2\tilde{r}_u) - 1}{\exp(2\tilde{r}_u) + 1}, \frac{\exp(2\tilde{r}_o) - 1}{\exp(2\tilde{r}_o) + 1}\right], \tag{2.91}$$

dass

$$\mathbb{P}_\rho\left(\kappa(r) \ni \rho\right) = \delta. \tag{2.92}$$

∘

Beweis. Mit Theorem 2.7 gilt durch Z-Transformation von r zunächst für $n \to \infty$, dass

$$\tilde{r}_z := \left(\tilde{r} - \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)\right)\sqrt{(n-3)} \overset{a}{\sim} N(0,1) \tag{2.93}$$

und damit asymptotisch, dass

$$\mathbb{P}(-z_\delta \leq \tilde{r}_z \leq z_\delta) = \delta. \tag{2.94}$$

Also gilt asymptotisch auch, dass

$$\begin{aligned}
\delta &= \mathbb{P}\left(-z_\delta \leq \tilde{r}_z \leq z_\delta\right) \\
&= \mathbb{P}\left(-z_\delta \leq \left(\tilde{r} - \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)\right)\sqrt{(n-3)} \leq z_\delta\right) \\
&= \mathbb{P}\left(-z_\delta\left(\sqrt{(n-3)}\right)^{-1} \leq \tilde{r} - \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right) \leq z_\delta\left(\sqrt{(n-3)}\right)^{-1}\right) \\
&= \mathbb{P}\left(-z_\delta\left(\sqrt{(n-3)}\right)^{-1} - \tilde{r} \leq -\frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right) \leq z_\delta\left(\sqrt{(n-3)}\right)^{-1} - \tilde{r}\right) \\
&= \mathbb{P}\left(\tilde{r} + z_\delta\left(\sqrt{(n-3)}\right)^{-1} \geq \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right) \geq \tilde{r} - z_\delta\left(\sqrt{(n-3)}\right)^{-1}\right) \\
&= \mathbb{P}\left(2\left(\tilde{r} + z_\delta\left(\sqrt{(n-3)}\right)^{-1}\right) \geq \ln\left(\frac{1+\rho}{1-\rho}\right) \geq 2\left(\tilde{r} - z_\delta\left(\sqrt{(n-3)}\right)^{-1}\right)\right) \\
&= \mathbb{P}\left(2\left(\tilde{r} - z_\delta\left(\sqrt{(n-3)}\right)^{-1}\right) \leq \ln\left(\frac{1+\rho}{1-\rho}\right) \leq 2\left(z_\delta\left(\sqrt{(n-3)}\right)^{-1} + \tilde{r}\right)\right) \\
&= \mathbb{P}\left(2\tilde{r}_u \leq \ln\left(\frac{1+\rho}{1-\rho}\right) \leq 2\tilde{r}_o\right) \\
&= \mathbb{P}\left(\exp\left(2\tilde{r}_u\right) \leq \frac{1+\rho}{1-\rho} \leq \exp\left(2\tilde{r}_o\right)\right) \\
&= \mathbb{P}\left(\exp\left(2\tilde{r}_u\right)(1-\rho) \leq 1+\rho \leq \exp\left(2\tilde{r}_o\right)(1-\rho)\right) \\
&= \mathbb{P}\left(\exp\left(2\tilde{r}_u\right) - \exp\left(2\tilde{r}_u\right)\rho \leq 1+\rho \leq \exp\left(2\tilde{r}_o\right) - \exp\left(2\tilde{r}_o\right)\rho\right) \\
&= \mathbb{P}\left(\exp\left(2\tilde{r}_u\right) - \exp\left(2\tilde{r}_u\right)\rho - 1 \leq \rho \leq \exp\left(2\tilde{r}_o\right) - \exp\left(2\tilde{r}_o\right)\rho - 1\right).
\end{aligned} \tag{2.95}$$

Weiterhin gilt aber

$$\begin{aligned}
&\exp\left(2\tilde{r}_u\right) - \exp\left(2\tilde{r}_u\right)\rho - 1 \leq \rho \\
&\quad\Leftrightarrow \exp\left(2\tilde{r}_u\right) - 1 \leq \exp\left(2\tilde{r}_u\right)\rho + \rho \\
&\quad\Leftrightarrow \exp\left(2\tilde{r}_u\right) - 1 \leq \left(\exp\left(2\tilde{r}_u\right) + 1\right)\rho \\
&\quad\Leftrightarrow \frac{\exp\left(2\tilde{r}_u\right) - 1}{\exp\left(2\tilde{r}_u\right) + 1} \leq \rho,
\end{aligned} \tag{2.96}$$

also

$$\delta = \mathbb{P}\left(\frac{\exp\left(2\tilde{r}_u\right)-1}{\exp\left(2\tilde{r}_u\right)+1} \leq \rho \leq \exp\left(2\tilde{r}_o\right) - \exp\left(2\tilde{r}_o\right)\rho - 1\right). \tag{2.97}$$

Analog gilt

$$\begin{aligned} \rho &\leq \exp\left(2\tilde{r}_o\right) - \exp\left(2\tilde{r}_o\right)\rho - 1 \\ \Leftrightarrow \exp\left(2\tilde{r}_o\right)\rho + \rho &\leq \exp\left(2\tilde{r}_o\right) \\ \Leftrightarrow \left(\exp\left(2\tilde{r}_o\right)+1\right)\rho &\leq \exp\left(2\tilde{r}_o\right) - 1 \\ \Leftrightarrow \rho &\leq \frac{\exp\left(2\tilde{r}_o\right)-1}{\exp\left(2\tilde{r}_o\right)+1} \end{aligned} \tag{2.98}$$

und damit schließlich

$$\begin{aligned} \delta &= \mathbb{P}\left(\frac{\exp\left(2\tilde{r}_u\right)-1}{\exp\left(2\tilde{r}_u\right)+1} \leq \rho \leq \frac{\exp\left(2\tilde{r}_o\right)-1}{\exp\left(2\tilde{r}_o\right)+1}\right) \\ &= \mathbb{P}\left(\left[\frac{\exp(2\tilde{r}_u)-1}{\exp(2\tilde{r}_u)+1}, \frac{\exp(2\tilde{r}_o)-1}{\exp(2\tilde{r}_o)+1}\right] \ni \rho\right) \\ &= \mathbb{P}_\rho\left(\kappa(r) \ni \rho\right). \end{aligned} \tag{2.99}$$

□

Beispiel

Wir demonstrieren Theorem 2.7 und Theorem 2.8 mithilfe eines Simulationsbeispiels. Dazu seien wie oben für $i = 1, ..., 30$

$$p(t_i) := N(t_i; 0, \sigma_\tau^2) \text{ und } p(y_{ij}|t_i) := N(y_{ij}; t_i, \sigma_\varepsilon^2) \tag{2.100}$$

für $j = 1, 2$ mit $\sigma_\tau^2 := 1.0$ und $\sigma_\varepsilon^2 := 0.2$. Folgender **R** Code generiert 10^5 Datensätze dieses Modells und evaluiert die entsprechenden Fisher-transformierten Stichprobenkorrelationen und Konfidenzintervalle für ein Konfidenzlevel von $\delta = 0.95$. Abbildung 2.4 A zeigt den Vergleich zwischen der so empirisch generierten Frequentistischen Verteilung der Fisher-transformierten Stichprobenkorrelationen und Abbildung 2.4 B zeigt die ersten 100 simulierten Konfidenzintervalle der Korrelation. In der vorliegenden Simulation zeigt sich im Vergleich zur ihrer analytischen Approximation empirisch eine leichte Verschiebung der Fisher-transformierten Stichprobenkorrelationen zu höheren Werten. Die betrachteten Konfidenzintervalle überdecken die wahre, aber unbekannte, Korrelation in zwei von 100 Fällen nicht.

```
set.seed(0)
nsim                = 1e5                                   # Reproduzierbarkeit
n                   = 30                                    # Realisierungsanzahl
m                   = 2                                     # Personenanzahl
sigsqr_tau          = 1                                     # Testmessungsanzahl
sigsqr_eps          = .2                                    # Wahrer Wert Varianz
R                   = sigsqr_tau/(sigsqr_tau+sigsqr_eps)    # Beobachteter Wert Varianz
delta               = 0.95                                  # Paralletestreliabilität
z_delta             = qnorm((1+delta)/2)                    # Konfidenzlevel
r                   = rep(NaN, nsim)                        # Konfidenzintervallskalierungsparameter
r_til               = rep(NaN, nsim)                        # Paralleltesterliabilitässchätzer Array
r_til_u             = rep(NaN, nsim)                        # Fisher Transformation von r
r_til_o             = rep(NaN, nsim)                        # Unterer Konfidenzintervallparameter
kappa               = matrix(rep(NaN,2*nsim), ncol= 2)      # Oberer  Konfidenzintervallparameter
for(s in 1:nsim){                                           # Konfidenzintervallarray
    T               = matrix(rep(NaN, n)  , nrow = n)       # Simulationsiterationen
    Y               = matrix(rep(NaN, n*m), nrow = n)       # Wahrer Wert Array
    E               = matrix(rep(NaN, n*m), nrow = n)       # Beobachteter Wert Array
    for(i in 1:n){                                          # Messfehler Array
        T[i]        = rnorm(1,1,sqrt(sigsqr_tau))           # Personeniterationen
        for(j in 1:m){                                      # Wahrer Wert Realisierung für j = 1,2
            Y[i,j]  = rnorm(1,T[i],sqrt(sigsqr_eps))        # Testmessungsiterationen
            E[i,j]  = Y[i,j] - T[i]}}                       # Beobachter Wert Realisierung
    r[s]            = cor(Y[,1],Y[,2])                      # Messfehler Realisierung
    r_til[s]        = 1/2*log((1+r[s])/(1-r[s]))            # Paralleltestreliabilitätsschätzer
    r_til_u[s]      = r_til[s] - z_delta*(1/sqrt(n-3))       # Fisher Transformation von r
                                                            # unterer Konfidenzintervallparameter
```

```
    r_til_o[s]      = r_til[s] + z_delta*(1/sqrt(n-3))              # oberer  Konfidenzintervallparameter
    kappa[s,1]      = (exp(2*r_til_u[s])-1)/(exp(2*r_til_u[s])+1)   # untere Konfidenzintervallgrenze
    kappa[s,2]      = (exp(2*r_til_o[s])-1)/(exp(2*r_til_o[s])+1)   # obere Konfidenzintervallgrenze
}
```

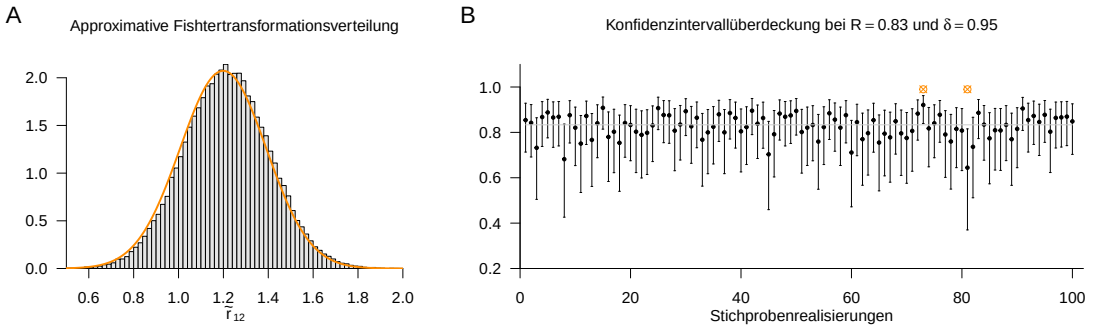

Abbildung 2.4 A Simulation der Verteilung der Fisher-transfomierten Stichprobenkorrelation und ihre analytische Approximation **B** Simulation der Überdeckungswahrscheinlichkeit des Konfidenzintervalls für die Stichprobenkorrelation bei einer wahren, aber unbekannten, Korrelation von $R = 0.833$ und einer gewünschten Überdeckungswahrscheinlichkeit von $\delta := 0.95$. Die Abbildung zeigt für jede Stichprobenrealisierung das Konfidenzintervall und die entsprechende Stichprobenkorrelation In der vorliegenden Simulation überdecken die Konfidenzintervalle die durch eine graue Linie eingezeichnete immer gleichen wahren, aber unbekannten, Korrelation $R := 0.833$ in 98 von 100 Fällen. Die Stichprobenrealisierungen, für die dies nicht der Fall sind, sind mit einen orangen Kreis markiert.

2.3 Interne Konsistenz

2.3.1 m-Komponententestmodelle

In den vorherigen Abschnitten bezeichnete y_{ij} die Zufallsvariable zur Modellierung des beobachteten Werts der $j$ten Testmessung der $i$ten Person, wobei wir das Vorliegen von m Testmessungen von n Personen angenommen haben. Wir haben dabei aber zunächst offen gelassen, ob mit der *jten Testmessung* das summative Ergebnis eines Tests oder das Ergebnis ein einzelnes Testitems gemeint ist, um die Theorie für beide Anwendungsfälle zu entwickeln. Als Grundlage der Bestimmung der internen Konsistenz eines Tests identifizieren wir in diesem Abschnitt nun die $j$te Testmessung mit dem $j$ten Item eines Tests. y_{ij} bezeichnet also im Folgenden die Zufallsvariable zur Modellierung des Beobachteten Werts des $j$ten Items der $i$ten Person in einem Test, wobei wir weiterhin m Items und n Personen annehmen. Weiterhin gehen wir in diesem Zusammenhang davon aus, dass für jede Person ein *Gesamt-Beobachteter-Wert* durch Summation über die Items eines Test gebildet wird. Die Zufallsvariable zur Modellierung dieses Gesamt-Beobachteten-Werts bezeichnen wir mit

$$y_i := \sum_{j=1}^{m} y_{ij}. \tag{2.101}$$

Wir fassen diese Vorüberlegungen in folgender Definition des *m-Komponententestmodells* zusammen.

Definition 2.8 (m-Komponententestmodell)**.** Gegeben sei das Modell multipler Testmessungen für $i = 1, ..., n$ Personen und $j = 1, ..., m$ Testmessungen. Für $i = 1, ..., n$ seien

- $y_i := \sum_{j=1}^{m} y_{ij}$ die *Summe der Beobachteten Werte* und
- $\tau_i := \sum_{j=1}^{m} \tau_{ij}$ die *Summe der Wahren Werte*.

Dann heißt die gemeinsame Verteilung der y_i und τ_i für $i = 1, ...n$ mit der Faktorisierungseigenschaft

$$\mathbb{P}(\tau_1, ..., \tau_n, y_1, ..., y_n) := \prod_{i=1}^{n} \mathbb{P}(\tau_i, y_i) = \prod_{i=1}^{n} \mathbb{P}(y_i|\tau_i)\mathbb{P}(\tau_i) \tag{2.102}$$

das m-Komponententestmodell, wenn gilt dass

$$\mathbb{P}(\tau_1, y_1) = \cdots = \mathbb{P}(\tau_n, y_n). \tag{2.103}$$

•

Zum Verständnis dazu, wie Cronbach's α die interne Konsistenz eines Tests misst, bietet es sich zunächst an, für das m-Komponententestmodell die Reliabilität im Sinne der Formulierung von Aussage (1) in Theorem 2.5 zu definieren.

Definition 2.9 (Reliabilität von m-Komponententestmodellen)**.** Gegeben sei ein m-Komponententestmodell mit der Summe der Beobachteten Werte y_i und der Summe der Wahren Werte τ_i für $i = 1, ..., n$ Personen. Dann ist die Reliabilität des Modells definiert als

$$\mathrm{R} := \frac{\mathbb{V}(\tau_i)}{\mathbb{V}(y_i)} \text{ für ein beliebiges } 1 \leq i \leq n. \tag{2.104}$$

•

2.3.2 Cronbach's α

Vor dem Hintergrund des m-Komponentenmodells definieren wir *Cronbach's α* wie folgt.

Definition 2.10 (Cronbach's α)**.** Gegeben sei ein m-Komponententestmodell. Dann heißt

$$\alpha := \frac{m}{m-1}\left(1 - \frac{\sum_{j=1}^{m} \mathbb{V}(y_{ij})}{\mathbb{V}(y_i)}\right) \tag{2.105}$$

Cronbach's α oder *Koeffizient α*.

•

Man beachte, dass in Definition 2.10 $\mathbb{V}(y_{ij})$ die Varianzen der Beobachteten Werte für Personen und Items und $\mathbb{V}(y_i)$ die Varianzen der Summen der Beobachteten Werte für Personen bezeichnen. Da nach Annahme des m-Komponententestmodells die Verteilungen der y_i identisch sind, wird Cronbach's α auch häufig in der Form

$$\alpha = \frac{m}{m-1}\left(1 - \frac{\sum_{j=1}^{m} \mathbb{V}(y_j)}{\mathbb{V}(y)}\right) \tag{2.106}$$

geschrieben, wobei $\mathbb{V}(y_j)$ die Varianzen der Items und $\mathbb{V}(y)$ die Varianz der Summen der Beobachten Werte bezeichnen.

Intuitiv nimmt Cronbach's α einen Wert nahe 1 an, wenn die Anzahl der Items m hoch und damit $\frac{m}{m-1} \approx 1$ ist und gleichzeitig das Verhältnis der summierten Itemvarianzen $\sum_{j=1}^{m} \mathbb{V}(y_j)$ zur Varianz der Summe der Itemwerte $\mathbb{V}(y)$ sehr gering und damit $\sum_{j=1}^{m} \mathbb{V}(y_j)/\mathbb{V}(y) \approx 0$ ist. Seine zentrale Bedeutung in der Klassischen Testtheorie bekommt Cronbach's α durch seinen Zusammenhang mit der Reliabilität eines m-Komponententests, welchen wir in unterem Theorem darstellen.

Theorem 2.9 (Cronbach's α und Reliabilität)**.** *Gegeben sei ein m-Komponententestmodell mit Reliabilität R. Dann gilt für Cronbach's α, dass*

$$\alpha \leq R \tag{2.107}$$

und Gleichheit tritt insbesondere dann ein, wenn die Testmessungen des m-Komponententestmodell parallel sind.

∘

Beweis. Zur Vereinfachung der Notation verzichten wir auf die explizite Auszeichung der Personen $i = 1, \ldots, n$ und setzen

$$y_j := y_{ij}, \tau_j := \tau_{ij}, y := \sum_{j=1}^{m} y_j \text{ und } \tau := \sum_{j=1}^{m} \tau_j, \tag{2.108}$$

sowie

$$\mathrm{R} := \frac{\mathbb{V}(\tau)}{\mathbb{V}(y)} \text{ und } \alpha := \frac{m}{m-1}\left(1 - \frac{\sum_{j=1}^{m} \mathbb{V}(y_j)}{\mathbb{V}(y)}\right). \tag{2.109}$$

Wir gehen in vier Schritten vor.

(1) (*Summendarstellung*) Wir halten zunächst folgende Darstellung der Summe von m Zahlen fest: für Zahlen $x_1, \ldots, x_m$ gilt, dass

$$\sum_{j=1}^{m} x_j = \frac{1}{m-1} \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} (x_j + x_k). \tag{2.110}$$

Anstelle eines Beweises betrachten wir den Fall $m := 4$. Dann gilt

$$\begin{aligned}
\frac{1}{3}\sum_{j=1}^{3}\sum_{k=j+1}^{4}(x_j + x_k) &= \frac{1}{3}\left(\sum_{k=1+1}^{4}(x_1 + x_k) + \sum_{k=2+1}^{4}(x_2 + x_k) + \sum_{k=3+1}^{4}(x_3 + x_k)\right) \\
&= \frac{1}{3}\left(\sum_{k=2}^{4}(x_1 + x_k) + \sum_{k=3}^{4}(x_2 + x_k) + \sum_{k=4}^{4}(x_3 + x_k)\right) \\
&= \frac{1}{3}\left((x_1 + x_2) + (x_1 + x_3) + (x_1 + x_4) + (x_2 + x_3) + (x_2 + x_4) + (x_3 + x_4)\right) \\
&= \frac{1}{3}\left((x_1 + x_1 + x_1) + (x_2 + x_2 + x_2) + (x_3 + x_3 + x_3) + (x_4 + x_4 + x_4)\right) \\
&= \frac{1}{3}\left(3x_1 + 3x_2 + 3x_3 + 3x_4\right) \\
&= x_1 + x_2 + x_3 + x_4 \\
&= \sum_{j=1}^{4} x_j
\end{aligned} \tag{2.111}$$

(2) (*True-Score-Kovarianzungleichung*) Wir leiten nun im Modell multipler Testmessungen eine Ungleichung her. Dazu betrachten wir in diesem Modell die Varianz der Differenz zweier wahrer Werte τ_j und τ_k. Mit der Nicht-Negativität der Varianz (Theorem 4.10) und dem Theorem zur Varianz spezieller Linearkombinationen von Zufallsvariablen (Theorem 4.21) ergibt sich

$$\mathbb{V}(\tau_j - \tau_k) \geq 0 \Leftrightarrow \mathbb{V}(\tau_j) + \mathbb{V}(\tau_k) - 2\mathbb{C}(\tau_j, \tau_k) \geq 0 \Leftrightarrow \mathbb{V}(\tau_j) + \mathbb{V}(\tau_k) \geq 2\mathbb{C}(\tau_j, \tau_k) \tag{2.112}$$

Weiterhin ergibt sich für beliebige $1 \leq j, k \leq m$ bei parallelen Testmessungen, dass

$$\mathbb{V}(\tau_j - \tau_k) = \mathbb{V}(f_j(\tau_1) - f_k(\tau_1)) = \mathbb{V}(\tau_1 - \tau_1) = \mathbb{V}(0) = 0. \tag{2.113}$$

Im Fall paralleler Testmessungen ergibt sich in obiger Ungleichung und ihrer Anwendung im Folgenden also Gleichheit.

(3) (*Summen-True-Score-Varianzungleichung*) Wir betrachten nun die Varianz der True-Score Summe im m-Komponententestmodell. Mit dem Theorem zur Varianz einer Linearkombination von Zufallsvariablen (Theorem 4.20), der Summendarstellung aus (1) und der True-Score-Kovarianzungleichung aus (2) ergibt sich zunächst

$$\begin{aligned}
\mathbb{V}(\tau) &= \mathbb{V}\left(\sum_{j=1}^{m} \tau_j\right) \\
&= \sum_{j=1}^{m} \mathbb{V}(\tau_j) + 2 \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} \mathbb{C}(\tau_j, \tau_k) \\
&= \frac{1}{m-1} \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} \left(\mathbb{V}(\tau_j) + \mathbb{V}(\tau_k)\right) + 2 \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} \mathbb{C}(\tau_j, \tau_k) \\
&\geq \frac{1}{m-1} \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} 2\mathbb{C}(\tau_j, \tau_k) + \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} 2\mathbb{C}(\tau_j, \tau_k) \\
&= \left(\frac{1}{m-1} + 1\right) \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} \mathbb{C}(\tau_j, \tau_k) \\
&= \left(\frac{1}{m-1} + \frac{m-1}{m-1}\right) \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} 2\mathbb{C}(\tau_j, \tau_k) \\
&= \frac{1+m-1}{m-1} \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} 2\mathbb{C}(\tau_j, \tau_k) \\
&= \frac{m}{m-1} \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} 2\mathbb{C}(\tau_j, \tau_k)
\end{aligned} \tag{2.114}$$

Mit Aussage (5) von Theorem 2.3 und wiederum mit Theorem 4.20 gilt dann

$$\begin{aligned}
\mathbb{V}(\tau) &\geq \frac{m}{m-1} \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} 2\mathbb{C}(\tau_j, \tau_k) \\
&= \frac{m}{m-1} \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} 2\mathbb{C}(y_j, y_k) \\
&= \frac{m}{m-1} \left(\mathbb{V}\left(\sum_{j=1}^{m} y_j\right) - \sum_{j=1}^{m} \mathbb{V}\left(y_j\right)\right) \\
&= \frac{m}{m-1} \left(\mathbb{V}(y) - \sum_{j=1}^{m} \mathbb{V}\left(y_j\right)\right).
\end{aligned} \tag{2.115}$$

(4) (*Reliabilität*) Wir betrachten schließlich die Reliabilität im m-Komponententestmodell. Es ergibt sich

$$\begin{aligned}
\mathrm{R} &= \frac{\mathbb{V}(\tau)}{\mathbb{V}(y)} \\
&\geq \frac{\frac{m}{m-1}\left(\mathbb{V}(y) - \sum_{j=1}^{m} \mathbb{V}\left(y_j\right)\right)}{\mathbb{V}(y)} \\
&= \frac{m}{m-1} \left(\frac{\mathbb{V}(y)}{\mathbb{V}(y)} - \frac{\sum_{j=1}^{m} \mathbb{V}\left(y_j\right)}{\mathbb{V}(y)}\right) \\
&= \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^{m} \mathbb{V}\left(y_j\right)}{\mathbb{V}(y)}\right) \\
&=: \alpha.
\end{aligned} \tag{2.116}$$

□

Cronbach's α ist also eine untere Grenze für die Reliabilität eines m-Komponententestmodells, die Relibabilität eines m-Komponententestmodells also ist mindestens so groß wie Cronbach's α, kann aber größer sein. Für parallele Testmessungen, also parallele Items, ist die Reliabilität eines m-Komponententestmodells sogar gleich α.

2.3.3 Schätzung von Cronbach's α

Die Schätzung von Cronbach's α greift auf die Stichprobenvarianzen der beobachteten Itemscores und der beobachteten Testsummenscores zurück. Für Daten von Personen $i = 1, ..., n$ ist ein Schätzer für die Varianz des $j$ten Itemscores durch

$$S_j^2 := \frac{1}{n-1} \sum_{i=1}^{n} \left(y_{ij} - \bar{y}_j\right)^2 \text{ mit } \bar{y}_j := \frac{1}{n} \sum_{i=1}^{n} y_{ij} \tag{2.117}$$

und ein Schätzer für die Varianz der Testsummenscores durch

$$S^2 := \frac{1}{n-1} \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2 \text{ mit } \bar{y} := \frac{1}{n} \sum_{i=1}^{n} y_i \tag{2.118}$$

gegeben. Ein Schätzer für α ergibt sich entsprechend zu

$$\hat{\alpha} = \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^{m} S_j^2}{S^2}\right). \tag{2.119}$$

Wir demonstrieren die Evaluation dieses Schätzers untenstehend anhand einer Simulation im Modell paralleler Testmessungen für $n := 30$ und $m := 21$ mit

$$\mathbb{P}(\tau_i) := N(1, 1) \text{ und } \mathbb{P}(y_{ij}|\tau_i) := N(\tau_i, 4) \text{ für } i = 1, ..., n, j = 1, ..., m. \tag{2.120}$$

```
library(psych)                                          # R Paket zur Testanalyse
set.seed(0)                                             # Reproduzierbarkeit
n   = 30                                                # Personenanzahl
m   = 21                                                # Itemanzahl
mu  = 1                                                 # Wahrer Wert Erwartungswertparameter
T   = matrix(rep(NaN, n)  , nrow = n)                   # Wahrer Wert Array
Y   = matrix(rep(NaN, n*m), nrow = n)                   # Beobachteter Wert Array
for(i in 1:n){                                          # Iteration über Personen
  T[i]  = rnorm(2,mu,sqrt(1))                           # Wahrer Wert Realisierung
  for(j in 1:m){                                        # Iteration über Items
    Y[i,j]  = rnorm(1,T[i],sqrt(4))}}                   # Beobachteter Wert Realisierung
vsi  = var(apply(Y,1,sum))                              # Stichprobenvarianz der Beobachten Wert Summen
siv  = sum(apply(Y,2,var))                              # Summe der Item-Stichprobenvarianzen
a    = (m/(m-1))*(1-(siv/vsi))                          # Direkte Berechnung von Cronbach's alphca
ap   = alpha(Y,warnings = F)                            # Berechnung von Cronbach's alpha mit psych
```

```
Cronbach's alpha (manuell) :  0.823
Cronbach's alpha (psych)   :  0.823
```

Die Frequentistische Verteilungs- und Inferenztheorie zu diesem Schätzer ist von Kristof (1963), Feldt (1965), Feldt et al. (1987), Van Zyl et al. (2000) und Yuan & Bentler (2002) ausgearbeitet worden.

3 Faktoranalyse

Mit dem Begriff *Faktoranalyse* werden verschiedene Inferenzverfahren bezeichnet, bei denen den zugrundeliegenden Modellen gemeinsam ist, dass sie mithilfe linear-affiner Transformationen nicht-beobachter Zufallsvektoren die Kovarianzstruktur eines beobachtbaren Zufallsvektors generieren. Dabei sind in der psychologischen Anwendung Realisierungen des beobachtbaren Zufallsvektors typischerweise die Itemwerte einer Gruppe von Proband:innen in einem psychologischen Test. Ziel einer Faktoranalyse ist es dann, die geschätzte Item $\times$ Item Kovarianzmatrix mithilfe eines Faktoranalysemodells von möglichst einfacher Struktur zu erklären. Stilprägend sind in dieser Hinsicht die Arbeit von Spearman (1904), in der versucht wird, die Kovarianzstruktur verschiedener Leistungstestwerte mithilfe eines *generellen Intelligenzfaktors* zu erklären und Arbeiten zur Analyse des 16PF-Fragebogens (Cattell (1957)) im Sinne der *Big Five* Faktoren der Persönlichkeit. Im Kontext der Fragebogenentwicklung hat es sich eingebürgert, reproduzierbare Ergebnisse von Faktoranalysen als Ausdruck der *faktoriellen Validität* eines Fragebogens zu interpretieren.

Anwendungsbeispiel

Wir betrachten einen simulierten Datensatz nach den Ergebnissen einer Studie zur faktoriellen Validität der deutschen Version des BDI-II (Hautzinger et al. (2006), Keller et al. (2008)). Keller et al. (2008) betrachten dabei eine Stichprobe von $n = 266$ depressiven Patient:innen, die zum Zeitpunkt der Bearbeitung des BDI-II die Kriterien einer depressiven Störung nach ICD-10/DSM-IV erfüllten. Untenstehende Tabelle zeigt die BDI-II-Item-Scores der ersten 12 Patiet:innen des simulierten Datensatzes, Abbildung 3.1 zeigt den gesamten Datensatz als kolorierte Datensatzmatrix. Abbildung 3.2 zeigt die Item $\times$ Item Kovarianz- und Korrelationsmatrizen des Datensatzes, die die primären Explananda einer faktoranalytischen Modellierung bilden.

Tabelle 3.1 BDI-II Item-Scores von 12 von 266 Patient:innen

	1	2	3	4	5	6	7	8	9	10	11	12
Traurigkeit	4	3	0	2	2	4	2	0	3	4	1	3
Pessimismus	2	2	1	2	3	3	3	0	3	2	2	1
Versagensgefühle	2	2	0	2	3	3	3	0	2	2	2	1
VerlustAnFreude	2	2	1	2	3	3	3	0	2	2	2	1
Schuldgefühle	3	3	0	2	2	3	1	1	3	2	0	4
Bestrafungsgefühle	3	3	0	2	3	3	1	1	2	3	0	4
Selbstablehnung	3	3	0	2	2	3	1	1	2	3	1	4
Selbstkritik	3	3	0	2	3	4	2	1	2	3	0	3
Suizidgedanken	3	3	1	2	2	3	1	2	2	3	0	4
Weinen	3	3	0	2	3	4	2	0	3	3	1	2
Unruhe	2	2	0	2	3	3	3	0	2	2	3	2
Interessenverlust	2	2	1	2	3	4	3	0	3	2	2	1
Entschlusslosigkeit	3	2	1	2	3	4	3	0	2	2	2	1
Wertlosigkeit	4	3	1	2	2	3	1	2	2	2	0	4
Energieverlust	2	2	1	2	3	3	3	0	3	2	2	1
Schlaf	2	2	1	2	3	3	3	0	2	3	2	1
Reizbarkeit	2	2	2	2	3	3	3	0	2	2	3	1
Appetitveränderung	2	2	1	2	3	3	3	0	2	2	2	1
Konzentrationsschwierigkeiten	2	2	1	2	3	3	3	0	2	2	2	1
Ermüdung	3	2	1	2	3	3	3	0	2	2	3	1
VerlustSexuellenInteresses	3	2	1	2	3	3	3	0	2	2	2	1

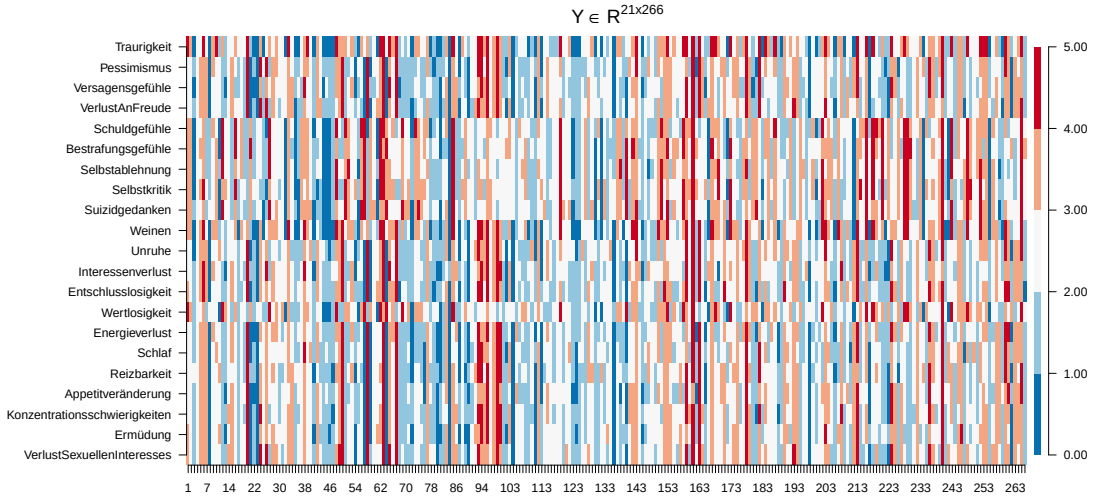

Abbildung 3.1 Simulierter BDI-II Datensatz nach Keller et al. (2008).

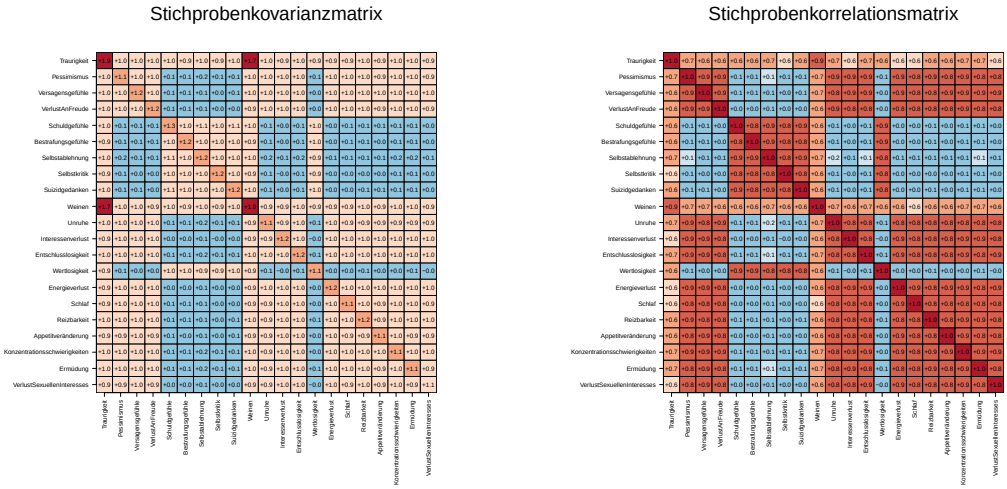

Abbildung 3.2 Stichprobenkovarianz- und Stichprobenkorrelationsmatrix des BDI-II Datensatz nach Keller et al. (2008).

3.1 Faktoranalysemodelle

Wir beginnen mit folgender Definition.

Definition 3.1 (Faktoranalysemodelle in struktureller Form)**.** Es sei

$$y = Lf + \varepsilon \tag{3.1}$$

wobei für $m > k$

- y ein m-dimensionaler beobachtbarer Zufallsvektor ist, der *Daten* genannt wird, ist,
- $L = (l_{ij}) \in \mathbb{R}^{m \times k}$ eine Matrix ist, die *Faktorladungsmatrix* genannt wird,
- f ein k-dimensionaler nicht-beobachtbarer Zufallsvektor ist, dessen Komponenten *Faktoren* genannt werden und für den gilt, dass

$$\mathbb{E}(f) = 0_k \text{ und } \mathbb{C}(f) = I_k, \tag{3.2}$$

- ε ein m-dimensionaler nicht-beobachtbarer und von f unabhängiger Zufallsvektor ist, der *Beobachtungsfehler* genannt wird und für den gilt, dass

$$\mathbb{E}(\varepsilon) = 0_m \text{ und } \mathbb{C}(\varepsilon) = \operatorname{diag}(\psi_1, ..., \psi_m) =: \Psi \text{ mit } \psi_i > 0 \text{ für } i = 1, ..., m. \tag{3.3}$$

Dann heißt (3.1) *Faktoranalysemodell in struktureller Form*. Gelten darüberhinaus, dass

- $f \sim N(0_k, \Phi)$ mit $\Phi \in \mathbb{R}^{k \times k}$ p.d. und
- $\varepsilon \sim N(0_m, \Psi)$ mit $\Psi := \operatorname{diag}(\psi_1, ..., \psi_m)$ für $\psi_i > 0$ und $i = 1, ..., m$,

dann heißt (3.1) *Normalverteilungsmodell der Faktoranalyse in struktureller Form*.

•

Die von den Komponenten des Datenvektors y modellierten Daten beziehen sich in der Analyse von psychologischen Testdaten in der Regel auf die Itemscores eines Tests. Die von den Komponenten des Vektors f modellierten *Faktoren* werden manchmal auch *gemeinsame Faktoren (common factors)* genannt und gegenüber den Komponenten des Zufallsvektors ε abgegrenzt, die als *spezifische Faktoren (unique factors)* bezeichnet werden. Gemäß des Matrixproduktes Lf in (3.1) wird der Eintrag l_{ij} für $i = 1, ..., m, j = 1, ..., k$ der Faktorladungsmatrix L für $i = 1, ..., m$ und $j = 1, ..., k$ die *Faktorladung des jten Faktors auf die ite Datenkompente* genannt. Wir schreiben das Faktoranalysemodell in der Folge meist in der Kurzform

$$y = Lf + \varepsilon \text{ mit } f \sim (0_k, I_k) \text{ und } \varepsilon \sim (0_m, \Psi) \tag{3.4}$$

Dabei soll die Notation $\zeta \sim (\mu, \Sigma)$ ausdrücken, dass $\mathbb{E}(\zeta) = \mu$ und $\mathbb{C}(\zeta) = \Sigma$ sind und f und ε sollen unabhängige Zufallsvektoren sein.

Verteilungsform des Faktoranalysemodells

In Verteilungsform kann das Faktoranalysemodell äquivalent geschrieben werden als

$$\mathbb{P}(f, y) = \mathbb{P}(f)\mathbb{P}(y|f) \text{ mit } f \sim (0_k, I_k) \text{ und } y \,|\, f \sim (Lf, \Psi). \tag{3.5}$$

Dabei erschließt sich die Bedingtheit der Verteilung von y auf f aus datengenerativer Sicht wie folgt: Es wird zunächst ein Wert von f anhand von $\mathbb{P}(f)$ realisiert, dieser Wert

wird dann durch L transformiert und bildet den Erwartungswert der bedingten Verteilung $\mathbb{P}(y|f)$ von y gegeben f. Dann wird ein Wert von ε realisiert und schließlich werden die Werte von Lf und ε werden zu einem Wert von y addiert. Entsprechend gilt für das Normalverteilungsmodell der Faktoranalyse die Wahrscheinlichkeitsdichte

$$p(f, y) = p(f)p(y|f) \text{ mit } p(f) = N(0_k, \Phi) \text{ und } p(y|f) = N(Lf, \Psi). \tag{3.6}$$

Datensatzmodell der Faktoranalyse

Im datenanalytischen Kontext betrachtet man die vorliegenden Itemscores einer Proband:in j, die einen psychologischen Fragebogen oder Test bearbeitet hat, als Realisierung eines Proband:innen-spezifischen und anhand der Verteilung des Faktoranalysemodells verteilten Zufallsvektors y_j. Gleichzeitig unterstellt man, dass diesem Zufallsvektor eine Realisierung des nicht-beobachtbaren Zufallsvektors f_j, also der Proband:inne-nspezifischen Faktorwerte, entspricht. Weiterhin nimmt man an, dass die Realisierungen von beobachtbaren Zufallsvektoren $y_1, ..., y_n$ und ihren entsprechenden nicht-beobachtbaren Zufallsvektoren $f_1, ..., f_n$ über Proband:innen unabhängig und identisch nach einem zugrundeliegenden proband:innenübergreifenden Faktoranalysemodell verteilt sind. Formal drückt man dies entweder als Faktorisierung der gemeinsamen Verteilung der beobachtbaren und nicht-beobachtbaren Zufallsvektoren in der Form

$$\mathbb{P}(f_1, y_1, ..., f_n, y_n) = \prod_{j=1}^{n} \mathbb{P}(f_j, y_j) \text{ mit } \mathbb{P}(f_j, y_j) = \mathbb{P}(f_k, y_k) \text{ für } 1 \leq j, k \leq n \tag{3.7}$$

oder im Sinne unabhängig und identisch verteilter Zufallsvektoren in der Form

$$(f_1, y_1), ..., (f_n, y_n) \sim \mathbb{P}(f, y) \tag{3.8}$$

aus. Entscheidend ist, dass man bei entsprechendem Vorliegen einer Datenmatrix $Y \in \mathbb{R}^{m \times n}$ wie in Abbildung 3.1 gezeigt, unterstellt, dass zu jedem Spalteneintrag ein virtueller, nicht-beobachteter, proband:innenspezifischer, Faktorenvektorwert korrespondiert. Inferenz bezüglich proband:innenspezifischer Faktorwerte, also die Modellschätzer-basierte Angabe der wahrscheinlichsten Werte des proband:innenspezifischen Faktorvektors, wird im Kontext der Faktoranalyse als *Evaluation von Factorscores* bezeichnet, steht in der Anwendung allerdings meist nicht im Vordergrund.

Beispiel (1) Spearmans g-Faktor Modell

Grundlage von Spearman's Intelligenzmodell ist die Korrelationsstruktur von Leistungswerten in $m = 6$ Themenbereichen (Classics, French, English, Mathematics, Pitch discrimination, Musical talent) in einer Stichprobe von ungefähr 30 Kindern im Alter von 9 und 13 Jahren (Yanai & Ichikawa (2007)). Spearman (1904) erklärt die entsprechenden Leistungswerte y_i mithilfe eines Faktoranalysemodells der Form

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} = \begin{pmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \\ l_5 \\ l_6 \end{pmatrix} f + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix} \Leftrightarrow y_i = l_i f + \varepsilon_i \text{ für } i = 1, ..., 6. \tag{3.9}$$

Dabei bezeichnet die latente Zufallsvariable f den *Allgemeinen Faktor der Intelligenz* und wird traditionell als *g-Faktor* bezeichnet. Für die Faktorladungen $l_i, i = 1, ..., 6$ fordert

Spearman (1904) $|l_i| < 1$. Die Komponenten ε_i werden im Kontext von Spearman (1904) als Themenbereichs-spezifische Faktoren bezeichnet. Entsprechend wird auch oft von *Spearman's Zweifaktoren Modell* gesprochen, wobei es im Sinne der modernen Faktoranalyse allerdings nur einen Faktor und einen Vektor von Zufallsfehlern in diesem Modell gibt.

Beispiel (2) Thurstones' Multifaktorenmodell

Thurstone (1947) schlägt mit dem Multifaktorenmodell ein generelles Modell von k Faktoren vor, um die Kovarianzstruktur eines m-dimensionalen Zufallskvetors, der m Leistungstestwerte einer Gruppe von Proband:innen modelliert, zu erklären. Dabei soll $k < m$ gelten. In struktureller Form hat dieses Modell die Form

$$y_i = l_{i1}f_1 + l_{i2}f_2 + \cdots + l_{ik}f_k + \varepsilon_i \text{ für } i = 1, ..., m \tag{3.10}$$

und entspricht damit der allgemeinen Form eines Faktoranalysemodells, wobei hier die Faktoren $f_1, ..., f_k$ als *gemeinsame Faktoren (common factors)* und die Zufallsfehler $\varepsilon_1, ..., \varepsilon_m$ als *spezifische Faktoren (unique factors)* bezeichnet werden. In Thurstone (1936) beispielsweise wird versucht zu zeigen, dass zur Erklärung der beobachteten Kovarianzstruktur von Leistungstestdaten von 240 Proband:innen sieben Faktoren (Memory, Word fluency, Verbal relation, Number, Perceptual Speed, Visualization, Reasoning), benötigt werden.

Beispiel (3) Das Modell paralleler Testmessungen

$y_1, ..., y_m$ seien die beobachtbaren Itemwerte eines psychologischen Tests und sei das Modell paralleler Testmessungen für den wahren Wert τ in der Form

$$y_i = \tau + \varepsilon_i \text{ für } i = 1, ..., m \tag{3.11}$$

mit Messfehler ε_i. Dann entspricht dieses Modell dem Faktoranalysemodell

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \tau + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix} \tag{3.12}$$

Insbesondere entspricht der Faktor f hier also dem wahren Wert τ und die Faktorladungsmatrix hat die als bekannt vorausgesetzte Form $L \coloneqq 1_m$.

Die zentrale Eigenschaft des Faktoranalysemodell ist die Beschaffenheit seiner marginalen Datenkovarianzmatrix, die wir im folgenden Theorem festhalten.

Theorem 3.1 (Marginale Datenkovarianzmatrix des Faktoranalysemodells)**.** *Gegeben sei ein Faktoranalysemodell*

$$y = Lf + \varepsilon \text{ mit } f \sim (0_k, I_k) \text{ und } \varepsilon \sim (0_m, \Psi). \tag{3.13}$$

Dann gilt für die marginale Kovarianzmatrix des Datenvektors

$$\mathbb{C}(y) = LL^T + \Psi. \tag{3.14}$$

∘

Beweis. Mit Theorem 4.23 gilt aufgrund der Unabhängigkeit von f und ε

$$\mathbb{C}(y) = L\mathbb{C}(f)L^T + \mathbb{C}(\varepsilon) = LI_kL^T + \Psi = LL^T + \Psi. \tag{3.15}$$

□

Die Diagonaleinträge der marginalen Datenkovarianzmatrix des Faktoranalysemodell lassen sich additiv durch die Einträge der Faktorladungsmatrix und der Kovarianzmatrix des Beobachtungsfehlers darstellen. Dies ist der Inhalt folgenden Theorems.

Theorem 3.2 (Varianzzerlegung der faktoranalytischen Datenkomponenten)**.** *Gegeben sei ein Faktoranalysemodell*

$$y = Lf + \varepsilon \text{ mit } f \sim (0_k, I_k) \text{ und } \varepsilon \sim (0_k, \Psi). \tag{3.16}$$

Dann ist für $i = 1, ..., m$ die Varianz der $i$ten Komponente von y gegeben durch

$$\mathbb{V}(y_i) = \sum_{j=1}^{k} l_{ij}^2 + \psi_i. \tag{3.17}$$

∘

Beweis. Mit Theorem 3.1 gilt

$$\begin{aligned}
\mathbb{C}(y) &= LL^T + \Psi \\
&= \begin{pmatrix} l_{11} & \cdots & l_{1k} \\ l_{21} & \cdots & l_{2k} \\ \vdots & \ddots & \vdots \\ l_{m1} & \cdots & l_{mk} \end{pmatrix} \begin{pmatrix} l_{11} & \cdots & l_{m1} \\ l_{12} & \cdots & l_{m2} \\ \vdots & \ddots & \vdots \\ l_{1k} & \cdots & l_{mk} \end{pmatrix} + \begin{pmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \psi_m \end{pmatrix} \\
&= \begin{pmatrix} \sum_{j=1}^k l_{1j}l_{1j} & \sum_{j=1}^k l_{1j}l_{2j} & \cdots & \sum_{j=1}^k l_{1j}l_{mj} \\ \sum_{j=1}^k l_{2j}l_{1j} & \sum_{j=1}^k l_{2j}l_{2j} & \cdots & \sum_{j=1}^k l_{2j}l_{mj} \\ \vdots & \cdots & \ddots & \vdots \\ \sum_{j=1}^k l_{mj}l_{1j} & \sum_{j=1}^k l_{mj}l_{2j} & \cdots & \sum_{j=1}^k l_{mj}l_{mj} \end{pmatrix} + \begin{pmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \psi_m \end{pmatrix} \\
&= \begin{pmatrix} \sum_{j=1}^k l_{1j}^2 + \psi_1 & \sum_{j=1}^k l_{1j}l_{2j} & \cdots & \sum_{j=1}^k l_{1j}l_{mj} \\ \sum_{j=1}^k l_{2j}l_{1j} & \sum_{j=1}^k l_{2j}^2 + \psi_2 & \cdots & \sum_{j=1}^k l_{2j}l_{mj} \\ \vdots & \cdots & \ddots & \vdots \\ \sum_{j=1}^k l_{mj}l_{1j} & \sum_{j=1}^k l_{mj}l_{2j} & \cdots & \sum_{j=1}^k l_{mj}^2 + \psi_m \end{pmatrix}.
\end{aligned} \tag{3.18}$$

Für den $i$ten Diagonaleintrag von $\mathbb{C}(y)$ aber gilt dann bekanntlich $\mathbb{C}(y_i, y_i) = \mathbb{V}(y_i)$. □

Die Terme in der Varianzdarstellung der faktoranalytischen Datenkomponenten von Theorem 3.2 erhalten im Kontext der Faktoranalyse spezielle Bezeichnungen.

Definition 3.2 (Kommunalität und Spezifität)**.** Gegeben sei ein Faktoranalysemodell

$$y = Lf + \varepsilon \text{ mit } f \sim (0_k, I_k) \text{ und } \varepsilon \sim (0_k, \Psi). \tag{3.19}$$

Dann werden in

$$\mathbb{V}(y_i) = \sum_{j=1}^{k} l_{ij}^2 + \psi_i \tag{3.20}$$

- $h_i^2 := \sum_{j=1}^k l_{ij}^2$ die *Kommunalität von y_i*

- ψ_i die *Spezifität von* y_i

genannt.

•

Die *Kommunalität* einer Datenkomponente y_i ist also der Varianzanteil von y_i, der durch die Faktorladungen erklärt wird. Die Spezifität einer Datenkomponente y_i dagegen ist der Varianzanteil von y_i, der nicht durch die Faktorladungen erklärt wird und damit spezifisch für diese Datenkomponente ist. Mnemonisch gilt für jede Datenkomponente eines Faktoranalysemodells also

$$\text{Varianz} \;=\; \text{Kommunalität} \;+\; \text{Spezifität}. \tag{3.21}$$

Auch die Summe der Varianzen der Datenkomponenten eines Faktoranalysemodells erhält eine eigene Bezeichnung.

Definition 3.3 (Gesamtvarianz)**.** Gegeben sei ein Faktoranalysemodell

$$y = Lf + \varepsilon \text{ mit } f \sim (0_k, I_k) \text{ und } \varepsilon \sim (0_k, \Psi). \tag{3.22}$$

Dann wird

$$\mathbb{G} := \sum_{i=1}^{m} \mathbb{V}(y_i) = \sum_{i=1}^{m} \sum_{j=1}^{k} l_{ij}^2 + \sum_{i=1}^{m} \psi_i \tag{3.23}$$

die *Gesamtvarianz von* y genannt.

•

Die Gesamtvarianz des beobachtbaren Datenvektors ist also definiert als die Summe der Varianzen der Datenkomponenten und entspricht damit der Spur der marginalen Datenkovarianzmatrix. Für die Gesamtvarianz gilt mnemonisch also

$$\text{Gesamtvarianz} \;=\; \text{Summe der Kommunalitäten} \;+\; \text{Summe der Spezifitäten}. \tag{3.24}$$

Wie wir später sehen werden kann die entsprechende Zerlegung der Gesamtstichprobenvarianz als Grundlage zur Evaluation der Modellgüte einer Faktoranalyse dienen.

Nichtidentifizierbarkeit

Eine fundamentale Eigenschaft des Faktoranalysemodells ist seine *Nichtidentifizierbarkeit*. Damit wird ausgedrückt, dass verschiedene Kombinationen von Faktorwerten und Faktorladungsmatrizen in der gleichen marginale Datenkovarianzmatrix resultieren und somit anhand einer Datenkovarianzmatrix bzw. ihrem Stichprobenäquivalent nicht eindeutig identizifiert werden können. Um diese Eigenschaft formal zu beschreiben, definieren wir zunächst den Begriff der orthogonalen Transformation eines Faktoranalysemodells

Definition 3.4 (Orthogonale Transformation eines Faktoranalysemodells)**.** Gegeben sei ein Faktoranalysemodell

$$y = Lf + \varepsilon \text{ mit } f \sim (0_k, I_k) \text{ mit } \varepsilon \sim (0_m, \Psi) \tag{3.25}$$

und es sei $Q \in \mathbb{R}^{k \times k}$ eine orthogonale Matrix. Dann nennen wir

$$\tilde{y} = \tilde{L}\tilde{f} + \varepsilon \text{ mit } \tilde{L} := LQ \text{ und } \tilde{f} := Q^T f \tag{3.26}$$

eine *orthogonale Transformation des Faktoranalysemodells.*

•

Die orthogonale Transformation eines Faktoranalysemodells lässt den Datenvektor und seine Kovarianzmatrix unberührt. Dies ist die Aussage folgenden Theorems.

Theorem 3.3 (Nichtidentifizierbarkeit und Kovarianzinvarianz der Faktoranalyse)**.** *Gegeben sei ein Faktoranalysemodell*

$$y = Lf + \varepsilon \text{ mit } f \sim (0_k, I_k) \text{ mit } \varepsilon \sim (0_m, \Psi) \tag{3.27}$$

sowie eine seiner orthogonale Transformationen

$$\tilde{y} = \tilde{L}\tilde{f} + \varepsilon \text{ mit } \tilde{L} := LQ \text{ und } \tilde{f} := Q^T f \tag{3.28}$$

für eine orthogonale Matrix $Q \in \mathbb{R}^{k \times k}$. Dann gelten

$$y = \tilde{y} \text{ und } \mathbb{C}(\tilde{y}) = \mathbb{C}(y) \tag{3.29}$$

∘

Beweis. Es gilt zum einen

$$\tilde{y} = \tilde{L}\tilde{x} + \varepsilon = LQQ^T f + \varepsilon = LI_k f + \varepsilon = Lf + \varepsilon = y. \tag{3.30}$$

Es gilt zum anderen

$$\mathbb{C}(\tilde{y}) = LQ(LQ)^T + \Psi = LQQ^T L^T + \Psi = LI_k L^T + \Psi = LL^T + \Psi = \mathbb{C}(y). \tag{3.31}$$

□

Mit

$$y = Lf + \varepsilon = \tilde{L}\tilde{f} + \varepsilon \tag{3.32}$$

folgt also unmittelbar, dass für einen festen Wert von y die Faktorladungsmatrix L und der Faktorvektorwert von f nicht eindeutig bestimmt sind. Verschiedene Faktorladungsmatrizen und Faktorwerte können also die gleichen Daten erklären und aus einer gegebenen Stichprobenkovarianzmatrix kann nicht eindeutig auf L und f geschlossen werden. Weiterhin folgt mit

$$\mathbb{C}(y) = LL^T + \Psi = \tilde{L}\tilde{L}^T + \Psi \tag{3.33}$$

aber auch, dass sich die Gesamtvarianz und die Kommunalitäten bei orthogonaler Transformation nicht ändern. Fundamental ist die Nichtidentifizierbarkeit des Faktoranalysemodells dadurch bedingt, dass nach Annahme weder die Faktorladungsmatrix, noch die Werte der Faktoren, noch die Beobachtungsfehler bekannt sind und Daten daher durch das Zusammenbeispiel dreier unbekannter Entitäten generiert werden.

Identifizierbarkeitsbedingungen

Zumindest im Kontext des Normalverteilungsmodells der Faktoranalyse kann man versuchen, die Identifizierbarkeit der Parameter dieses Modells durch eine Reihe von Nebenbedingungen zu gewährleisten. Allerdings sind auch in diesem Modell allgemeine hinreichende und notwendige Bedingungen für die Modellidentifizierbarkeit nicht vollumfänglich bekannt, die Modellidentifizierbarkeit im Bereich der Faktoranalyse und dem eng verwandten Feld der Strukturgleichungsmodelle bleibt also in aktives Forschungsfeld (vgl. Hyvärinen et al. (2024)). In der Anwendung sind deshalb simulationsbasierte Parameterrecoverystudien sicherlich eine gute Forschungspraxis. In diesem Abschnitt wollen wir die Identifizierbarkeit des Normalverteilungsmodells der Faktoranalyse genauer betrachten. Dazu definieren wir zunächst den Begriff der Identifizierbarkeit für das Normalverteilungsmodell der Faktoranalyse und führen dann mit der *Ordnungsbedingung* eine häufig verwendete Heuristik zur Modellidentifizierbarkeit ein, die die Intuition formalisiert, dass ein Modell im Sinne der datenanalytischen Datenreduktion nicht mehr Parameter haben sollte als Datenstatistiken betrachtet werden. Dazu zählen wir zunächst die Anzahl der unikalen Parameter und der Statistiken des Normalverteilungsmodells der Faktoranalyse.

Definition 3.5 (Identifizierbares Normalverteilungsmodell der Faktoranalyse)**.** Gegeben sei ein Normalverteilungsmodell der Faktoranalyse

$$y = Lf + \varepsilon \text{ mit } f \sim N(0_k, \Phi) \text{ und } \varepsilon \sim N(0_m, \Psi). \tag{3.34}$$

Weiterhin sei der *Parametervektor* dieses Modells definiert als die spaltenweise Konkatenation der Parametermatrizen L, Φ, Ψ, also

$$\theta := \text{vec}(L, \Phi, \Psi), \tag{3.35}$$

so dass die marginale Datenkovarianz geschrieben werden kann als

$$\Sigma_\theta := L\Phi L^T + \Psi. \tag{3.36}$$

Dann heißen das Normalverteilungsmodell der Faktoranalyse und der Parametervektor θ *identifizierbar*, wenn für alle $\theta_1, \theta_2 \in \Theta$ gilt, dass

$$\Sigma_{\theta_1} = \Sigma_{\theta_2} \Leftrightarrow \theta_1 = \theta_2. \tag{3.37}$$

Wenn für $\theta_1 \neq \theta_2$ gilt, dass $\Sigma_{\theta_1} = \Sigma_{\theta_2}$, so heißen das Modell und θ *nicht identifizierbar*.

•

Bei nicht identifizierbaren Modellen ergeben unterschiedliche Parameterwerte also die gleiche Datenverteilung und es kann folglich aus Daten nicht eindeutig auf Parameterwerte zurückgeschlossen werden. Allerdings gibt es bisher keine allgemein gültigen notwendigen und hinreichenden Bedingungen für die Identifizierbarkeit von Normalverteilungsmodell der Faktoranalyseen. Die unten vorgestellte *Ordnungsbedingung* ist lediglich eine notwendige Bedingung für die Identifizierbarkeit. Um sie diskutieren zu können, zählen wir zunächst die Anzahl an einzelnen (unikalen) skalaren Parametern und Stichprobenstatistiken eines Normalverteilungsmodell der Faktoranalyses. Dabei ist zentral, dass die Symmetrie von Kovarianzmatrizen bedeutet, dass nur die Einträge über und inklusive ihrer Hauptdiagonale unikal sind.

Theorem 3.4 (Anzahl unikaler skalarer Parameter und Statistiken)**.** *Gegeben sei ein Normalverteilungsmodell der Faktoranalyse*

$$y = Lf + \varepsilon \text{ mit } f \sim N(0_k, \Phi) \text{ und } \varepsilon \sim N(0_m, \Psi). \tag{3.38}$$

ein konfirmatorisches Faktoranalysemodell. Dann gilt für die Anzahl an unikalen skalaren Parametern des Modells

$$n_\theta = mk + \frac{k(k+1)}{2} + m \tag{3.39}$$

Weiterhin sei C sei die Stichprobenkovarianzmatrix eines Datensatzes von n unabhängigen Beobachtungen von y. Dann gilt für die Anzahl an unikalen skalaren Stichprobenstatistiken des Normalverteilungsmodell der Faktoranalyses

$$n_c = \frac{m(m+1)}{2}. \tag{3.40}$$

∘

Beweis. Die Anzahl der Einträge der Faktorladungsmatrix $L \in \mathbb{R}^{m\times k}$ ist mk.

Die Anzahl der Einträge einer symmetrischen Matrix $S \in \mathbb{R}^{k\times k}$ ist k^2. Aufgrund der Symmetrie von S sind dabei allerdings nur die k Einträge der Hauptdiagonale und die Einträge oberhalb (oder unterhalb) der Hauptdiagonalen unikal. Die Anzahl der Einträge oberhalb (oder unterhalb) der Hauptdiagonalen ist die Hälfte aller $k^2 - k$ nicht-diagonalen Einträge von S, also $(k^2 - k)/2$. Zusammen mit den Einträgen auf der Hauptdiagonalen ergibt sich damit für die Anzahl unikaler Einträge einer symmetrischen Matrix

$$\frac{k^2-k}{2} + k = \frac{k^2-k}{2} + \frac{2k}{2} = \frac{k^2+k}{2} = \frac{k(k+1)}{2}. \tag{3.41}$$

Als Kovarianzmatrix ist $\Phi \in \mathbb{R}^{k\times k}$ symmetrisch und hat damit $\frac{k(k+1)}{2}$ unikale Einträge. Die Anzahl der von Null verschiedenden Einträge von $\Psi \in \mathbb{R}^{m\times m}$ ist m.

Für die Anzahl an unikalen skalaren Parametern des Normalverteilungsmodell der Faktoranalyses e rgibt sich also zusammenfassend

$$n_\theta = mk + \frac{k(k+1)}{2} + m. \tag{3.42}$$

Die Stichprobenkovarianzmatrix $C \in \mathbb{R}^{m\times m}$ eines Datensatzes ist symmetrisch. Mit obigen Überlegungen zu den unikalen Einträgen einer symmetrischen Matrix ergibt sich also direkt

$$n_c = \frac{m(m+1)}{2}. \tag{3.43}$$

□

Nach Theorem 3.4 bezeichnet n_θ also die Dimension des unikalen Parametervektors θ, es gilt also $\theta \in \mathbb{R}^{n_\theta}$ und n_c ist die Anzahl der unikalen skalaren Einträge von C. Das Verhältnis von n_θ und n_c ist die Grundlage der Ordnungsbedingung zur Identifizierbarkeit von Normalverteilungsmodell der Faktoranalyseen.

Definition 3.6 (Ordnungsbedingung)**.** Gegeben sei ein Normalverteilungsmodell der Faktoranalyse

$$y = Lf + \varepsilon \text{ mit } f \sim N(0_k, \Phi) \text{ und } \varepsilon \sim N(0_m, \Psi) \tag{3.44}$$

Dann sagt man, dass das Modell der *Ordnungsbedingung* genügt, wenn für die Anzahl n_θ der unikalen skalaren Parameter und für die Anzahl n_c der unikalen skalaren Statistiken gilt, dass

$$n_\theta \leq n_c. \tag{3.45}$$

•

Die Ordnungsbedingung besagt also, dass die Anzahl der unbekannten und damit zu schätzenden Parameter des betrachteten Modells kleiner oder gleich der unikalen Einträge der Stichprobenkovarianzmatrix ist. Softwarelösungen wie Lavaan implementieren die Ordnungsbedingung meist per default (vgl. Rosseel (2021))).

3.2 Schätzverfahren

Zur Schätzung der Parameter eines Faktoranalysemodells stehen verschiedene Schätzverfahren zur Verfügung die in unterschiedlichen Softwareumgebungen unterschiedlich implementiert sein könnnen und so zu numerisch unterschiedlichen Schätzern für die Parameter des Faktoranalysemodells führen können. Gemeinsam ist den hier betrachteten Schätzverfahren, dass sie von einem Faktoranalysemodell fester Ordnung, d.h. mit einer prädefinierten Anzahl k von Faktoren ausgehen.

3.2.1 Hauptkomponentenschätzung

Mit der *Hauptkomponentenschätzung* wollen wir zunächst ein grundlegendes Verfahren zur Schätzung der Parameter eines Faktoranalysemodells diskutieren. Zentrale Motivation ist dabei die Approximation der Stichprobenkovarianzmatrix eines durch ein Faktoranalysemodell generierten Datensatzes anhand von Theorem 3.1 durch

$$C \approx \hat{L}\hat{L}^T + \hat{\Psi}. \tag{3.46}$$

Die Hauptkomponentenschätzung vernachlässigt dabei zunächst $\hat{\Psi}$ und nutzt die Orthonormalzerlegung

$$C = Q\Lambda Q^T = Q\Lambda^{1/2}\Lambda^{1/2}Q^T = \left(Q\Lambda^{1/2}\right)\left(Q\Lambda^{1/2}\right)^T \tag{3.47}$$

zur Darstellung von C. Die Hauptkomponentenschätzung vernachlässigt dann weiterhin die $k+1, ..., m$ Spalten von Q sowie die $k+1, ..., m$ Zeilen und Spalten von Λ und setzt

$$\hat{L}\hat{L}^T = Q_k\Lambda_k^{1/2}\left(Q_k\Lambda_k^{1/2}\right)^T \text{ mit } \hat{L} \in \mathbb{R}^{m\times k}. \tag{3.48}$$

Für die Diagonalelemente c_{ii}, $\hat{h}_i^2$ und $\hat{\psi}_i$ von $C, \hat{L}\hat{L}^T$ bzw. $\hat{\Psi}$ folgt dann, dass

$$c_{ii} = \sum_{j=1}^{k}\hat{l}_{ij}^2 + \hat{\psi}_i \Leftrightarrow \hat{\psi}_i = c_{ii} - \sum_{j=1}^{k}\hat{l}_{ij}^2, \tag{3.49}$$

worauf sich die entsprechenden Spezifitätsschätzer bei Hauptkomponentenschätzung gründen. Wir fassen das skizzierte Vorgehen in folgenden Definitionen zusammen.

Definition 3.7 (Hauptkomponentenschätzer $k$ter Ordnung von L und Ψ)**.** Gegeben sei ein Datensatz $Y \in \mathbb{R}^{m\times n}$ von n unabhängigen Beobachtungen eines Faktoranalysemodell

$$y = Lf + \varepsilon \text{ mit } f \sim (0_k, I_k) \text{ und } \varepsilon \sim (0_m, \Psi). \tag{3.50}$$

$C \in \mathbb{R}^{m\times m}$ sei die Stichprobenkovarianzmatrix von Y und

$$C = Q\Lambda Q^T \tag{3.51}$$

sei ihre Orthonormalzerlegung mit spaltenweise der Größe nach sortierten Eigenwerten und zugehörigen Eigenvektoren. Dann sind die *Hauptkomponentenschätzer kter Ordnung von L und Ψ* definiert als

$$\hat{L} := Q_k \Lambda_k^{1/2} \in \mathbb{R}^{m \times k} \text{ und } \hat{\Psi} := \text{diag}\left(\hat{\psi}_1, ..., \hat{\psi}_m\right), \tag{3.52}$$

wobei $Q_k \in \mathbb{R}^{m \times k}$ die Matrix bestehend aus den ersten k Spalten von $Q \in \mathbb{R}^{m \times m}$, $\Lambda_k \in \mathbb{R}^{k \times k}$ die Matrix bestehend aus den ersten k Zeilen und k Spalten von $\Lambda \in \mathbb{R}^{m \times m}$ und für $i = 1, ..., m$

$$\hat{\psi}_i := c_{ii} - \sum_{j=1}^{k} \hat{l}_{ij}^2 \tag{3.53}$$

mit den Diagonaleinträgen c_{ii} von C bezeichnen.

•

Die Selektion der ersten k Spalten von C und der ersten k Zeilen und Spalten von Λ in Definition 3.7 impliziert ein Faktoranalysemodell mit k Faktoren. Vor dem Hintergrund der Bezeichnungen von Definition 3.2 und Definition 3.3 ergeben sich auf Grundlage von Definition 3.7 folgende weitere Schätzer.

Definition 3.8 (Varianz-, Kommunalitäts-, Spezifitäts- und Gesamtvarianzschätzer)**.** Für einen Datensatz $Y \in \mathbb{R}^{m \times n}$ von n unabhängigen Beobachtungen und ein festes $k < m$ seien $\hat{L} = (\hat{l}_{ij})_{1 \leq i \leq m, 1 \leq j \leq k} \in \mathbb{R}^{m \times k}$ und $\hat{\Psi} = \text{diag}(\hat{\psi}_1, ..., \hat{\psi}_m) \in \mathbb{R}^{m \times m}$ die auf der Stichprobenkovarianzmatrix $C = (c_{ij})_{1 \leq i,j \leq m}$ basierenden Hauptkomponentenschätzer $k$ter Ordnung. Dann ergeben sich für $i = 1, ..., m$,

- c_{ii} als Schätzer von $\mathbb{V}(y_i)$ (*Varianzschätzer*),
- $\hat{h}_i^2 := \sum_{j=1}^{k} \hat{l}_{ij}^2$ als Schätzer von h_i^2 (*Kommunalitätsschätzer*),
- $\hat{\psi}_i$ als Schätzer von ψ_i (*Spezifitätsschätzer*),
- $G := \text{tr}(C)$ als Schätzer von $\mathbb{G}$ (*Gesamtvarianzschätzer*).

•

Anwendungsbeispiel

Folgender **R** Code demonstriert die Hauptkomponentenschätzung für ein Faktoranalysemodell mit $k = 2$ anhand des in Abbildung 3.1 dargestellen Datensatzes nach Keller et al. (2008). Abbildung 3.3 zeigt die resultierenden Faktorladungsmatrix- und Spezifitätsschätzer.

```
Y         = as.matrix(t(read.csv("./_data/fa-keller.csv")))        # Y \in \mathbb{R}^{m x n}
m         = nrow(Y)                                                # Datendimension
n         = ncol(Y)                                                # Datenpunktanzahl
k         = 2                                                      # Faktoranzahl
I_n       = diag(n)                                                # Einheitsmatrix I_n
J_n       = matrix(rep(1,n^2), nrow = n)                           # 1_{nn}
C         = (1/(n-1))*(Y %*% (I_n-(1/n)*J_n) %*% t(Y))             # Stichprobenkovarianzmatrix
EA        = eigen(C)                                               # Eigenanalyse von R
lambda_k  = EA$values[1:k]                                         # k größte Eigenwerte von R
Q_k       = EA$vectors[,1:k]                                       # k zugehörige Eigenvektoren von R
L_hat     = Q_k %*% diag(sqrt(lambda_k))                           # Faktorladungsmatrixschätzer
Psi_hat   = diag(diag(C) - diag(L_hat %*% t(L_hat)))               # Beobachtungsrauschenkovarianzmatrixschätzer
V_i_hat   = diag(C)                                                # Varianzschätzer
h2_i_hat  = rowSums(L_hat^2)                                       # Kommunalitätsschätzer
psi_i_hat = diag(Psi_hat)                                          # Spezifitätsschätzer
```

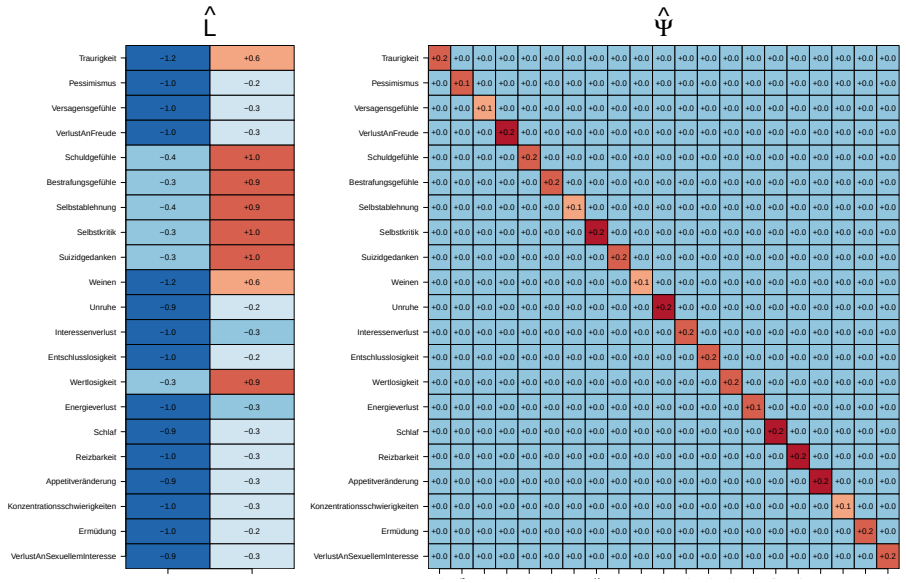

Abbildung 3.3 Faktorladungsmatrix- und Spezifitätsschätzer für den BDI-II Datensatz nach Keller et al. (2008).

3.2.2 Iterierte Hauptachsenfaktorisierung

Die Grundidee der iteriertern Hauptachsenfaktorisierung (IHAF) zur Schätzung der der Faktorladungsmatrix besteht darin, eine Schätzung $\hat{L}$ der Faktorladungsmatrix L auf Basis der Stichproben-Korrelationsmatrix $R \in \mathbb{R}^{m \times m}$ zu berechnen. Zu diesem Zweck versucht der Algorithmus der IHAF, iterativ verbesserte Kommunalitätsschätzungen zu erzeugen, indem die geschätzten Kommunalitäten sukzessive wieder in die Stichproben-Korrelationsmatrix eingesetzt werden. In folgender Definition halten wir die algorithmische Ausgestaltung der IHAF wie im R-Paket `psych` (Revelle (2024)) fest, die wir nachfolgend erläutern wollen.

Definition 3.9 (Iterierte Hauptachsenfaktorisierung nach Revelle (2024))**.** Gegeben sei eine Stichprobenkorrelationsmatrix $R \in \mathbb{R}^{m \times m}$. Dann hat der Algorithmus der iterierten Hauptachsenfaktorisierung nach Revelle (2024) die folgende Form.

Initialisierung

(S0) Definiere $\epsilon_{\min}$, $i_{\max}$, und setze $R^{(0)} := R$, $h^{(0)} := \sum_{j=1}^{m} R_{jj}^{(0)}$, $\epsilon^{(1)} := h^{(0)}$ und $i := 1$

Iterationen

Solange $\epsilon^{(i)} > \epsilon_{\min}$ und $i \leq i_{\max}$

(S1) Zerlege $R^{(i-1)} = Q^{(i)} \Lambda^{(i)} Q^{(i)^T}$ und setze $\hat{L}^{(i)} := Q_k^{(i)} \Lambda_k^{(i)^{1/2}}$

(S2) Setze $R^{(i)} := R^{(i-1)}$ und $R_{jj}^{(i)} := \left(\hat{L}^{(i)} \hat{L}^{(i)^T} \right)_{jj}$ für $j = 1, \ldots, m$

(S3) Setze $h^{(i)} := \sum_{j=1}^{m} R_{jj}^{(i)}$, $\epsilon^{(i+1)} := |h^{(i)} - h^{(i-1)}|$ und $i := i + 1$

Finalisierung

(S4) Setze $\hat{L} := \hat{L}^{(i)} D$ mit $D := \operatorname{diag}\left(\operatorname{sgn}\left(\sum_{j=1}^{m} \hat{l}_{j1}^{(i)} \right), \ldots, \operatorname{sgn}\left(\sum_{j=1}^{m} \hat{l}_{jk}^{(i)} \right) \right)$

-

Im Initialisierungsschritt (S0) ist $\epsilon_{\min}$ eine Schranke für das Konvergenzkriterium, wie weiter unten erläutert, $i_{\max}$ ist die maximale Anzahl an Iterationen, die durchgeführt werden sollen, $R^{(0)}$ ist die iterierte Korrelationsmatrix, die mit der experimentell beobachteten Korrelationsmatrix R initialisiert wird, $h^{(0)}$ ist der Anfangswert der Summe der Kommunalitätsschätzungen, also die Summe der Diagonalelemente von $R^{(0)}$, $\epsilon^{(1)}$ ist das Konvergenzkriterium und i ist der Iterationszähler. Die Iterationen werden solange durchgeführt, bis entweder das Konvergenzkriterium $\epsilon^{(i)}$ einen Wert annimmt, der kleiner oder gleich der Schranke $\epsilon_{\min}$ ist, oder mehr Iterationen als die maximale Iterationsanzahl $i_{\max}$ durchgeführt wurden.

Im ersten Iterationsschritt (S1) wird die Eigenwertzerlegung der iterierten Korrelationsmatrix durchgeführt, und die Schätzung der Faktorladungsmatrix k-ter Ordnung wird basierend auf den k Eigenvektoren mit den k größten Eigenwerten der iterierten Korrelationsmatrix berechnet. Genauer gesagt sind $Q_k^{(i)} \in \mathbb{R}^{m \times k}$ und $\Lambda_k^{(i)} \in \mathbb{R}^{m \times k}$ die Matrizen, die aus den ersten k Spalten von $Q^{(i)} \in \mathbb{R}^{m \times m}$ bzw. $\Lambda^{(i)} \in \mathbb{R}^{m \times m}$ bestehen, wobei angenommen wird, dass die Spalten entsprechend der absteigenden Reihenfolge der Eigenwerte sortiert sind. Entscheidend ist, dass im zweiten Iterationsschritt (S2) die Diagonaleinträge der iterierten Korrelationsmatrix durch die aus der aktuellen Schätzung der Faktorladungsmatrix resultierenden Kommunalitätsschätzungen ersetzt werden. Dadurch entsteht eine aktualisierte, iterierte Korrelationsmatrix, die die Grundlage für die nächste Schätzung der Faktorladungsmatrix bildet. Im Iterationsschritt (S3) wird die Summe der Kommunalitätsschätzungen neu berechnet, das Konvergenzkriterium als absoluter Unterschied zwischen den Summen der aktuellen und vorherigen Iteration bestimmt, und der Iterationszähler wird erhöht. Der Algorithmus konvergiert also, wenn sich die Kommunalitätsschätzungen zwischen zwei aufeinanderfolgenden Iterationen nicht wesentlich ändern oder die maximale Anzahl erlaubter Iterationen erreicht wurde.

Abschließend werden angesichts der fundamentalen Unbestimmtheit der Parameter des Faktorenanalysemodell im Schritt (S4) die Vorzeichen der Einträge der geschätzten Faktorladungsmatrix so gewählt, dass die Spaltensummen der Faktorladungsmatrix größer oder gleich null sind. Zu diesem Zweck wird die bei Konvergenz erhaltene Faktorladungsmatrix $\hat{L}^{(i)}$ von rechts mit einer Diagonalmatrix multipliziert, deren Diagonalelemente den Werten der Signumfunktion sgn der Spaltensummen von $\hat{L}^{(i)}$ entsprechen, wobei $\text{sgn}(x) = +1$ für $x \geq 0$ und $\text{sgn}(x) = -1$ für $x < 0$ gilt.

Anwendungsbeispiel

Folgender **R** Code demonstriert die iterierte Hauptachsenfaktorisierung zur Schätzung der Parameter eines Faktoranalysemodells mit Faktoranalysemodell mit $k = 2$ anhand des in Abbildung 3.1 dargestellen Datensatzes nach Keller et al. (2008).

```
Y            = as.matrix(read.csv("./_data/fa-keller.csv"))          # Y \in \mathbb{R}^{m x n}
R            = cor(Y)                                                # m x m Korrelationsmatrix
k            = 2                                                     # Faktorenzahl
eps_min      = 0.01                                                  # Konvergenzkriterium
i_max        = 25                                                    # Maximale Anzahl an IPAF Iterationen
i            = 0                                                     # Iterationszähler
R_i          = R                                                     # Initiale Korrelationsmatrix R^{(0)}
h_i          = sum(diag(R_i))                                        # Summe der Kommunalitätsschätzer
eps_i        = h_i                                                   # Fehlerinitialisierung
while(eps_i > eps_min && i <= i_max){                                # Iterationen
    QLQ_i      = eigen(R_i)                                          # Eigenanalyse
    L_hat_i    = QLQ_i$vectors[,1:k] %*% diag(sqrt(QLQ_i$values[1:k])) # Hauptkomponentenschätzer kter Ordnung
    diag(R_i)  = diag(L_hat_i %*% t(L_hat_i) )                       # Kommunalitätsschätzer Update
    h_ii       = sum(diag(R_i))                                      # Kommunalitätschätzersummen Update
    eps_i      = abs(h_ii - h_i)                                     # Fehlerkonvergenkriterium
    h_i        = h_ii                                                # Kommunalitätsschätzersummen Update
```

```
    i           = i + 1}                                    # Iterationszähler Update
D               = diag(sign(colSums(L_hat_i)))              # Vorzeichenfinalisierung
L_hat           = L_hat_i %*% D                             # IHAF Faktorladungsmatrixschätzer
```

3.2.3 Maximum-Likelihood-Schätzung

Im Normalverteilungsmodell der Faktoranalyse können die Parameter mithilfe des Maximum-Likelihood-Verfahrens geschätzt werden. Traditionell werden die Parameter dabei durch Minimierung der sogenannten *Diskrepanzfunktion* geschätzt (vgl. Lawley (1940), Jöreskog (1967) und Jöreskog (1969)). Die funktionale Form der Diskrepanzfunktion ist dabei durch ein Log-Likelihood-Kriterium bei Betrachtung der Frequentistischen Verteilung der Stichprobenkovarianz motiviert. Diese ist nach Wishart (1928) benannt. Modernere Sichtweisen im Kontext von Strukturgleichungsmodellen motivieren die funktionale Form der Diskrepanzfunktion direkt durch ein Log-Likelihood-Kriterium bei Betrachtung der multivariaten Datennormalverteilung (vgl. Bollen (1989) und Rosseel (2021)). Wir wollen hier diesen modernen Weg nachzeichnen und dabei auf eine Einführung der Wishart-Verteilung verzichten. Zu diesem Zweck nehmen wir durchgängig die *Zentrierung* des betrachteten Datensatzes $Y \in \mathbb{R}^{m \times n}$, also $\bar{y} = 0_m$, sowie die Identifizierbarkeit des Modells an. Für einen alternativen zeitgemäßen Zugang zur Parameterschätzung im Normalverteilungsmodell der Faktoranalyse mithilfe des Expectation-Maximization Algorithmus der variationalen Inferenz, siehe z.B. Rubin & Thayer (1982), Roweis & Ghahramani (1999) und Ghojogh et al. (2022). Die genauen Bezüge und qualitativen Eigenschaften der traditionellen und modernen Faktoranalyseschätzverfahren sind dabei eine offene Forschungsfrage.

Zur Diskussion des Maximum-Likelihood-Verfahrens im Normalverteilungsmodell der Faktoranalyse gehen wir wie folgt vor: Wir evaluieren zunächst die Log-Likelihood-Funktion des Normalverteilungsmodells der Faktoranalyse und definieren dann die funktionale Form der traditionellen Diskrepanzfunktion nach Lawley (1940). Wir zeigen dann, dass die Minimumstellen der Diskrepanzfunktion Maximum-Likelihood-Schätzer für das Normalverteilungsmodell der Faktoranalyse sind. Die Bestimmung von Minimumstellen der Diskrepanzfunktion mithilfe von Standardverfahren der nichtlinearen Optimierung (vgl. Rosseel (2012), Rosseel (2021)) ist also äquivalent zur Bestimmung von Maximumstellen der Log-Likelihood-Funktion des Normalverteilungsmodells der Faktoranalyse. Eine weitere Motivation für die Betrachtung der Diskrepanzfunktion ergibt sich im Kontext von Likelihood-Ration-Kriterium-basierten Modellvergleichen, wie in Kapitel 3.3.2 diskutiert.

Für die Log-Likelihood-Funktion des Normalverteilungsmodells der Faktoranalyse gilt zunächst folgendes Theorem.

Theorem 3.5 (Log-Likelihood-Funktion und Maximum-Likelihood-Schätzer des Normalverteilungsmodells der Faktoranalyse)**.** *Gegeben sei ein Faktoranalysemodell der Form*

$$y = Lf + \varepsilon \text{ mit } f \sim N(0_k, \Phi) \text{ und } \varepsilon \sim N(0_m, \Psi) \tag{3.54}$$

mit Parametervektor

$$\theta := vec(L, \Phi, \Psi), \tag{3.55}$$

und marginaler Datenkovarianzmatrix

$$\Sigma_\theta := L\Phi L^T + \Psi. \tag{3.56}$$

Weiterhin sei $Y := (y_1, ..., y_n) \in \mathbb{R}^{m \times n}$ *ein zentrierter Datensatz von* n *unabhängigen Beobachtungen von* y, C *seine Stichprobenkovarianzmatrix und*

$$S := \frac{n-1}{n} C \tag{3.57}$$

seine verzerrte Stichprobenkovarianzmatrix. Dann kann die Log-Likelihood-Funktion von Y *geschrieben werden als*

$$\ell_Y : \Theta \to \mathbb{R}, \theta \mapsto \ell_Y(\theta) := -\frac{n}{2} \ln |\Sigma_\theta| - \frac{n}{2} tr\left(S\Sigma_\theta^{-1}\right) - \frac{nm}{2} \ln(2\pi) \tag{3.58}$$

und eine Maximumstelle von ℓ_Y, *also ein Wert*

$$\hat{\theta}_{ML} = \arg\max_{\theta \in \Theta} \ell_Y(\theta), \tag{3.59}$$

heißt Maximum-Likelihood-Schätzer *für* θ.

∘

Beweis. Mit der Definition der Log-Likelihood-Funktion als und der marginalen Datenverteilung des Normalverteilungsmodells der Faktorenanalyse ergibt sich

$$\begin{aligned}
\ell_Y(\theta) &= \ln \prod_{i=1}^{n} p_\theta(y_i) \\
&= \prod_{i=1}^{n} \ln N\left(y_i; 0_m, L\Phi L^T + \Psi\right) \\
&= \ln \left(\prod_{i=1}^{n} (2\pi)^{-m/2} |\Sigma_\theta|^{-1/2} \exp\left(-\frac{1}{2} (y_i - 0_m)^T \Sigma_\theta^{-1} (y_i - 0_m)) \right) \right) \\
&= -\frac{mn}{2} \ln 2\pi - \frac{n}{2} \ln |\Sigma_\theta| - \frac{1}{2} \sum_{i=1}^{n} y_i^T \Sigma_\theta^{-1} y_i.
\end{aligned} \tag{3.60}$$

Um die Gleicheit des letzten Terms auf der rechten Seite mit dem letzten Term in der postulierten Funktionsform zu zeigen, halten wir zunächst fest, dass mit elementaren Eigenschaften der Matrixspur gilt, dass

$$\operatorname{tr}\left(\sum_{i=1}^{n} y_i^T \Sigma_\theta^{-1} y_i \right) = \sum_{i=1}^{n} \operatorname{tr}\left(y_i^T \Sigma_\theta^{-1} y_i \right) = \sum_{i=1}^{n} \operatorname{tr}\left(\Sigma_\theta^{-1} y_i y_i^T \right) = \operatorname{tr}\left(\Sigma_\theta^{-1} \sum_{i=1}^{n} y_i y_i^T \right). \tag{3.61}$$

Weiterhin gilt mit dem Binomischen Lehrsatz

$$\begin{aligned}
\sum_{i=1}^{n} y_i^T \Sigma_\theta^{-1} y_i &= \sum_{i=1}^{n} (y_i - \bar{y} + \bar{y})^T \Sigma_\theta^{-1} (y_i - \bar{y} + \bar{y}) \\
&= \sum_{i=1}^{n} (y_i - \bar{y})^T \Sigma_\theta^{-1} (y_i - \bar{y}) + \sum_{i=1}^{n} \bar{y}^T \Sigma_\theta^{-1} \bar{y} + 2 \sum_{i=1}^{n} (y_i - \bar{y})^T \Sigma_\theta^{-1} \bar{y} \\
&= \sum_{i=1}^{n} (y_i - \bar{y})^T \Sigma_\theta^{-1} (y_i - \bar{y}) + \sum_{i=1}^{n} \bar{y}^T \Sigma_\theta^{-1} \bar{y} + 2 \left(\sum_{i=1}^{n} \left(y_i - \frac{1}{n} \sum_{i=1}^{n} y_i \right)^T \right) \Sigma_\theta^{-1} \bar{y} \\
&= \sum_{i=1}^{n} (y_i - \bar{y})^T \Sigma_\theta^{-1} (y_i - \bar{y}) + \sum_{i=1}^{n} \bar{y}^T \Sigma_\theta^{-1} \bar{y} + 2 \left(\sum_{i=1}^{n} y_i^T - \frac{n}{n} \sum_{i=1}^{n} y_i^T \right) \Sigma_\theta^{-1} \bar{y} \\
&= \sum_{i=1}^{n} (y_i - \bar{y})^T \Sigma_\theta^{-1} (y_i - \bar{y}) + \sum_{i=1}^{n} \bar{y}^T \Sigma_\theta^{-1} \bar{y} + 2 \left(0_m^T \Sigma_\theta^{-1} \bar{y} \right) \\
&= \sum_{i=1}^{n} (y_i - \bar{y})^T \Sigma_\theta^{-1} (y_i - \bar{y}) + \sum_{i=1}^{n} \bar{y}^T \Sigma_\theta^{-1} \bar{y}
\end{aligned} \tag{3.62}$$

Also folgt mit obiger Eigenschaft der Matrixspur, der Zentrierung des Datensatzes $\bar{y} = 0_m$ und der Definition von S

$$\begin{aligned}
\sum_{i=1}^{n} y_i^T \Sigma_\theta^{-1} y_i &= \sum_{i=1}^{n} (y_i - \bar{y})^T \Sigma_\theta^{-1} (y_i - \bar{y}) + \sum_{i=1}^{n} \bar{y}^T \Sigma_\theta^{-1} \bar{y} \\
&= \sum_{i=1}^{n} y_i^T \Sigma_\theta^{-1} y_i \\
&= \operatorname{tr}\left(\sum_{i=1}^{n} y_i^T \Sigma_\theta^{-1} y_i \right) \\
&= \operatorname{tr}\left(\Sigma_\theta^{-1} \sum_{i=1}^{n} y_i y_i^T \right) \\
&= \operatorname{tr}\left(\Sigma_\theta^{-1} n S \right) \\
&= n \operatorname{tr}\left(S \Sigma_\theta^{-1} \right)
\end{aligned} \tag{3.63}$$

Substitution in $\ell_Y(\theta)$ von oben ergibt dann die postulierte funktionale Form der Log-Likelihood Funktion. □

Die Diskrepanzfunktion der Faktoranalyse und basierend auf ihr definierte Parameterschätzer sind nun wie folgt definiert.

Definition 3.10 (Diskrepanzfunktion der Faktoranalyse und diskrepanzfunktionsbasierte Faktoranalyseparameterschätzer)**.** Gegeben sei ein Faktoranalysemodell der Form

$$y = Lf + \varepsilon \text{ mit } f \sim N(0_k, \Phi) \text{ und } \varepsilon \sim N(0_m, \Psi) \tag{3.64}$$

mit Parametervektor

$$\theta := \operatorname{vec}(L, \Phi, \Psi), \tag{3.65}$$

und marginaler Datenkovarianzmatrix

$$\Sigma_\theta := L\Phi L^T + \Psi. \tag{3.66}$$

Weiterhin sei für einen Datensatz $Y \in \mathbb{R}^{m \times n}$ von n unabhängigen Beobachtungen von y S die verzerrte Stichprobenkovarianzmatrix von Y. Dann heißt die Funktion

$$F_Y : \Theta \to \mathbb{R}, \theta \mapsto F_Y(\theta) := n \ln |\Sigma_\theta| + n \operatorname{tr}\left(S \Sigma_\theta^{-1} \right) - n \ln |S| - nm \tag{3.67}$$

die *Diskrepanzfunktion der Faktoranalyse*. Weiterhin heißt ein Wert $\hat{\theta} \in \Theta$ mit

$$\hat{\theta} = \arg\min_{\theta \in \Theta} F_Y(\theta), \tag{3.68}$$

also eine Minimumstelle von F_Y, ein *diskrepanzfunktionsbasierte Faktoranalyseparameterschätzer*.

•

Der Vergleich mit Theorem 3.5 zeigt, dass die Log-Likelihood-Funktion und die Diskrepanzfunktion des Normalverteilungsmodells der Faktoranalyse nicht identisch sind. Allerdings kann man zeigen, dass Minimumstellen der Diskrepanzfunktion Maximumstellen der Log-Likelihood-Funktion sind und damit die Minimierung der Diskrepanzfunktion Maximum-Likelihood-Schätzer für die Parameter des Faktoranalysenormalverteilungsmodell ergibt. Die Motivation dafür, die Diskrepanzfunktion in modernen Betrachtungen der Faktoranalyse beizubehalten und nicht zugunsten der Log-Likelihood-Funktion aufzugeben (vgl. Rosseel (2021)) erschließt sich dann vor dem Hintergrund der Modellevaluation.

Theorem 3.6 (Äquivalenz von diskrepanzfunktionsbasierten und Maximum-Likelihood--Schätzern)**.** *Jede Minimumstelle der Diskrepanzfunktion der Faktoranalyse maximiert die Log-Likelihood-Funktion des Normalverteilungsmodells der Faktoranalyse.*

∘

Beweis. Wir halten zunächst fest, dass mit von θ unabhängigen Konstanten $a, b \in \mathbb{R}$ gilt, dass

$$\ell_Y(\theta) = a\left(-F_Y(\theta)\right) + b. \tag{3.69}$$

Die Log-Likelihood-Funktion ist also eine linear-affine und damit insbesondere monotone Transformation von F_Y. Da monotone Transformationen Extremstellen unverändert lassen und ein negatives Vorzeichen eine Minimumstelle in eine Maximumstelle transformiert, ergibt sich das Theorem direkt. □

Minima von F_Y werden in populären Analyseprogrammen zur Faktorenanalyse mit iterativen Standardverfahren der nichtlinearen Optimierung bestimmt. Allgemein gehen diese Verfahren von einem Startwert $\hat{\theta}^{(0)}$ aus und evaluieren weitere Iteranden rekursiv durch

$$\hat{\theta}^{(i+1)} = g\left(\hat{\theta}^{(i)}, Y\right) \text{ für } i = 0, 1, ... \tag{3.70}$$

mit einer entsprechend gewählten Funktion g des vorherigen Iteranden $\hat{\theta}^{(i)}$ und des Datensatzes Y solange, bis ein entsprechend gewähltes Abbruchkriterium erfüllt ist. Einen Überblick über die zum Beispiel im populärem **R** Paket lavaan implementierten Optimierungsalgorithmen geben Rosseel (2012) und Rosseel (2021). Wir wollen die numerische Minimierung der Diskrepanzfunktion hier nicht weiter vertiefen.

Anwendungsbeispiel

Folgender **R** Code demonstriert die Maximum-Likelihood-Schätzung der Parameter eines Faktoranalysemodells mit Faktoranalysemodell mit $k = 2$ anhand des in Abbildung 3.1 dargestellen Datensatzes nach Keller et al. (2008) mithilfe des `psych` Paketes nach Revelle (2024). Abbildung 3.4 stellt die Ergebnisse von Hauptkomponentenschätzung, iterierter Hauptachsenfaktorisierung und Maximum-Likelihood-Schätzung der Faktorladungsmatrix nebeneinander dar. Es ergeben sich qualitativ ähnliche Ladungen der beiden Faktoren auf die Items des BDI-II, allerdings mit feineren quantitativen Unterschieden.

```r
library(psych)
Y       = as.matrix(read.csv("./_data/fa-keller.csv"))          # Y \in \mathbb{R}^{m x n}
M       = fa(r = Y, nfactors = 2, rotate = "none", fm = "ml")   # Modellformulierung und -schätzung
L_hat   = M$loadings[,1:2]                                      # Faktorladungsmatrixschätzer
```

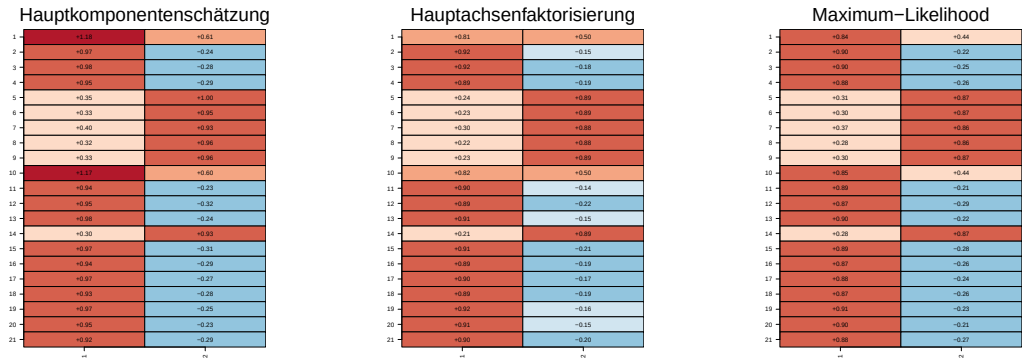

Abbildung 3.4 Faktorladungsmatrixschätzer für den BDI-II Datensatz nach Keller et al. (2008) basierend auf Hauptkomponenten-, iterierter Hauptachsenfaktorisierung- und Maximum-Likelihood-Schätzung

3.3 Modellevaluation

Grundlegende Fragen der Modellevaluation im Kontext der Faktoranalyse betreffen die Frage der Anzahl der zur Erklärung eines Datensatzes benötigten Faktoren, sowie die Überlegenheit eines spezifizierten Modells gegenüber einem parametrischen Nullmodell. Erstere Frage behandeln wir aus varianzanalytischer Sicht in Kapitel 3.3.1. Zweitere Frage behandlen wir aus Likelihood-Ratio-Sicht in Kapitel 3.3.2.

3.3.1 Stichprobenvarianzbasierte Wahl der Faktoranzahl

Die quantitative Grundlage der stichprobenvarianzbasierten Faktorzahlwahl ist die additive Zerlegung der *Gesamtstichprobenvarianz* in eine *faktorenbasierte Stichprobenvarianz* und eine *fehlerbasierte Stichprobenvarianz*. Man bestimmt die Faktorenanzahl dann anhand des Prinzips, dass bei möglichst geringer Faktorenzahl das Verhältnis von faktorenbasierter Stichprobenvarianz und Fehlervarianz möglichst groß ist. Traditionell gibt es zu diesem Zweck eine Reihe von Heuristiken. Im Folgenden zeigen zunächst die Validität der skizzierten additiven Stichprobenvarianzzerlegung und diskutieren dann einige Möglichkeiten zur Wahl der Faktorenzahl.

Definition 3.11 (Stichprobenvarianzzerlegung der Faktoranalyse)**.** $Y \in \mathbb{R}^{m\times n}$ sei ein Datensatz von n unabhängigen Beobachtungen eines Faktoranalysemodells, $C \in \mathbb{R}^{m\times m}$ sei seine Stichprobenkovarianzmatrix und $\hat{L} \in \mathbb{R}^{m\times k}$ und $\hat{\Psi} \in \mathbb{R}^{m\times m}$ seien die Hauptkomponentenschätzer $k$ter Ordnung. Dann werden

- die Summe der Diagonalemente von C als *Gesamtstichprobenvarianz*,
- die Summe der Diagonalelemente von $\hat{L}\hat{L}^T$ als *faktorbasierte Stichprobenvarianz* und
- die Summe der Diagonalelemente von $\hat{\Psi}$ als *fehlerbasierte Stichprobenvarianz*

bezeichnet.

•

Es gilt nun folgendes Theorem.

Theorem 3.7 (Stichprobenvarianzzerlegung der Faktoranalyse)**.** *Basierend auf n unabhängigen Beobachtungen eines Faktoranalysemodells seien*

- $C = (c_{ij})_{1\leq i,j\leq m} \in \mathbb{R}^{m\times m}$ *die Stichprobenkovarianzmatrix,*
- $\hat{L} = (\hat{l}_{ij})_{1\leq i\leq m, 1\leq j\leq k} \in \mathbb{R}^{m\times k}$ *der Hauptkomponentenschätzer kter Ordnung von L,*
- $\hat{\Psi} = diag(\hat{\psi}_i, ..., \hat{\psi}_m) \in \mathbb{R}^{m\times m}$ *der Hauptkomponentenschätzer kter Ordnung von Ψ,*

sowie

- $G := \sum_{i=1}^{m} c_{ii}$ *die Gesamtstichprobenvarianz,*
- $F := \sum_{i=1}^{m}\sum_{j=1}^{k} \hat{l}_{ij}^2$ *die Faktorbasierte Stichprobenvarianz,*
- $R := \sum_{i=1}^{m} \hat{\psi}_i$ *die Beobachtungsrauschenbasierte Stichprobenvarianz.*

Dann gilt

$$G = F + R. \tag{3.71}$$

Außerdem gilt mit den Eigenwerten $\lambda_1, ..., \lambda_k$ von C, dass

$$F = \sum_{j=1}^{k} \lambda_j, \text{ wobei } \lambda_j = \sum_{i=1}^{m} \hat{l}_{ij}^2 \tag{3.72}$$

für $j = 1, ..., k$ der Anteil des $j$ten Faktors an F ist.

∘

Beweis. Wir erinnern zunächst daran, dass die Diagonalelemente von $\hat{L}\hat{L}^T$ durch

$$\sum_{j=1}^{k} \hat{l}_{ij}^2 \tag{3.73}$$

gegeben sind, wovon man sich durch Betrachtung der Einträge von $\hat{L}\hat{L}^T$ überzeugt:

$$\begin{aligned} \hat{L}\hat{L}^T &= \begin{pmatrix} \hat{l}_{11} & \cdots & \hat{l}_{1k} \\ \hat{l}_{21} & \cdots & \hat{l}_{2k} \\ \vdots & \ddots & \vdots \\ \hat{l}_{m1} & \cdots & \hat{l}_{mk} \end{pmatrix} \begin{pmatrix} \hat{l}_{11} & \cdots & \hat{l}_{m1} \\ \hat{l}_{12} & \cdots & \hat{l}_{m2} \\ \vdots & \ddots & \vdots \\ \hat{l}_{1k} & \cdots & \hat{l}_{mk} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{j=1}^{k} l_{1j}^2 & \sum_{j=1}^{k} l_{1j}l_{2j} & \cdots & \sum_{j=1}^{k} l_{1j}l_{mj} \\ \sum_{j=1}^{k} l_{2j}l_{1j} & \sum_{j=1}^{k} l_{2j}^2 & \cdots & \sum_{j=1}^{k} l_{2j}l_{mj} \\ \vdots & \cdots & \ddots & \vdots \\ \sum_{j=1}^{k} l_{mj}l_{1j} & \sum_{j=1}^{k} l_{mj}l_{2j} & \cdots & \sum_{j=1}^{k} l_{mj}^2 \end{pmatrix} \end{aligned} \tag{3.74}$$

Die Identität von G und $F + R$ folgt dann direkt aus der Identität der Diagonalelemente von C, $\hat{L}\hat{L}^T$ und $\hat{\Psi}$, die im Rahmen der Hauptkomponentenschätzung mithilfe von

$$\hat{\psi}_i := c_{ii} - \sum_{j=1}^{k} \hat{l}_{ij}^2 \text{ für } i = 1, ..., m \tag{3.75}$$

konstruiert wird. Um als nächstes

$$F = \sum_{j=1}^{k} \lambda_j \tag{3.76}$$

zu zeigen halten zunächst fest, dass mit der Definition des Hauptkomponentenschätzer $\hat{L}$ die Summe der quadrierten Einträge in der $j$ten Spalte von $\hat{L}$ gleich der Summe der quadrierten Einträge in der $j$ten Spalte von $Q_k\Lambda_k^{1/2}$ ist. Dies mag man sich zum Beispiel für $m = 5$ und $k = 2$ verdeutlichen:

$$\hat{L} = Q_k\Lambda_k^{1/2} \Leftrightarrow \begin{pmatrix} \hat{l}_{11} & \hat{l}_{12} \\ \hat{l}_{21} & \hat{l}_{22} \\ \hat{l}_{31} & \hat{l}_{32} \\ \hat{l}_{41} & \hat{l}_{42} \\ \hat{l}_{51} & \hat{l}_{52} \end{pmatrix} = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \\ q_{31} & q_{32} \\ q_{41} & q_{42} \\ q_{51} & q_{52} \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix} = \begin{pmatrix} \sqrt{\lambda_1}q_{11} & \sqrt{\lambda_2}q_{12} \\ \sqrt{\lambda_1}q_{21} & \sqrt{\lambda_2}q_{22} \\ \sqrt{\lambda_1}q_{31} & \sqrt{\lambda_2}q_{32} \\ \sqrt{\lambda_1}q_{41} & \sqrt{\lambda_2}q_{42} \\ \sqrt{\lambda_1}q_{51} & \sqrt{\lambda_2}q_{52} \end{pmatrix} \tag{3.77}$$

Weiterhin halten wir fest, dass, wenn q_j für $j = 1, .., k$ die $j$te Spalte von Q_k bezeichnet aufgrund der Orthonormalität von Q folgt, dass

$$q_j^T q_j = \sum_{i=1}^{m} q_{ij}^2 = 1. \tag{3.78}$$

Dann ergibt sich für die Summe der Diagonalelemente von $\hat{L}\hat{L}^T$ aber

$$F = \sum_{i=1}^{m}\sum_{j=1}^{k} \hat{l}_{ij}^2 = \sum_{j=1}^{k}\sum_{i=1}^{m} \hat{l}_{ij}^2 = \sum_{j=1}^{k}\sum_{i=1}^{m} \left(\sqrt{\lambda}_j q_{ij}\right)^2 = \sum_{j=1}^{k} \lambda_j \sum_{i=1}^{m} q_{ij}^2 = \sum_{j=1}^{k} \lambda_j \tag{3.79}$$

Die Tatsache, dass der $j$te Eigenwert λ_j von C dabei der Anteil der durch den $j$ten Faktor erklärten Gesamtstichprobenvarianz ist ergibt sich dabei durch die Einsicht, dass der Beitrag des $j$ten Faktors in der $j$ten Spalte von $\hat{L}$ enkodiert ist und obige Gleichungskette impliziert, dass

$$\sum_{i=1}^{m} \hat{l}_{ij}^2 = \lambda_j \text{ für } j = 1, \ldots, k. \tag{3.80}$$

□

Anwendungsbeispiel

Folgender **R** Code bestimmt die Gesamtstichprobenvarianz, faktorbasierte Stichprobenvarianz und fehlerbasierte Stichprobenvarianz sowie ihre Beziehungen nach Theorem 3.2 für ein Faktoranalysemodell mit $k = 2$ anhand des in Abbildung 3.1 dargestellen Datensatzes nach Keller et al. (2008).

```
# EFA mit Hauptkomponentenschätzung für k = 2
YT        = read.csv("./_data/fa-keller.csv")                  # Y^T \in \mathbb{R}^{n x m}
Y         = as.matrix(t(YT))                                   # Y \in \mathbb{R}^{m x n}
m         = nrow(Y)                                            # Datendimension
n         = ncol(Y)                                            # Datenpunktanzahl
k         = 2                                                  # Faktoranzahl
I_n       = diag(n)                                            # Einheitsmatrix I_n
J_n       = matrix(rep(1,n^2), nrow = n)                       # 1_{nn}
C         = (1/(n-1))*(Y %*% (I_n-(1/n)*J_n) %*% t(Y))         # Stichprobenkovarianzmatrix
EA        = eigen(C)                                           # Eigenanalyse von R
lambda_k  = EA$values[1:k]                                     # k größte Eigenwerte von R
Q_k       = EA$vectors[,1:k]                                   # k zugehörige Eigenvektoren von R
L_hat     = Q_k %*% diag(sqrt(lambda_k))                       # Faktorladungsmatrixschätzer
Psi_hat   = diag(diag(C) - diag(L_hat %*% t(L_hat)))           # Beobachtungsrauschenkovarianzmatrixschätzer
GG        = sum(diag(C))                                       # Gesamtstichprobenvarianz
FF        = sum(diag(L_hat %*% t(L_hat)))                      # Faktorenbasierte Stichprobenvarianz
RR        = sum(diag(Psi_hat))                                 # Beobachtungsrauschenbasierter Stichprobenvarianz
FF_lambda = sum(lambda_k)                                      # Summe der Eigenwerte \lambda_1,...,\lambda_k
```

```
G   = 25.84
F   = 22.51
R   = 3.33
F+B = 25.84
```

```
Faktorbasierte Stichprobenvarianz F        = 22.51
Summe der Eigenwerte lambda_1,...,lambda_k = 22.51
```

Basierend auf obiger Stichprobenvarianzzerlegung kann man nun die Faktorzahl k so wählen, dass ein vorgegebener Anteil der Gesamtstichprobenvarianz durch das entsprechende Faktoranalysemodell erklärt wird. Obige Stichprobenvarianzzerlegung besagt ja gerade, dass der durch den $j$ten Faktor erklärte Anteil an der Gesamtstichprobenvarianz G durch λ_j, der durch den $j$ten Faktor erklärte relative Anteil an der Gesamtstichprobenvarianz demnach durch λ_j/G und der durch die $j = 1, ..., k$ Faktoren erklärte relative Anteil an der Gesamtstichprobenvarianz damit durch $\sum_{j=1}^{k} \lambda_j/G$ gegeben ist. Es macht also Sinn, $\lambda_j, \lambda_j/G$ und $\sum_{j=1}^{k} \lambda_j/G$ zu untersuchen und zu visualisieren. Basierend auf einer solchen Darstellung mag man dann k so wählen, dass k möglichst klein und $\sum_{j=1}^{k} \lambda_j/G$ möglichst groß ist. Die Visualisierung der λ_j wird in diesem Kontext *Scree-Plot* genannt, wobei sich das englische Wort *Scree* auf die geologische Formation einer Geröllhalde oder eines Talus bezieht und den charakteristischen zunächst schnell und dann langsam abfallenenden Anteil an erklärter Stichprobenvarianz durch die der Größe nach geordneten Eigenwerte beschreibt.

Awendungsbeispiel

```
# FA mit Hauptkomponentenschätzung für k = 5
YT        = read.csv("./_data/fa-keller.csv")                    # Y^T \in \mathbb{R}^{n x m}
Y         = as.matrix(t(YT))                                     # Y   \in \mathbb{R}^{m x n}
m         = nrow(Y)                                              # Datendimension
n         = ncol(Y)                                              # Datenpunktanzahl
k         = 5                                                    # Faktoranzahl
I_n       = diag(n)                                              # Einheitsmatrix I_n
J_n       = matrix(rep(1,n^2), nrow = n)                         # 1_{nn}
C         = (1/(n-1))*(Y %*% (I_n-(1/n)*J_n) %*% t(Y))           # Stichprobenkovarianzmatrix
EA        = eigen(C)                                             # Eigenanalyse von R
lambda_k  = EA$values[1:k]                                       # k größte Eigenwerte von R
Q_k       = EA$vectors[,1:k]                                     # k zugehörige Eigenvektoren von R
L_hat     = Q_k %*% diag(sqrt(lambda_k))                         # Faktorladungsmatrixschätzer
Psi_hat   = diag(diag(C) - diag(L_hat %*% t(L_hat)))             # Beobachtungsrauschenkovarianzmatrixschätzer
G         = sum(diag(C))                                         # Gesamtstichprobenvarianz
```

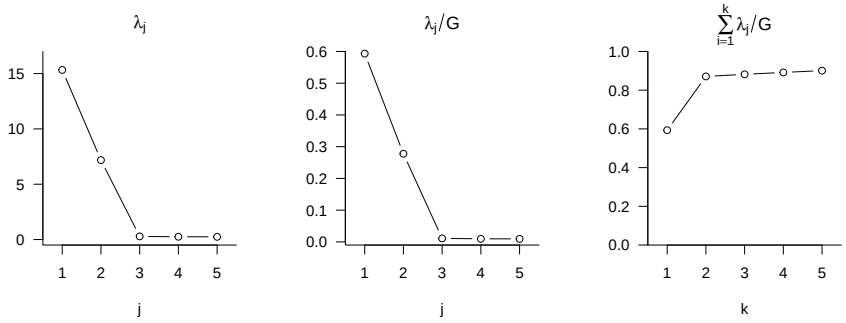

Abbildung 3.5 Scree-Plot und relative und kumulative Gesamtstichprobenvarianz durch steigende Faktorenanzahlen für den BDI-II Datensatz nach Keller et al. (2008).

Aus Abbildung 3.5 liest man ab, dass mit einer Faktoranzahl von $k = 1$ 56 % der Gesamtstichprobenvarianz, mit $k = 2$ können 87 % der Gesamtstichprobenvarianz und mit $k = 3$ 88 % der Gesamtstichprobenvarianz erklärt werden können. Der relative Anteil der durch Hinzunahme des dritten Faktors ist mit 1% also gering, so dass im Sinne der Modellparsimonität die Wahl von $k = 2$ Faktoren sinnvoll erscheint.

3.3.2 Likelihood-basierter Nullmodellvergleich

Die Normalverteilungsannahme im Faktoranalysemodell ermöglicht es, strukturierte Modellvergleiche mithilfe eines Log-Likelihood-Ratio-Kriteriums durchzuführen. Zu dessen Erläuterung seien $Y := (y_1, ..., y_n) \in \mathbb{R}^{m \times n}$ sei ein Datensatz von n unabhängigen Beobachtungen eines Faktoranalysemodells mit $\bar{y} = 0_m$ und $\text{uvec}(A, B, ...)$ sei die konkatenisierte Vektorisierung der unikalen Werte der Matrizen $A, B, ...$ sowie $\text{uvec}^{-1}(A, B, ...)$ ihre Umkehrung. Weiterhin seien M1 ein Normalverteilungsmodell der Faktoranalyse, das die Ordnungsrelation erfüllt und identifizierbar ist,

$$y \sim N(0, \Sigma_\theta) \tag{3.81}$$

mit

$$\Sigma_\theta = L\Phi L^T + \Psi, \theta = \text{uvec}(L, \Phi, \Psi) \in \Theta, \Theta \subset \mathbb{R}^p \text{ und } p \leq m(m+1)/2 \tag{3.82}$$

und M2 ein multivariates Normalverteilungsmodell mit beliebigem Kovarianzmatrixparameter,

$$y \sim N(0, \Sigma_\gamma) \text{ mit } \Sigma_\gamma = \text{uvec}^{-1}(\gamma), \gamma \in \Gamma \text{ und } \Gamma \subset \mathbb{R}^{m(m+1)/2}. \tag{3.83}$$

Das Log-Likelihood-Ratio-Kriterium zum Vergleich von M1 und M2 setzt bekanntlich die maximierten Wahrscheinlichkeitsdichtenvon Y unter M1 und M2 ins Verhältnis. Im vorliegenden Fall hat es die Form

$$\Lambda_Y := \ln\left(\frac{\max_{\theta\in\Theta}\prod_{i=1}^{n} N(y_i; 0_m, \Sigma_\theta)}{\max_{\gamma\in\Gamma}\prod_{i=1}^{n} N(y_i; 0_m, \Sigma_\gamma)}\right). \tag{3.84}$$

Große Werte von Λ_Y bedeuten, dass Y unter M1 eine größere Wahrscheinlichkeitsdichte besitzt als unter M2. Dies wird allgemein als Evidenz dafür verstanden, dass Y eher von M1 als von M2 generiert wurden. Dabei ist Λ_Y letztlich die zentrale Motivation für die funktionale Form der Diskrepanzfunktion (cf. Lawley (1940)). Folgendes Theorem zeigt nun zunächst den Zusammenhang zwischen dem Likelihood-Ratio-Kriterium Λ_Y und der im Kontext der Maximum-Likelihood-Schätzung der Faktoranalyse eingeführten Diskrepanzfunktion.

Theorem 3.8 (Diskrepanzfunktion)**.** *Für einen Datensatz $Y := (y_1, ..., y_n) \in \mathbb{R}^{m\times n}$ mit $\bar{y} = 0_m$ von n unabhängigen Beobachtungen eines Zufallsvektors y sei das Log-Likelihood-Ratio-Kriterium der Faktoranalyse gegeben durch*

$$\Lambda_Y := \ln\left(\frac{\max_{\theta\in\Theta}\prod_{i=1}^{n} N(y_i; 0_m, \Sigma_\theta)}{\max_{\gamma\in\Gamma}\prod_{i=1}^{n} N(y_i; 0_m, \Sigma_\gamma)}\right), \tag{3.85}$$

wobei

$$\Sigma_\theta = L\Phi L^T + \Psi, \theta = uvec(L, \Phi, \Psi) \in \Theta, \Theta \subset \mathbb{R}^p \text{ und } p \leq m(m+1)/2 \tag{3.86}$$

und

$$\Sigma_\gamma = uvec^{-1}(\gamma), \gamma \in \Gamma \text{ und } \Gamma \subset \mathbb{R}^{m(m+1)/2} \tag{3.87}$$

seien. Weiterhin sei für die verzerrte Stichprobenkovarianzmatrix S von Y

$$F_Y(\theta) := n\ln|\Sigma_\theta| + ntr(S\Sigma_\theta^{-1}) - n\ln|S| - nm \tag{3.88}$$

die Diskrepanzfunktion und $\hat{\theta}$ eine Minimumstelle von F_Y. Dann gilt

$$-2\Lambda_Y = F_Y(\hat{\theta}) \tag{3.89}$$

∘

Beweis. Wir halten zunächst fest, dass

$$\max_{\theta\in\Theta}\prod_{i=1}^{n} N(y_i; 0_m, \Sigma_\theta) = \prod_{i=1}^{n} N(y_i; 0_m, \Sigma_{\hat{\theta}}) \tag{3.90}$$

weil eine Minimumstelle $\hat{\theta}$ von F_Y wie in Theorem 3.6 gesehen die Log-Likelihood-Funktion und damit auch die Likelihood-Funktion eines Normalverteilungsmodells der Faktoranalyse maximiert. Weiterhin halten wir fest, dass

$$\max_{\gamma\in\Gamma}\prod_{i=1}^{n} N(y_i; 0_m, \Sigma_\gamma) = \prod_{i=1}^{n} N(y_i; 0_m, S) \tag{3.91}$$

weil die verzerrte Stichprobenkovarianz der Maximum-Likelihood-Schätzer des Kovarianzmatrixparameters einer multivariaten Normalverteilung ist. Die Logarithmuseigenschaften ergeben dann

$$\Lambda_Y = \ln\left(\frac{\prod_{i=1}^{n} N(y_i; 0_m, \Sigma_{\hat{\theta}})}{\prod_{i=1}^{n} N(y_i; 0_m, S)}\right) = \sum_{i=1}^{n}\ln N(y_i; 0_m, \Sigma_{\hat{\theta}}) - \sum_{i=1}^{n}\ln N(y_i; 0_m, S) \tag{3.92}$$

Substition der funktionalen Form der Log-Likelihood-Funktion der konfirmatorischen Faktoranalyse und der funktionalen Form der WDF der multivariaten Normalverteilung ergibt

$$\begin{aligned}
\Lambda_Y &= \sum_{i=1}^{n} \ln N\left(y_i; 0_m, \Sigma_{\hat{\theta}}\right) - \sum_{i=1}^{n} \ln N(y_i; 0_m, S) \\
&= -\frac{mn}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma_\theta| - \frac{n}{2}\mathrm{tr}\left(S\Sigma_\theta^{-1}\right) - \frac{n}{2}\bar{y}^T\Sigma_\theta^{-1}\bar{y} \\
&\quad + \frac{mn}{2}\ln(2\pi) + \frac{n}{2}\ln|S| + \frac{n}{2}\mathrm{tr}\left(SS^{-1}\right) + \frac{n}{2}\bar{y}^T S^{-1}\bar{y} \\
&= -\frac{n}{2}\ln|\Sigma_\theta| - \frac{n}{2}\mathrm{tr}\left(S\Sigma_\theta^{-1}\right) - \frac{n}{2}0_m^T\Sigma_\theta^{-1}0_m \\
&\quad + \frac{n}{2}\ln|S| + \frac{n}{2}\mathrm{tr}\left(SS^{-1}\right) + \frac{n}{2}0_m^T S^{-1}0_m \\
&= -\frac{n}{2}\ln|\Sigma_\theta| - \frac{n}{2}\mathrm{tr}\left(S\Sigma_\theta^{-1}\right) + \frac{n}{2}\ln|S| + \frac{mn}{2}
\end{aligned} \tag{3.93}$$

Multiplikation mit -2 ergibt schließlich

$$-2\Lambda_Y = n\ln|\Sigma_\theta| + n\,\mathrm{tr}\left(S\Sigma_\theta^{-1}\right) - n\ln|S| - mn = F_Y(\theta). \tag{3.94}$$

□

3.4 Modellinterpretation

Da die Werte des beobachtbaren Zufallsvektors eines Faktorenanalyse-Modells als Punkte in einem kanonischen Koordinatensystem aufgefasst werden, entspricht der j-te Eintrag in der i-ten Zeile einer Faktorenladungsmatrix dem Wert der beobachteten Variable y_i für f, unter der Annahme, dass f den Wert des j-ten kanonischen Einheitsvektors annimmt. Wie in Theorem 3.3 gesehen, bleibt das grundlegende Theorem der Faktorenanalyse jedoch erfüllt, wenn sowohl die Faktorkoordinaten als auch die kanonischen Beobachtungen in den Zeilen der Faktorenladungsmatrix auf eine beliebige andere orthogonale Basis projiziert werden (Lawley & Maxwell (1971)). Mit anderen Worten: In Bezug auf die Korrelationsmatrix des beobachtbaren Zufallsvektors ist der Parameter der Faktorenladungsmatrix eben nur bis zu einer beliebigen orthogonalen Koordinatentransformation eindeutig bestimmt — im Jargon der Faktorenanalyse als eine *orthogonale Rotation* bezeichnet. Diese Eigenschaft, die die Parameter des Faktorenanalysemodells nicht-identifizierbar macht, hat Forschende dazu angeregt, Kriterien zur Auswahl einer bestimmten Faktorenladungsmatrix aus ihrer unendlichen Menge gleichermaßen gültiger Werte zu entwickeln.

Varimax-Rotation

Eine Kriterium, das eine *einfache Struktur (simple structure)* nach Thurstone (1937), der geschätzten Faktorladungsmatrix begünstigen soll, ist das Varimax-Kriterium nach Kaiser (1958). Im Wesentlichen formuliert das Varimax-Kriterium eine Zielfunktion, um die Summe der Varianzen der quadrierten Einträge der Spalten der Faktorenladungsmatrix zu maximieren. Auf diese Weise sollen Spalten bevorzugt werden, deren Einträge entweder nahe am Spaltenmittelwert der Quadrate oder stark positiv bzw. negativ davon abweichend sind und damit ein Muster bevorzugt werden, das in modernen Begriffen als *sparse matrix* bezeichnet wird (Rohe & Zeng (2023)). Entscheidend ist, dass das Varimax-Kriterium eine Funktion der auf die ursprüngliche Faktorenladungsmatrix angewandten orthogonalen Rotation ist, dargestellt durch die Multiplikation mit einer orthogonalen

Matrix (Deisenroth et al. (2020)). Das Varimax-Problem besteht somit darin, die orthogonale Matrix zu identifizieren, die das Varimax-Kriterium für eine gegebene Anfangs-Faktorenladungsmatrix maximiert. Eine so erhaltene Version der Faktorenladungsmatrix ist hoffentlich leicht interpretierbar, da sie direkt Gruppen von beobachtbaren Variablen mit kanonischen Einheitsvektor-Faktorwerten verknüpft und somit den Faktoren im Hinblick auf die von ihnen getragenen beobachtbaren Variablen Bedeutung verleiht. Der spezifische Varimax-Ansatz der **R** Funktion `varimax()` implementiert einen Algorithmus, der als *simultane Faktoren-Varimax-Lösung* bezeichnet wird und von Horst (1965) und Lawley & Maxwell (1971) entwickelt wurde.

Im Folgenden geben wir daher zunächst einen kurzen Überblick über die Varimax-Zielfunktion, die verwendet wird, um eine einfache Struktur der Faktorenladungsmatrix zu erreichen, betrachten dann das daraus resultierende Optimierungsproblem mit Nebenbedingungen, gegen dann die notwendige Bedingung für ein Maximum der Lagrange-Funktion dieses Problems an und betrachten schließlich über den iterativen Algorithmus, der in `varimax()` implementiert ist, um die notwendige Bedingung für eine orthogonale Koordinatentransformation zu lösen. Zu diesem Zweck sei durchgängig

$$\hat{L}_M = \left(\hat{l}_{M_{ij}}\right) 1 \leq i \leq m, 1 \leq j \leq k := \hat{L}M \tag{3.95}$$

die koordinatentransformierte Version einer Schätzung der Faktorenladungsmatrix $\hat{L} \in \mathbb{R}^{m \times k}$, sodass die Koordinaten der Zeilenvektoren von $\hat{L}$ bezüglich der Basisvektoren ausgedrückt werden, die die Spalten der orthogonalen Matrix $M \in \mathbb{R}^{k \times k}$ bilden. Im Jargon der Faktorenanalyse wird $\hat{L}_M$ üblicherweise als die *rotierte Faktorenladungsmatrix* bezeichnet und M als die *Rotationsmatrix*.

Definition 3.12. Gegeben sei ein Schätzer $\hat{L}$ für die Faktorladungsmatrix eines Faktoranalysemodells. Dann ist die Varimax-Zielfunktion definiert als

$$v : \mathbb{R}_{\mathrm{O}}^{k \times k} \to \mathbb{R}, M \mapsto v(M) := \sum_{j=1}^{k} \sum_{i=1}^{m} \left(\hat{l}_{M_{ij}}^2 - \frac{1}{m} \sum_{r=1}^{m} \hat{l}_{M_{rj}}^2 \right)^2 = \sum_{j=1}^{k} \sum_{i=1}^{m} \left(\hat{l}_{M_{ij}}^4 - \frac{1}{m} \left(\sum_{r=1}^{m} \hat{l}_{M_{rj}}^2 \right)^2 \right), \tag{3.96}$$

wobei $\mathbb{R}_{\mathrm{O}}^{k \times k}$ die Menge aller orthogonalen $k \times k$-Matrizen bezeichnen soll.

•

Für eine gegebene Faktorenladungsmatrix bewertet die Varimax-Zielfunktion v also die Summe der Abweichungsquadrate der quadrierten Einträge von $\hat{L}_M$, wobei jede Abweichung vom entsprechenden Spaltenmittelwert als Funktion der Rotationsmatrix M gemessen wird. Es sei angemerkt, dass die Gleichheit auf der rechten Seite von Gleichung 3.96 einige algebraische Umformungen erfordert, die hier nicht ausgeführt werden.

Die Grundidee des Varimax-Ansatzes besteht nun darin, v bezüglich der orthogonalen Matrix M zu maximieren. Dies führt auf das nichtlineare Optimierungsproblem

$$\max v(M) \text{ unter der Nebenbedingung } M^T M = I_k. \tag{3.97}$$

Mittels der Differentialrechnung für Matrizen konnte gezeigt werden (vgl. Horst (1965), Lawley & Maxwell (1971), Neudecker (1969), Magnus & Neudecker (1989)), dass die notwendige Bedingung für ein Maximum der Lagrange-Funktion des Optimierungsproblems Gleichung 3.97 in der Gleichung

$$MA_M = B_M \tag{3.98}$$

ausgedrückt werden kann, wobei $A_M \in \mathbb{R}^{k \times k}$ eine symmetrische und positiv definite Matrix unbekannter Lagrange-Multiplikatoren ist und

$$B_M := \hat{L}^T \left(C_M - \frac{1}{m} \hat{L}_M D_M \right) \tag{3.99}$$

mit

$$\hat{L}_M := \hat{L}M, C_M := \left(\hat{l}^3_{M_{ij}} \right)_{1 \leq i \leq m, 1 \leq j \leq k} \text{ und } D_M := \text{diag} \left(\sum_{i=1}^{m} \hat{l}^2_{M_{i1}}, ..., \sum_{i=1}^{m} \hat{l}^2_{M_{ik}} \right) \tag{3.100}$$

ist. Bemerkenswert ist, dass Gleichung 3.98 lediglich eine implizite Definition des optimalen M liefert, da beide Seiten der Gleichung von M abhängen. Um dieses Gleichungssystem iterativ nach M zu lösen, geht der in `varimax()` implementierte Ansatz der simultanen Faktoren-Varimax-Lösung nach folgendem Algorithmus vor.

Definition 3.13 (Simultane-Faktoren-Varimax-Lösungsalgorithmus)**.** Gegeben sei ein Schätzer $\hat{L}$ für die Faktorladungsmatrix eines Faktoranalysemodells. Dann hat der Algorithmus der **R** Funktion `varimax()` die folgende Form:

Initialisierung

(SN) Setze $\hat{L}_N := N\hat{L}$ mit $N := \text{diag} \left(\left(\sum_{j=1}^{k} \hat{l}^2_{1j} \right)^{-\frac{1}{2}}, ..., \left(\sum_{j=1}^{k} \hat{l}^2_{mj} \right)^{-\frac{1}{2}} \right)$

(S0) Setze $M^{(0)} := I_k, \delta^{(0)} := 0, \delta^{(1)} := 1$ und definiere $\epsilon^{(0)}$

Iterationen

Für $i = 1, ..., 1000$ und solange $d^{(i)} \geq d^{(i-1)}(1 + \epsilon)$

(S1) Setze $\hat{L}_{M^{(i)}} = \hat{L}_N M^{(i-1)}$

(S2) Setze $B_{M^{(i)}} = \hat{L}_N^T \left(C_{M^{(i)}} - \frac{1}{m} \hat{L}_{M^{(i)}} D_{M^{(i)}} \right)$

(S3) Evaluiere $B_{M^{(i)}} = U^{(i)} \Delta^{(i)} V^{(i)^T}$

(S4) Setze $M^{(i)} = U^{(i)} V^{(i)^T}$

(S5) Evaluiere $d^{(i)} := \text{tr}(\Delta^{(i)})$

Finalisierung

(S6) Setze $\hat{L}_{NM} := \hat{L}_N M^{(i)}$ und $\hat{L}_M := N^{-1} \hat{L}_{NM}$

•

In Definition 3.13 normalisiert der Initialisierungsschritt (SN) jede Zeile der Faktorladungsmatrixschätzung $\hat{L}$ auf die euklidische Einheitslänge durch Multiplikation mit einer $m \times m$ diagonalen Normalisierungsmatrix N, deren Diagonalelemente den reziproken euklidischen Längen der Zeilen von $\hat{L}$ entsprechen. Obwohl nicht unumstritten (vgl. Kaiser (1970), Kaiser & Rice (1974), Rohe & Zeng (2023)), wurde diese zeilenweise Normalisierung bei der Einführung der Varimax-Rotation durch Kaiser (1958) vorgeschlagen und bildet das Standardverfahren in der **R** Implementierung der Varimax-Prozedur. Die Rotationsmatrix wird dann als $k \times k$ Einheitsmatrix initialisiert, die Konvergenzvariablen δ werden auf 0 gesetzt und ein geeigneter Wert für das Konvergenzkriterium ϵ wird definiert. Die Iterationsschritte (S1) und (S2) evaluieren dann die relevanten Iterandenmatrizen für

die rechte Seite von Gleichung 3.98, $L_{M^{(i)}}$, $D_{M^{(i)}}$, $C_{M^{(i)}}$, und $B_{M^{(i)}}$, wodurch die rekursive Gleichung

$$M^{(i+1)} A_{M^{(i)}} = B_{M^{(i)}} \tag{3.101}$$

implizit definiert wird. Diese implizite rekursive Gleichung wird dann für $M^{(i+1)}$ mittels einer Singulärwertzerlegung von $B_{M^{(i)}}$ gelöst, wobei $M^{(i+1)}$ in den Iterationsschritten (S3) und (S4) auf $U^{(i)}V^{(i)}$ gesetzt wird. Hierbei sind $U^{(i)}$ und $V^{(i)}$ die Matrizen der linken bzw. rechten Singulärvektoren von $B_{M^{(i)}}$, und wir begründen diese Lösung untenstehend. Die Konvergenz des Algorithmus wird dann anhand der Spur der Singulärwertmatrix $\Delta^{(i)}$ im Iterationsschritt (S5) bewertet. Schließlich wird nach der Konvergenz die optimierte Transformationsmatrix $M^{(i)}$ auf die normalisierte Faktorladungsmatrixschätzung angewendet und die Normalisierungsskalierung in Schritt (S6) rückgängig gemacht.

Wir wollen schließlich den Iterationsschritt (S4) begründen, welche die notwendige Bedingung für ein Maximum der Lagrange-Funktion des restringierten Optimierungsproblems Gleichung Gleichung 3.97 in der Form von Gleichung 3.98 iterativ mittels einer Singulärwertzerlegung der Matrix auf der rechten Seite von Gleichung 3.98 löst. Dieser Ansatz nutzt die Eigenschaften der beteiligten Matrizen und deren Beziehungen. Zur Vereinfachung der Notation verzichten wir im Folgenden auf die Matrixabhängigkeit von M sowie auf den Iterations-Superskript und betrachten die Lösung von $MA = B$ für bekanntes B und unbekannte (aber positiv-definite und symmetrische) Lagrange-Multiplikatormatrix A. Zunächst stellen wir fest, dass die Prämultiplikation von $MA = B$ mit der Transponierten von B ergibt, dass

$$MA = B \Leftrightarrow B^T MA = B^T B \Leftrightarrow (MA)^T MAB^T B \Leftrightarrow A^T M^T MA = B^T B \Leftrightarrow A^2 = B^T B, \tag{3.102}$$

wobei die letzte Äquivalenz aus der Orthogonalität von M und der Symmetrie von A folgt.

Basierend auf Gleichung 3.102 formulieren wir als Nächstes zwei Ausdrücke für die unbekannte Matrix A^2. Erstens betrachten wir die Singulärwertzerlegung von $B = USV^T$, wobei U die orthogonale Matrix der linken Singulärvektoren, S die Diagonalmatrix der Singulärwerte und V die orthogonale Matrix der rechten Singulärvektoren ist. Wir erhalten dann:

$$A^2 = B^T B = \left(USV^T\right)^T USV^T = VSU^T USV^T = VS^2V^T. \tag{3.103}$$

Da A eine positiv-definite und symmetrische Matrix ist, kann sie als Zerlegung in ihre orthogonale Matrix von Eigenvektoren Q und ihre Diagonalmatrix von Eigenwerten Λ geschrieben werden. Wir haben also auch:

$$A^2 = AA = Q\Lambda Q^T Q\Lambda Q^T = Q\Lambda^2 Q^T \tag{3.104}$$

und wir erhalten:

$$A^2 = VS^2V^T = Q\Lambda^2 Q^T = A^2. \tag{3.105}$$

Ohne Beschränkung der Allgemeinheit können wir $V = Q$ annehmen, sodass $S^2 = \Lambda^2$ und damit $S = \Lambda$ bis auf Vorzeichenpermutationen gilt. Wir können nun Gleichung 3.98 wie folgt lösen:

$$MA = B \Leftrightarrow M = BA^{-1} \Leftrightarrow M = USV^T Q\Lambda^{-1} Q^T \Leftrightarrow M = USQ^T QS^{-1}V^T \Leftrightarrow M = UV^T, \tag{3.106}$$

was schließlich den Iterationsschritt (S4) rechtfertigt.

Anwendungsbeispiel

Folgender **R** Code demonstriert die Implemtnation von Definition 3.13.

```
N             = solve(diag(sqrt(diag(L_hat %*% t(L_hat)))))     # Zeilennormalisierende Matrix
L_hat_N       = N %*% L_hat                                     # zeilennormalisierte Faktorladungsmatrixschätzer
m             = nrow(L_hat_N)                                   # Zeilenanzahl
k             = ncol(L_hat_N)                                   # Spaltenanzahl
Mi            = diag(k)                                         # Transformationsmatrixinitialisierung
d             = 0                                               # Konvergenzvariableninitialisierung
eps           = 1e-05                                           # Konvergenzkriterium (change factor)
for (i in 1L:100L){                                             # Iterationen
L_hat_Mi  = L_hat_N %*% Mi                                      # \hat{L}_M^{(i)}
    Ci        = L_hat_Mi^3                                      # C^{(i)}
    Di        = diag(drop(rep(1, m) %*% L_hat_Mi^2))            # D^{(i)}
    Bi        = t(L_hat_N) %*% (Ci - (1/m) * L_hat_Mi %*% Di)   # B^{(i)}
    UDVi      = svd(Bi)                                         # Singulärwertzerlegung von B^{(i)}
    Mi        = UDVi$u %*% t(UDVi$v)                            # Transformationsmatrix Update M^{(i)}
    dd        = sum(UDVi$d)                                     # Konvergenzkriterium Update
    if (dd < d * (1 + eps)){break}                              # Konvergenztest
    d         = dd}                                             # Konvergenzvariablen Update
L_hat_NM      = L_hat_N %*% Mi                                  # Transformierter Faktorladungsmatrixschätzer
L_hat_M       = solve(N) %*% L_hat_NM                           # Zeilenormalisierungsaufhebung
```

4 Theoretische Ergänzungen

4.1 Erwartungswerte

Wir erinnern zunächst an den Begriff des Erwartungswerts einer Zufallsvariable.

Definition 4.1 (Erwartungswert einer Zufallsvariable)**.** $(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum und ξ sei eine Zufallsvariable. Dann ist der *Erwartungswert von* ξ definiert als

- $\mathbb{E}(\xi) := \sum_{x \in \mathcal{X}} x\, p(x)$, wenn $\xi : \Omega \to \mathcal{X}$ diskret mit WMF p ist.
- $\mathbb{E}(\xi) := \int_{-\infty}^{\infty} x\, p(x)\, dx$, wenn $\xi : \Omega \to \mathbb{R}$ kontinuierlich mit WDF p ist.

•

Der Erwartungswert ist also eine skalare Zusammenfassung der Verteilung einer Zufallsvariable. Eine Definition des Erwartungswertes, die ohne eine Fallunterscheidung in kontinuierliche und diskrete Zufallsvariablen auskommt, ist möglich, erfordert aber mit der Einführung des Lebesgue-Integrals einigen technischen Aufwand. Wir verweisen dafür auf die weiterführende Literatur (vgl. Schmidt (2009), Meintrup & Schäffler (2005)). Intuitiv entspricht der Erwartungswert einer Zufallsvariable dem im langfristigen Mittel zu erwartenden Wert der Zufallsvariable. Wir betrachten drei Beispiele für Definition 4.1.

Beispiele

Theorem 4.1 (Erwartungswert einer diskreten Zufallsvariable)**.** *ξ sei eine Zufallsvariable mit Ergebnisraum $\mathcal{X} := \{-1, 0, 1\}$ und WMF*

$$p(-1) = \frac{1}{4}, \quad p(0) = \frac{1}{2}, \quad p(1) = \frac{1}{4}. \tag{4.1}$$

Dann gilt

$$\mathbb{E}(\xi) = 0. \tag{4.2}$$

∘

Beweis. Nach Definition 4.1 ergibt sich

$$\begin{aligned} \mathbb{E}(\xi) &= \sum_{x \in \mathcal{X}} x p(x) \\ &= -1 \cdot p(-1) + 0 \cdot p(0) + 1 \cdot p(1) \\ &= -1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} \\ &= 0. \end{aligned} \tag{4.3}$$

□

Theorem 4.2 (Erwartungswert einer Bernoulli Zufallsvariable)**.** *Es sei* $\xi \sim Bern(\mu)$*. Dann gilt* $\mathbb{E}(\xi) = \mu$*.*

∘

Beweis. ξ ist diskret mit $\mathcal{X} = \{0, 1\}$. Also gilt nach Definition 4.1

$$\begin{aligned} &= \sum_{x \in \{0,1\}} x \operatorname{Bern}(x; \mu) \\ &= 0 \cdot \mu^0 (1-\mu)^{1-0} + 1 \cdot \mu^1 (1-\mu)^{1-1} \\ &= 1 \cdot \mu^1 (1-\mu)^0 \\ &= \mu. \end{aligned} \tag{4.4}$$

□

Theorem 4.3 (Erwartungswert einer normalverteilten Zufallsvariable)**.** *Es sei* $\xi \sim N(\mu, \sigma^2)$*. Dann gilt* $\mathbb{E}(\xi) = \mu$*.*

∘

Beweis. Wir verzichten auf einen Beweis. □

In Verallgemeinerung von Definition 4.1 gilt folgende Definition für den Erwartungswert einer Funktion einer Zufallsvariable.

Definition 4.2 (Erwartungswert einer Funktion einer Zufallsvariable)**.** $(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum, ξ sei eine Zufallsvariable mit Ergebnisraum $\mathcal{X}$ und $f : \mathcal{X} \to \mathcal{Z}$ sei eine Funktion mit Zielmenge $\mathcal{Z}$. Dann ist der *Erwartungswert der Funktion* f *der Zufallsvariable* ξ definiert als

- $\mathbb{E}(f(\xi)) := \sum_{x \in \mathcal{X}} f(x)\, p(x)$, wenn $\xi : \Omega \to \mathcal{X}$ diskret mit WMF p ist,
- $\mathbb{E}(f(\xi)) := \int_{-\infty}^{\infty} f(x)\, p(x)\, dx$, wenn $\xi : \Omega \to \mathbb{R}$ kontinuierlich mit WDF p ist.

•

Der Erwartungswert einer Zufallsvariable ergibt sich anhand von Definition 4.2 als der Spezialfall, in dem gilt, dass

$$f : \mathcal{X} \to \mathcal{Z}, x \mapsto f(x) := x. \tag{4.5}$$

In der englischsprachigen Literatur ist Definition 4.2 auch als *law of the unconscious statistician* bekannt und wird oft auch direkt zur Definition des Erwartungswertes herangezogen. Folgendes Theorem gibt ein erstes Beispiel für den Erwartungswert einer Funktion einer Zufallsvariable.

Theorem 4.4 (Erwartungswert bei linear-affiner Transformation einer Zufallsvariable)**.** *ξ sei eine Zufallsvariable mit Ergebnisraum $\mathcal{X}$ und es sei*

$$f : \mathcal{X} \to \mathcal{Z}, x \mapsto f(x) := ax + b \text{ für } a, b \in \mathbb{R} \tag{4.6}$$

eine linear-affine Funktion. Dann gilt

$$\mathbb{E}(f(\xi)) = \mathbb{E}(a\xi + b) = a\mathbb{E}(\xi) + b. \tag{4.7}$$

∘

Beweis. Die Aussage des Theorems folgt mit den Linearitätseigenschaften von Summen und Integralen. Wir betrachten den Fall einer diskreten Zufallsvariable ξ mit endlichem Ergebnisraum $\mathcal{X}$ und WMF p. Es gilt

$$\begin{aligned}\mathbb{E}(f(\xi)) &= \mathbb{E}(a\xi + b) \\ &= \sum_{x\in\mathcal{X}} (ax + b)p(x) \\ &= \sum_{x\in\mathcal{X}} axp(x) + bp(x) \\ &= a\sum_{x\in\mathcal{X}} xp(x) + b\sum_{x\in\mathcal{X}} p(x) \\ &= a\mathbb{E}(\xi) + b.\end{aligned} \tag{4.8}$$

□

In Analogie zu Definition 4.2 definiert man für die Funktion eines Zufallsvektors den Erwartungswert dieser Transformation wie folgt.

Definition 4.3 (Erwartungswert einer Funktion eines Zufallsvektors)**.** $(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum, ξ sei ein Zufallsvektor mit Ergebnisraum $\mathcal{X}$ und $f : \mathcal{X} \to \mathcal{Z}$ sei eine Funktion mit Zielmenge $\mathcal{Z}$. Dann ist der *Erwartungswert der Funktion f des Zufallsvektors ξ* definiert als

- $\mathbb{E}(f(\xi)) := \sum_{x\in\mathcal{X}} f(x)\,p(x)$, wenn $\xi : \Omega \to \mathcal{X}$ diskret mit WMF p ist,
- $\mathbb{E}(f(\xi)) := \int_{-\infty}^{\infty} f(x)\,p(x)\,dx$, wenn $\xi : \Omega \to \mathbb{R}$ kontinuierlich mit WDF p ist.

•

Theorem 4.5 (Erwartungswert bei linear-affiner Kombination)**.** *ξ sei ein n-dimensionaler Zufallsvektor mit Komponenten $\xi_1, ..., \xi_n$ und Ergebnisraum $\mathcal{X} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$. Weiterhin sei*

$$f : \mathcal{X} \to \mathcal{Z}, x \mapsto f(x) := a_0 + \sum_{i=1}^{n} a_i x_i \text{ für } a_0, ..., a_n \in \mathbb{R}. \tag{4.9}$$

eine linear-affine Kombination der Komponenten von ξ. Dann gilt

$$\mathbb{E}(f(\xi)) = \mathbb{E}\left(a_0 + \sum_{i=1}^{n} a_i\xi_i\right) = a_0 + \sum_{i=1}^{n} a_i\mathbb{E}(\xi_i). \tag{4.10}$$

∘

Beweis. Wie unten gezeigt, folgt das Theorem mit den Linearitätseigenschaften von Summen und Integralen, sowie Fubini's Theorem zur Vertauschung von Summations- bzw. Integrationsreihenfolge. Zur

Vereinfachung der Notation schreiben wir untenstehend $\sum_{x_i \in \mathcal{X}_i}$ als $\sum_{x_i}$. Es gilt dann

$$\begin{aligned}
\mathbb{E}(f(\xi)) &= \mathbb{E}\left(a_0 + \sum_{i=1}^{n} a_i \xi_i\right) \\
&= \mathbb{E}(a_0 + a_1\xi_1 + \cdots + a_n\xi_n) \\
&= \sum_{x_1} \cdots \sum_{x_n} (a_0 + a_1 x_1 + \cdots a_n x_n) p(x_1, \ldots, x_n) \\
&= \sum_{x_1} \cdots \sum_{x_n} a_0 p(x_1, \ldots, x_n) + a_1 x_1 p(x_1, \ldots, x_n) + \cdots + a_n x_n p(x_1, \ldots, x_n) \\
&= \sum_{x_1} \cdots \sum_{x_n} a_0 p(x_1, \ldots, x_n) + \sum_{x_1} \cdots \sum_{x_n} a_1 x_1 p(x_1, \ldots, x_n) + \cdots + \sum_{x_1} \cdots \sum_{x_n} a_n x_n p(x_1, \ldots, x_n) \\
&= a_0 \sum_{x_1} \cdots \sum_{x_n} p(x_1, \ldots, x_n) + a_1 \sum_{x_1} \cdots \sum_{x_n} x_1 p(x_1, \ldots, x_n) + \cdots + a_n \sum_{x_n} \cdots \sum_{x_1} x_n p(x_1, \ldots, x_n) \\
&= a_0 + a_1 \sum_{x_1} x_1 \sum_{x_2} \cdots \sum_{x_n} p(x_1, \ldots, x_n) + \cdots + a_n \sum_{x_n} x_n \sum_{x_{n-1}} \cdots \sum_{x_1} p(x_1, \ldots, x_n) \\
&= a_0 + a_1 \sum_{x_1} x_1 p(x_1) + \cdots + a_n \sum_{x_n} x_n p(x_n) \\
&= a_0 + a_1 \mathbb{E}(\xi_1) + \cdots + a_n \mathbb{E}(\xi_n) \\
&= a_0 + \sum_{i=1}^{n} a_i \mathbb{E}(\xi_i)
\end{aligned} \tag{4.11}$$

□

Bedingter Erwartungswert

Betrachtet man bei der Bildung eines Erwartungswertes nun anstelle der Verteilung einer Zufallsvariable die bedingte Verteilung einer Zufallsvariable, so ist man entsprechend auf den Begriff des bedingten Erwartungswerts geführt.

Definition 4.4 (Bedingter Erwartungswert)**.** Gegeben sei ein Zufallsvektor $\xi := (\xi_1, \xi_2)$ mit Ergebnisraum $\mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2$, WMF oder WDF $p(x_1, x_2)$ und bedingter WMF oder WDF $p(x_1|x_2)$ für alle $x_2 \in \mathcal{X}_2$. Dann ist der *bedingte Erwartungswert* von ξ_1 gegeben $\xi_2 = x_2$ definiert als

- $\mathbb{E}(\xi_1|\xi_2 = x_2) := \sum_{x_1 \in \mathcal{X}_1} x_1 p(x_1|x_2)$, wenn ξ ein diskreter Zufallsvektor ist.
- $\mathbb{E}(\xi_1|\xi_2 = x_2) := \int_{\mathcal{X}_1} x_1 p(x_1|x_2)\, dx_1$, wenn ξ ein kontinuierlicher Zufallsvektor ist.

•

Der bedingte Erwartungswert einer Zufallsvariable ist also der Erwartungswert einer Zufallsvariable in einer bedingten Verteilung. Man beachte, dass wir in Definition 4.4 den bedingten Erwartungswert für einen festen Wert x_2 von ξ_2 definiert habe. Bei einem festen Wert x_2 von ξ_2 ist $\mathbb{E}(\xi_1|\xi_2 = x_2)$ damit ein fester Wert und durch Austauschen der Subskripte erhält man entsprechend $\mathbb{E}(\xi_2|\xi_1 = x_1)$.

Allgemein ist $\mathbb{E}(\xi_1|\xi_2)$ allerdings eine Zufallsvariable, da ξ_2 eine Zufallsvariable ist und nur mit einer gewissen Wahrscheinlichkeit einen Wert $\xi_2 = x_2$ annimmt. Wir werden später sehen, dass analog zu Definition 4.4 bedingte Varianzen, bedingte Kovarianzen und bedingte Korrelationen definiert werden. In der Psychologie ist der Begriff des bedingten Erwartungswerts zentral, denn in der klassischen Testtheorie sind die wahren Werte von

Personen bei Testmessungen
als bedingte Erwartungswerte definiert. Zur Verdeutlichung von Definition 4.4 betrachten wir ein Beispiel für einen diskreten Zufallsvektor.

Beispiel

Für einen zweidimensionalen Zufallsvektor $\xi := (\xi_1, \xi_2)$, der Werte in $\mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2$ mit $\mathcal{X}_1 := \{1, 2, 3\}$ und $\mathcal{X}_2 = \{1, 2, 3, 4\}$ annehme seien bedingte WMFen für $p(x_2|x_1)$ für alle $x_1 \in \mathcal{X}_1$ wie in Tabelle 4.1 spezifiziert

Tabelle 4.1 Bedingte Wahrscheinlichkeitsverteilung von x_2 gegeben x_1

$p(x_2\|x_1)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$
$p(x_2\|x_1 = 1)$	$\frac{1}{4}$	0	$\frac{1}{2}$	$\frac{1}{4}$
$p(x_2\|x_1 = 2)$	$\frac{1}{3}$	$\frac{2}{3}$	0	0
$p(x_2\|x_1 = 3)$	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Dann ergeben sich die bedingten Erwartungswerte

$$\begin{aligned}
\mathbb{E}(\xi_2|\xi_1 = 1) &= \sum_{x_2 \in \mathcal{X}_2} x_2 p(x_2|x_1 = 1) = 1 \cdot \frac{1}{4} + 2 \cdot 0 + 3 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} = \frac{11}{4} \\
\mathbb{E}(\xi_2|\xi_1 = 2) &= \sum_{x_2 \in \mathcal{X}_2} x_2 p(x_2|x_1 = 2) = 1 \cdot \frac{1}{3} + 2 \cdot \frac{2}{3} + 3 \cdot 0 + 4 \cdot 0 = \frac{5}{3} \\
\mathbb{E}(\xi_2|\xi_1 = 3) &= \sum_{x_2 \in \mathcal{X}_2} x_2 p(x_2|x_1 = 3) = 1 \cdot 0 + 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{3} + 4 \cdot \frac{1}{3} = \frac{8}{3}.
\end{aligned} \tag{4.12}$$

Das folgende Theorem, der sogenannte *Satz vom iterierten Erwartungswert*, stellt einen Zusammenhang zwischen dem (marginalen) Erwartungswert einer Zufallsvariable und dem bedingten Erwartungswert dieser Zufallsvariable her. Das Theorem wird manchmal auch als *Satz vom totalen Erwartungswert* bezeichnet.

Theorem 4.6 (Satz vom iterierten Erwartungswert)**.** *Gegeben sei ein Zufallsvektor* $\xi := (\xi_1, \xi_2)$*. Dann gilt*

$$\mathbb{E}(\xi_1) = \mathbb{E}\left(\mathbb{E}\left(\xi_1|\xi_2\right)\right). \tag{4.13}$$

∘

Beweis. Wir beschränken uns auf den Beweis für einen diskreten Zufallsvektor, der Beweis für einen kon-

tinuierlichen Zufallsvektor folgt analog. Es gilt

$$\begin{aligned}
\mathbb{E}(\xi_1) &= \sum_{x_1 \in \mathcal{X}_1} x_1 p(x_1) \\
&= \sum_{x_1 \in \mathcal{X}_1} x_1 \sum_{x_2 \in \mathcal{X}_2} p(x_1, x_2) \\
&= \sum_{x_1 \in \mathcal{X}_1} x_1 \sum_{x_2 \in \mathcal{X}_2} p(x_1|x_2) p(x_2) \\
&= \sum_{x_1 \in \mathcal{X}_1} \sum_{x_2 \in \mathcal{X}_2} x_1 p(x_1|x_2) p(x_2) \\
&= \sum_{x_2 \in \mathcal{X}_2} \sum_{x_1 \in \mathcal{X}_1} x_1 p(x_1|x_2) p(x_2) \\
&= \sum_{x_2 \in \mathcal{X}_2} \mathbb{E}(\xi_1|\xi_2 = x_2) p(x_2) \\
&= \mathbb{E}\left(\mathbb{E}(\xi_1|\xi_2)\right).
\end{aligned} \tag{4.14}$$

□

Offenbar bezeichnen die verschiedenen Erwartungswerte in Theorem 4.6 Erwartungswert bezüglich unterschiedlicher Verteilungen: der Erwartungswert auf der linken Seite der Gleichung bezeichnet den Erwartungswert bezüglich der marginalen Verteilung von ξ_1, der äußere Erwartungswert auf der rechten Seite der Gleichung bezeichnet den Erwartungswert bezüglich der marginalen Verteilung von ξ_2 und der innere Erwartungswert auf der rechten Seite der Gleichung bezeichnet den bedingten Erwartungswert von ξ_1 gegeben ξ_2

4.2 Varianzen

Definition 4.5 (Varianz)**.** ξ sei eine Zufallsvariable mit endlichem Erwartungswert $\mathbb{E}(\xi)$. Dann ist die *Varianz von* ξ definiert als

$$\mathbb{V}(\xi) := \mathbb{E}\left((\xi - \mathbb{E}(\xi))^2\right). \tag{4.15}$$

•

Die Varianz einer Zufallsvariable ξ mit Ergebnisraum $\mathcal{X}$ ist nach Definition 4.2 also der Erwartungswert der Funktion

$$f : \mathcal{X} \to \mathcal{Z}, x \mapsto f(x) := (x - \mathbb{E}(\xi))^2 \tag{4.16}$$

und damit die erwartete quadrierte Abweichung einer Zufallsvariable von ihrem Erwartungswert. Die Quadrierung der Abweichung der Zufallsvariable von ihrem Erwartungswert ist dabei nötig, da andernfalls mit Theorem 4.4 immer gelten würde, dass

$$\mathbb{E}(\xi - \mathbb{E}(\xi)) = \mathbb{E}(\xi) - \mathbb{E}(\xi) = 0. \tag{4.17}$$

Intuitiv misst die Varianz also die *Streuung* oder *Variabilität* einer Zufallvariable. Insbesondere besagt die *Chebyshev Ungleichung*

$$\mathbb{P}(|\xi - \mathbb{E}(\xi)| \geq x) \leq \frac{\mathbb{V}(\xi)}{x^2}. \tag{4.18}$$

Beispielweise gilt für eine Zufallsvariable immer, dass die Wahrscheinlichkeit für eine absolute Abweichung vom doppelten ihrer Standardabweichung höchstens 1/4 ist, also

Frequentistisch betrachtet nur für etwa ein Viertel ihrer Realisierungen zutrifft, und die Wahrscheinlichkeit für eine absolute Abweichung vom dreifachen ihrer Standardabweichung höchstens 1/9 ist, also Frequentistisch betrachtet nur für etwa ein Zehntel ihrer Realisierungen zutrifft, jeweils unabhängig davon, von welcher genauen Form die Verteilung der Zufallsvariable ist. Dies folgt mit der Chebyshev Ungleichung aus

$$\mathbb{P}\left(|\xi - \mathbb{E}(\xi)| \geq 2\sqrt{\mathbb{V}(\xi)}\right) \leq \frac{\mathbb{V}(\xi)}{\left(2\sqrt{\mathbb{V}(\xi)}\right)^2} = \frac{1}{4} \tag{4.19}$$

und

$$\mathbb{P}\left(|\xi - \mathbb{E}(\xi)| \geq 3\sqrt{\mathbb{V}(\xi)}\right) \leq \frac{\mathbb{V}(\xi)}{\left(3\sqrt{\mathbb{V}(\xi)}\right)^2} = \frac{1}{9}. \tag{4.20}$$

Neben der Varianz gibt es viele weitere Maße für die Variabilität von Zufallsvariablen. Hier seien beispielsweise die erwartete absolute Abweichung einer Zufallsvariable von ihrem Erwartungswert $\mathbb{E}(|\xi - \mathbb{E}(\xi)|)$ und die sogenannte *Entropie* $-\mathbb{E}(\ln p(x))$ genannt. Wir verdeutlichen Definition 4.5 zunächst an einigen Beispielen.

Beispiele

Theorem 4.7. *ξ sei eine Zufallsvariable mit Ergebnisraum $\mathcal{X} := \{-1, 0, 1\}$ und WMF*

$$p(-1) = \frac{1}{4} \quad p(0) = \frac{1}{2} \quad p(1) = \frac{1}{4}. \tag{4.21}$$

Dann gilt

$$\mathbb{V}(\xi) = \frac{1}{2}. \tag{4.22}$$

∘

Beweis. Nach Definition 4.5 ergibt sich mit dem in Theorem 4.1 bestimmten Erwartungswert derselben Zufallsvariable

$$\begin{aligned}
\mathbb{V}(\xi) &= \mathbb{E}\left((\xi - \mathbb{E}(\xi))^2\right) \\
&= \mathbb{E}\left((\xi - 0)^2\right) \\
&= \mathbb{E}\left(\xi^2\right) \\
&= \sum_{x \in \mathcal{X}} x^2 \, p_\xi(x) \\
&= (-1)^2 \cdot p_\xi(-1) + 0^2 \cdot p_\xi(0) + 1^2 \cdot p_\xi(1) \\
&= 1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} \\
&= \frac{1}{2}.
\end{aligned} \tag{4.23}$$

□

Theorem 4.8 (Varianz einer Bernoulli-Zufallsvariable)**.** *Es sei $\xi \sim Bern(\mu)$. Dann ist die Varianz von ξ gegeben durch*

$$\mathbb{V}(\xi) = \mu(1 - \mu). \tag{4.24}$$

∘

Beweis. ξ ist eine diskrete Zufallsvariable und es gilt $\mathbb{E}(\xi) = \mu$. Also gilt

$$\begin{aligned}
\mathbb{V}(\xi) &= \mathbb{E}\left((\xi - \mu)^2\right) \\
&= \sum_{x \in \{0,1\}} (x - \mu)^2 \text{Bern}(x; \mu) \\
&= (0 - \mu)^2 \mu^0 (1 - \mu)^{1-0} + (1 - \mu)^2 \mu^1 (1 - \mu)^{1-1} \\
&= \mu^2 (1 - \mu) + (1 - \mu)^2 \mu \\
&= \left(\mu^2 + (1 - \mu)\mu\right)(1 - \mu) \\
&= \left(\mu^2 + \mu - \mu^2\right)(1 - \mu) \\
&= \mu(1 - \mu).
\end{aligned} \tag{4.25}$$

□

Theorem 4.9 (Varianz einer normalverteilten Zufallsvariable)**.** *Es sei* $\xi \sim N(\mu, \sigma^2)$*. Dann ist die Varianz von* ξ *gegeben durch*

$$\mathbb{V}(\xi) = \sigma^2. \tag{4.26}$$

∘

Beweis. Wir verzichten auf einen Beweis. □

Die Varianz hat die Eigenschaft, keine negativen Werte anzunehmen.

Theorem 4.10 (Nichtnegativität der Varianz)**.** ξ *sei eine Zufallsvariable. Dann gilt*

$$\mathbb{V}(\xi) \geq 0. \tag{4.27}$$

∘

Beweis. Wir betrachten den Fall einer diskreten Zufallsvariable. Dann gilt zunächst

$$(x - \mathbb{E}(\xi))^2 \geq 0 \text{ für alle } x \in \mathcal{X}. \tag{4.28}$$

Weiterhin gilt für die WMF p von ξ, dass

$$p(x) \geq 0 \text{ für alle } x \in \mathcal{X}. \tag{4.29}$$

Also folgt

$$p(x)(x - \mathbb{E}(\xi))^2 \geq 0 \text{ für alle } x \in \mathcal{X}. \tag{4.30}$$

Damit gilt dann aber

$$\mathbb{V}(\xi) = \sum_{x \in \mathcal{X}} p(x)(x - \mathbb{E}(\xi))^2 \, dx \geq 0, \tag{4.31}$$

Analog zeigt man die Nichtnegativität der Varianz bei kontinuierlichen Zufallsvariablen. □

Das Berechnen von Varianzen wird durch folgendes Theorem, den sogenannten *Varianzverschiebungssatz* oft erleichtert, insbesondere, wenn der Erwartungswert der quadrierten Zufallsvariable leicht zu bestimmen oder bekannt ist.

Theorem 4.11 (Varianzverschiebungssatz)**.** ξ *sei eine Zufallsvariable. Dann gilt*

$$\mathbb{V}(\xi) = \mathbb{E}\left(\xi^2\right) - \mathbb{E}(\xi)^2. \tag{4.32}$$

∘

Beweis. Mit der Definition der Varianz und der Linearität des Erwartungswerts gilt

$$\begin{aligned}\mathbb{V}(\xi) &= \mathbb{E}\left((\xi - \mathbb{E}(\xi))^2\right) \\ &= \mathbb{E}\left(\xi^2 - 2\xi\mathbb{E}(\xi) + \mathbb{E}(\xi)^2\right) \\ &= \mathbb{E}(\xi^2) - 2\mathbb{E}(\xi)\mathbb{E}(\xi) + \mathbb{E}\left(\mathbb{E}(\xi)^2\right) \\ &= \mathbb{E}(\xi^2) - 2\mathbb{E}(\xi)^2 + \mathbb{E}(\xi)^2 \\ &= \mathbb{E}(\xi^2) - \mathbb{E}(\xi)^2.\end{aligned} \tag{4.33}$$

□

In Analogie zu Theorem 4.4 gilt für die Varianz folgendes Resultat.

Theorem 4.12 (Varianz bei linear-affiner Tansformation einer Zufallsvariable)**.** *ξ sei eine Zufallsvariable und es sei*

$$f : \mathcal{X} \to \mathcal{Z}, x \mapsto f(x) := ax + b \text{ für } a, b \in \mathbb{R} \tag{4.34}$$

eine linear-affine Funktion. Dann gilt

$$\mathbb{V}(f(\xi)) = \mathbb{V}(a\xi + b) = a^2\mathbb{V}(\xi). \tag{4.35}$$

∘

Beweis. Es ergibt sich

$$\begin{aligned}\mathbb{V}(f(\xi)) &= \mathbb{V}(a\xi + b) \\ &= \mathbb{E}\left((a\xi + b - \mathbb{E}(a\xi + b))^2\right) \\ &= \mathbb{E}\left((a\xi + b - a\mathbb{E}(\xi) - b)^2\right) \\ &= \mathbb{E}\left((a\xi - a\mathbb{E}(\xi))^2\right) \\ &= \mathbb{E}\left(a(\xi - \mathbb{E}(\xi))^2\right) \\ &= \mathbb{E}\left(a^2(\xi - \mathbb{E}(\xi))^2\right) \\ &= a^2\mathbb{E}\left((\xi - \mathbb{E}(\xi))^2\right) \\ &= a^2\mathbb{V}(\xi).\end{aligned} \tag{4.36}$$

□

Bedingte Varianz

Definition 4.6 (Bedingte Varianz)**.** Gegeben sei ein Zufallsvektor $\xi := (\xi_1, \xi_2)$ mit Ergebnisraum $\mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2$. Dann ist die *bedingte Varianz von ξ_1 gegeben $\xi_2 = x_2$* definiert als

$$\mathbb{V}(\xi_1|\xi_2 = x_2) = \mathbb{E}\left(\left(\xi_1 - \mathbb{E}(\xi_1|\xi_2 = x_2)\right)^2 |\xi_2 = x_2\right) \tag{4.37}$$

und die *bedingte Standardabweichung von ξ_1 gegeben $\xi_2 = x_2$* ist definiert als

$$\mathbb{S}(\xi_1|\xi_2 = x_2) = \sqrt{\mathbb{V}(\xi_1|\xi_2 = x_2)}. \tag{4.38}$$

•

Die bedingte Varianz ist im Sinne des bedingten Erwartungswerts definiert, ebenso wie ein bedingter Erwartungswert ist eine bedingte Varianz also eine Zufallsvariable. Auch für die bedingte Varianz gilt der Verschiebungssatz, wie folgendes Theorem besagt.

Theorem 4.13 (Kovarianzverschiebungssatz der bedingten Varianz)**.** *Gegeben sei ein Zufallsvektor* $\xi := (\xi_1, \xi_2)$*. Dann gilt*

$$\mathbb{V}(\xi_1|\xi_2 = x_2) = \mathbb{E}\left(\xi_1^2|\xi_2 = x_2\right) - \mathbb{E}(\xi_1|\xi_2 = x_2)^2 \tag{4.39}$$

∘

Beweis. Wir beschränken uns auf den Beweis für einen diskreten Zufallsvektor. Der Beweis für einen kontinuierlichen Zufallsvektor folgt analog. Mit der Definition des bedingten Erwartungswerts (Definition 4.4) gilt

$$\begin{aligned}
\mathbb{V}(\xi_1|\xi_2 = x_2) &= \mathbb{E}\left(\left(\xi_1 - \mathbb{E}(\xi_1|\xi_2 = x_2)\right)^2|\xi_2 = x_2\right) \\
&= \mathbb{E}\left(\xi_1^2 - 2\xi_1\mathbb{E}(\xi_1|\xi_2 = x_2) + \mathbb{E}(\xi_1|\xi_2 = x_2)^2|\xi_2 = x_2\right) \\
&= \sum_{x_1 \in \mathcal{X}_1} p(x_1|x_2)\left(x_1^2 - 2x_1\mathbb{E}(\xi_1|\xi_2 = x_2) + \mathbb{E}(\xi_1|\xi_2 = x_2)^2\right) \\
&= \sum_{x_1 \in \mathcal{X}_1} x_1^2 p(x_1|x_2) - \sum_{x_1 \in \mathcal{X}_1} 2x_1\mathbb{E}(\xi_1|\xi_2 = x_2)p(x_1|x_2) + \sum_{x_1 \in \mathcal{X}_1} \mathbb{E}(\xi_1|\xi_2 = x_2)^2 p(x_1|x_2) \\
&= \mathbb{E}(\xi_1^2|\xi_2 = x_2) - 2\mathbb{E}(\xi_1|\xi_2 = x_2)\sum_{x_1 \in \mathcal{X}_1} x_1 p(x_1|x_2) + \mathbb{E}(\xi_1|\xi_2 = x_2)^2 \sum_{x_1 \in \mathcal{X}_1} p(x_1|x_2) \\
&= \mathbb{E}(\xi_1^2|\xi_2 = x_2) - 2\mathbb{E}(\xi_1|\xi_2 = x_2)\mathbb{E}(\xi_1|\xi_2 = x_2) + \mathbb{E}(\xi_1|\xi_2 = x_2)^2 \cdot 1 \\
&= \mathbb{E}(\xi_1^2|\xi_2 = x_2) - 2\mathbb{E}(\xi_1|\xi_2 = x_2)^2 + \mathbb{E}(\xi_1|\xi_2 = x_2)^2 \\
&= \mathbb{E}(\xi_1^2|\xi_2 = x_2) - \mathbb{E}(\xi_1|\xi_2 = x_2)^2
\end{aligned} \tag{4.40}$$

□

Folgendes Theorem, der sogenannte *Satz von der iterierten Varianz* stellt einen Zusammenhang zwischen der Varianz einer Zufallsvariable, ihrer bedingten Varianz und der Varianz ihres bedingten Erwartungswerts her.

Theorem 4.14 (Satz von der iterierten Varianz)**.** *Gegeben sei ein Zufallsvektor* $\xi := (\xi_1, \xi_2)$*. Dann gilt*

$$\mathbb{V}(\xi_1) = \mathbb{E}(\mathbb{V}(\xi_1|\xi_2)) + \mathbb{V}(\mathbb{E}(\xi_1|\xi_2)) \tag{4.41}$$

∘

Beweis. Wir beschränken uns auf den Beweis für einen diskreten Zufallsvektor. Der Beweis für einen kontinuierlichen Zufallsvektor folgt analog. Mit Kovarianzverschiebungssatz (Theorem 4.16), dem Satz vom iterierten Erwartungswert (Theorem 4.6) und dem Kovarianzverschiebungssatz der bedingten Varianz (Theorem 4.13) gilt

$$\begin{aligned}
\mathbb{V}(\xi_1) &= \mathbb{E}(\xi_1^2) - \mathbb{E}(\xi_1)^2 \\
&= \mathbb{E}\left(\mathbb{E}(\xi_1^2|\xi_2)\right) - \mathbb{E}\left(\mathbb{E}(\xi_1|\xi_2)\right)^2 \\
&= \sum_{x_2 \in \mathcal{X}_2} p(x_2)\mathbb{E}(\xi_1^2|\xi_2 = x_2) - \mathbb{E}\left(\mathbb{E}(\xi_1|\xi_2)\right)^2 \\
&= \sum_{x_2 \in \mathcal{X}_2} p(x_2)\left(\mathbb{V}(\xi_1|\xi_2 = x_2) + \mathbb{E}(\xi_1|\xi_2 = x_2)^2\right) - \mathbb{E}\left(\mathbb{E}(\xi_1|\xi_2)\right)^2 \\
&= \sum_{x_2 \in \mathcal{X}_2} p(x_2)\mathbb{V}(\xi_1|\xi_2 = x_2) + \sum_{x_2 \in \mathcal{X}_2} p(x_2)\mathbb{E}(\xi_1|\xi_2 = x_2)^2 - \mathbb{E}\left(\mathbb{E}(\xi_1|\xi_2)\right)^2 \\
&= \mathbb{E}(\mathbb{V}(\xi_1|\xi_2)) + \left(\mathbb{E}\left(\mathbb{E}(\xi_1|\xi_2)^2\right) - \mathbb{E}\left(\mathbb{E}(\xi_1|\xi_2)\right)^2\right) \\
&= \mathbb{E}(\mathbb{V}(\xi_1|\xi_2)) + \mathbb{V}\left(\mathbb{E}(\xi_1|\xi_2)\right)
\end{aligned} \tag{4.42}$$

□

4.3 Kovarianzen

Definition 4.7 (Kovarianz)**.** Die *Kovarianz* zweier Zufallsvariablen ξ_1 und ξ_2 ist definiert als

$$\mathbb{C}(\xi_1, \xi_2) := \mathbb{E}\left((\xi_1 - \mathbb{E}(\xi_1))(\xi_2 - \mathbb{E}(\xi_2))\right). \tag{4.43}$$

•

Die Definition der Kovarianz nach Definition 4.7 lässt implizit, dass die in ihr auftauchenden Erwartungswerte unterschiedlicher Natur sind. Dabei geht die Definition der Kovarianz zunächst einmal davon aus, dass ξ_1 und ξ_2 die Komponenten eines Zufallsvektors $\xi := (\xi_1, \xi_2)$ mit Ergebnisraum $\mathcal{X}_1 \times \mathcal{X}_2$, gemeinsamer Verteilung $\mathbb{P}(\xi_1, \xi_2)$ und marginalen Verteilungen $\mathbb{P}(\xi_1)$ und $\mathbb{P}(\xi_2)$ sind. Bei den Erwartungswerten $\mathbb{E}(\xi_1)$ und $\mathbb{E}(\xi_2)$ handelt es sich um die Erwartungswerte der Komponenten ξ_1 und ξ_2 bezüglich dieser Marginalverteilungen. Die Kovarianz selbst ist dann der Erwartungswert der Funktion

$$f : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Z}, (x_1, x_2) \mapsto f(x_1, x_2) := (x_1 - \mathbb{E}(\xi_1))(x_2 - \mathbb{E}(\xi_2)) \tag{4.44}$$

bezüglich der gemeinsamen Verteilung $\mathbb{P}(\xi_1, \xi_2)$. Wir wollen dies am Beispiel eines diskreten Zufallsvektors verdeutlichen.

Beispiel

Es $\xi := (\xi_1, \xi_2)$ ein diskreter Zufallsvektor mit Ergebnisraum $\mathcal{X} := \{1, 2\} \times \{1, 2, 3\}$ und der in Tabelle 4.2 dargestellten gemeinsamen WMF $p(x_1, x_2)$ und marginalen WMFen $p(x_1)$ und $p(x_2)$.

Tabelle 4.2 Gemeinsame und marginale WMFen des Zufallsvektors ξ

$p(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$p(x_1)$
$x_1 = 1$	0.10	0.05	0.15	0.30
$x_1 = 2$	0.60	0.05	0.05	0.70
$p(x_2)$	0.70	0.10	0.20	

Mit der Definition der Kovarianz von ξ_1 und ξ_2 gilt dann unter Beachtung von Gleichung 4.44

$$\begin{aligned}
\mathbb{C}(\xi_1, \xi_2) &= \mathbb{E}(f(\xi_1, \xi_2)) \\
&= \sum_{x_1=1}^{2} \sum_{x_2=1}^{3} f(x_1, x_2) p(x_1, x_2) \\
&= \sum_{x_1=1}^{2} \sum_{x_2=1}^{3} (x_1 - \mathbb{E}(\xi_1))(x_2 - \mathbb{E}(\xi_2)) p(x_1, x_2)
\end{aligned} \tag{4.45}$$

Einsetzen der Werte aus Tabelle 4.2 ergibt dann zunächst

$$\mathbb{E}(\xi_1) = \sum_{x_1=1}^{2} x_1 p(x_1) = 1 \cdot 0.3 + 2 \cdot 0.7 = 1.7 \tag{4.46}$$

und

$$\mathbb{E}(\xi_2) = \sum_{x_2=1}^{3} x_2 p(x_2) = 1 \cdot 0.7 + 2 \cdot 0.1 + 3 \cdot 0.2 = 1.5. \tag{4.47}$$

und damit dann schließlich

$$\begin{aligned}
\mathbb{C}(\xi_1, \xi_2) &= \sum_{x_1=1}^{2} \sum_{x_2=1}^{3} (x_1 - 1.7)\,(x_2 - 1.5)\,p(x_1, x_2) \\
&= \sum_{x_1=1}^{2} (x_1 - 1.7)(1 - 1.5)p(x_1, 1) + (x_1 - 1.7)(2 - 1.5)p(x_1, 2) + (x_1 - 1.7)(3 - 1.5)p(x_1, 3) \\
&= \quad (1 - 1.7)(1 - 1.5)p(1, 1) + (1 - 1.7)(2 - 1.5)p(1, 2) + (1 - 1.7)(3 - 1.5)p(1, 3) \\
&\quad\; (2 - 1.7)(1 - 1.5)p(2, 1) + (2 - 1.7)(2 - 1.5)p(2, 2) + (2 - 1.7)(3 - 1.5)p(2, 3) \\
&= (-0.7) \cdot (-0.5) \cdot 0.10 + (-0.7) \cdot 0.5 \cdot 0.05 + (-0.7) \cdot 1.5 \cdot 0.15 \\
&\quad + 0.3 \cdot (-0.5) \cdot 0.60 \; + 0.3 \cdot 0.5 \cdot 0.05 \quad + 0.3 \cdot 1.5 \cdot 0.05 \\
&= 0.035 - 0.0175 - 0.1575 - 0.09 + 0.0075 + 0.0225 \\
&= -0.2.
\end{aligned} \tag{4.48}$$

Die Kovarianz der Zufallsvariablen ξ_1 und ξ_2 mit der in obiger Tabelle festgelegter Verteilung ist also $\mathbb{C}(\xi_1, \xi_2) = -0.2$. Im Gegensatz zur Varianz kann die Kovarianz also offenbar auch negative Werte annehmen.

Wir betrachten im Folgenden einige grundlegende Eigenschaften der Kovarianz.

Theorem 4.15 (Symmetrie der Kovarianz)**.** *ξ_1 und ξ_2 seien zwei Zufallsvariablen. Dann gilt*

$$\mathbb{C}(\xi_1, \xi_2) = \mathbb{C}(\xi_2, \xi_1) \tag{4.49}$$

∘

Beweis. Mit der Kommutativität der Multiplikation gilt

$$\mathbb{C}(\xi_1, \xi_2) = \mathbb{E}\left((\xi_1 - \mathbb{E}(\xi_1))(\xi_2 - \mathbb{E}(\xi_2))\right) = \mathbb{E}\left((\xi_2 - \mathbb{E}(\xi_2))(\xi_1 - \mathbb{E}(\xi_1))\right) = \mathbb{C}(\xi_2, \xi_1) \tag{4.50}$$

□

Wie das Berechnen von Varianzen wird auch das Berechnen von Kovarianzen manchmal durch folgendes Theorem erleichtert.

Theorem 4.16 (Kovarianzverschiebungssatz)**.** *ξ_1 und ξ_2 seien Zufallsvariablen. Dann gilt*

$$\mathbb{C}(\xi_1, \xi_2) = \mathbb{E}(\xi_1 \xi_2) - \mathbb{E}(\xi_1)\mathbb{E}(\xi_2). \tag{4.51}$$

∘

Beweis. Mit Definition 4.7 gilt

$$\begin{aligned}
\mathbb{C}(\xi_1, \xi_2) &= \mathbb{E}\left((\xi_1 - \mathbb{E}(\xi_1))(\xi_2 - \mathbb{E}(\xi_2))\right) \\
&= \mathbb{E}\left(\xi_1 \xi_2 - \xi_1 \mathbb{E}(\xi_2) - \mathbb{E}(\xi_1)\xi_2 + \mathbb{E}(\xi_1)\mathbb{E}(\xi_2)\right) \\
&= \mathbb{E}(\xi_1 \xi_2) - \mathbb{E}(\xi_1)\mathbb{E}(\xi_2) - \mathbb{E}(\xi_1)\mathbb{E}(\xi_2) + \mathbb{E}(\xi_1)\mathbb{E}(\xi_2) \\
&= \mathbb{E}(\xi_1 \xi_2) - \mathbb{E}(\xi_1)\mathbb{E}(\xi_2).
\end{aligned} \tag{4.52}$$

□

Natürlich ist Theorem 4.16 nur dann wirklich nützlich, wenn $\mathbb{E}(\xi_1\xi_2)$ leicht zu berechnen sind. Der Varianzverschiebungssatz nach Theorem 4.11 ergibt sich aus Theorem 4.16 für den Fall $\xi_1 = \xi_2 := \xi$ anhand von

$$\mathbb{V}(\xi) = \mathbb{C}(\xi, \xi) = \mathbb{E}(\xi\xi) - \mathbb{E}(\xi)\mathbb{E}(\xi) = \mathbb{E}(\xi^2) - \mathbb{E}(\xi)\mathbb{E}(\xi). \tag{4.53}$$

Auch in Hinblick auf die Kovarianz $\mathbb{C}(\xi_1, \xi_2)$ ist man daran interessiert, wie sich diese unter Anwendung linear-affiner Transformationen auf ξ_2 und ξ_2 verhält. Folgendes Resultat zeigt, dass die Kovarianz zweier Zufallsvariablen von ihrem Maßstab abhängt.

Theorem 4.17 (Kovarianz bei linear-affinen Transformationen)**.** *ξ_1 und ξ_2 seien Zufallsvariablen mit gemeinsamen Ergebnisraum $\mathcal{X}_1 \times \mathcal{X}_2$ und es seien*

$$\begin{aligned} f_1 : \mathcal{X}_1 \to \mathcal{Z}_1, x_1 \mapsto f(x_1) := ax_1 + b \\ f_2 : \mathcal{X}_2 \to \mathcal{Z}_2, x_2 \mapsto f(x_2) := cx_2 + d \end{aligned} \tag{4.54}$$

für $a, b, c, d \in \mathbb{R}$ zwei linear-affine Funktionen. Dann gilt

$$\mathbb{C}(f_1(\xi_1), f(\xi_2)) = \mathbb{C}(a\xi_1 + b, c\xi_2 + d) = ac\mathbb{C}(\xi_1, \xi_2). \tag{4.55}$$

∘

Beweis. Es gilt

$$\begin{aligned} \mathbb{C}(a\xi_1 + b, c\xi_2 + d) &= \mathbb{E}((a\xi_1 + b - \mathbb{E}(a\xi_1 + b))(c\xi_2 + d - \mathbb{E}(c\xi_2 + d))) \\ &= \mathbb{E}((a\xi_1 + b - a\mathbb{E}(\xi_1) - b)(c\xi_2 + d - c\mathbb{E}(\xi_2) - d)) \\ &= \mathbb{E}(a(\xi_1 - \mathbb{E}(\xi_1))(c(\xi_2 - c\mathbb{E}(\xi_2))) \\ &= \mathbb{E}(ac((\xi_1 - \mathbb{E}(\xi_1))(\xi_2 - c\mathbb{E}(\xi_2)))) \\ &= ac\mathbb{C}(\xi_1, \xi_2). \end{aligned} \tag{4.56}$$

□

Mithilfe des Begriffes des Kovarianz ist es möglich, weitgreifenderer Aussagen über die Varianzen von Summen und Differenzen von Zufallsvariablen zu treffen als es in Kapitel 4.2 der Fall war, wo lediglich *unabhängige* Zufallsvariablen betrachtetet wurden. Mit dem folgenden Theorem betrachten wir zunächst den sehr allgemeinen Fall der Kovarianz zweier Zufallsvariablen, die sich als Linearkombinationen von n Zufallsvariablen $\xi_1, ..., \xi_n$ und m Zufallsvariablen $\zeta_1, ..., \zeta_m$ unter Addition reeller Konstanten ergeben. Hieraus ergeben sich dann eine Reihe von in der Anwendung, insbesondere in der Klassischen Testtheorie, wichtiger Spezialfälle.

Theorem 4.18 (Kovarianz von Linearkombinationen von Zufallsvariablen)**.** *Gegeben seien n Zufallsvariablen $\xi_1, ..., \xi_n$ und $n + 1$ reelle Konstanten $a_0, a_1, ..., a_n$ sowie m Zufallsvariablen $\zeta_1, ..., \zeta_m$ und $m + 1$ reelle Konstanten $b_0, b_1, ..., b_m$. Dann gilt*

$$\mathbb{C}\left(a_0 + \sum_{i=1}^{n} a_i\xi_i, b_0 + \sum_{j=1}^{m} b_j\zeta_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_j\mathbb{C}(\xi_i, \zeta_j). \tag{4.57}$$

∘

Beweis.

$$
\begin{aligned}
&\mathbb{C}\left(a_0 + \sum_{i=1}^{n} a_i\xi_i, b_0 + \sum_{j=1}^{m} b_j\zeta_j\right) \\
&= \mathbb{E}\left(\left(a_0 + \sum_{i=1}^{n} a_i\xi_i - \mathbb{E}\left(a_0 + \sum_{i=1}^{n} a_i\xi_i\right)\right)\left(b_0 + \sum_{j=1}^{m} b_j\zeta_j - \mathbb{E}\left(b_0 + \sum_{j=1}^{m} b_j\zeta_j\right)\right)\right) \\
&= \mathbb{E}\left(\left(a_0 + \sum_{i=1}^{n} a_i\xi_i - a_0 - \mathbb{E}\left(\sum_{i=1}^{n} a_i\xi_i\right)\right)\left(b_0 + \sum_{j=1}^{m} b_j\zeta_j - b_0 - \mathbb{E}\left(\sum_{j=1}^{m} b_j\zeta_j\right)\right)\right) \\
&= \mathbb{E}\left(\left(\sum_{i=1}^{n} a_i\xi_i - \sum_{i=1}^{n} a_i\mathbb{E}(\xi_i)\right)\left(\sum_{j=1}^{m} b_j\zeta_j - \sum_{j=1}^{m} b_j\mathbb{E}(\zeta_j)\right)\right) \\
&= \mathbb{E}\left(\sum_{i=1}^{n} a_i\xi_i \sum_{j=1}^{m} b_j\zeta_j - \sum_{i=1}^{n} a_i\xi_i \sum_{j=1}^{m} b_j\mathbb{E}(\zeta_j) - \sum_{i=1}^{n} a_i\mathbb{E}(\xi_i) \sum_{j=1}^{m} b_j\zeta_j + \sum_{i=1}^{n} a_i\mathbb{E}(\xi_i) \sum_{j=1}^{m} b_j\mathbb{E}(\zeta_j)\right) \\
&= \mathbb{E}\left(\sum_{i=1}^{n} a_i\xi_i \sum_{j=1}^{m} b_j\zeta_j\right) - \mathbb{E}\left(\sum_{i=1}^{n} a_i\xi_i \sum_{j=1}^{m} b_j\mathbb{E}(\zeta_j)\right) - \mathbb{E}\left(\sum_{i=1}^{n} a_i\mathbb{E}(\xi_i) \sum_{j=1}^{m} b_j\zeta_j\right) + \mathbb{E}\left(\sum_{i=1}^{n} a_i\mathbb{E}(\xi_i) \sum_{j=1}^{m} b_j\mathbb{E}(\zeta_j)\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_j\mathbb{E}(\xi_i\zeta_j) - 2\sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_j\mathbb{E}(\xi_i)\mathbb{E}(\zeta_j) + \sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_j\mathbb{E}(\xi_i)\mathbb{E}(\zeta_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_j\left(\mathbb{E}(\xi_i\zeta_j) - 2\mathbb{E}(\xi_i)\mathbb{E}(\zeta_j) + \mathbb{E}(\xi_i)\mathbb{E}(\zeta_j)\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_j\mathbb{E}\left((\xi_i - \mathbb{E}(\xi_i))(\zeta_j - \mathbb{E}(\zeta_j))\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_j\mathbb{C}(\xi_i, \zeta_j)
\end{aligned}
$$

□

Mit folgenden Theorem betrachten wir nun einen ersten Spezialfall von Theorem 4.18.

Theorem 4.19 (Kovarianz bei paarweiser Addition von Zufallsvariablen)**.** *$\xi_1, \xi_2, \zeta_1, \zeta_2$ seien vier Zufallsvariablen. Dann gilt*

$$\mathbb{C}(\xi_1 + \xi_2, \zeta_1 + \zeta_2) = \mathbb{C}(\xi_1, \zeta_1) + \mathbb{C}(\xi_1, \zeta_2) + \mathbb{C}(\xi_2, \zeta_1) + \mathbb{C}(\xi_2, \zeta_2). \tag{4.58}$$

∘

Beweis. Es seien $n := 2, m := 2, a_0 = b_0 := 0$ und $a_i = b_i := 1$ für $i = 1, 2$. Dann gilt mit Theorem 4.18

$$
\begin{aligned}
\mathbb{C}\left(\xi_1 + \xi_2, \zeta_1 + \zeta_2\right) &= \mathbb{C}\left(a_0 + \sum_{i=1}^{2} a_i\xi_i, b_0 + \sum_{j=1}^{2} b_j\zeta_j\right) \\
&= \sum_{i=1}^{2}\sum_{j=1}^{2} a_ib_j\mathbb{C}\left(\xi_i, \zeta_j\right) \\
&= \sum_{i=1}^{2}\sum_{j=1}^{2} \mathbb{C}\left(\xi_i, \zeta_j\right) \\
&= \mathbb{C}\left(\xi_1, \zeta_1\right) + \mathbb{C}\left(\xi_1, \zeta_2\right) + \mathbb{C}\left(\xi_2, \zeta_1\right) + \mathbb{C}\left(\xi_2, \zeta_2\right).
\end{aligned} \tag{4.59}
$$

□

Folgendes Theorem besagt nun, wie man im Allgemeinen die Varianz einer Linearkombinationen von Zufallsvariablen unter Addition einer Konstante bestimmt werden kann.

Theorem 4.20 (Varianz einer Linearkombination von Zufallsvariablen)**.** *Gegeben seien n Zufallsvariablen $\xi_1, ..., \xi_n$ und $n+1$ reelle Konstanten $a_0, a_1, ..., a_n$. Dann gilt*

$$\mathbb{V}\left(a_0 + \sum_{i=1}^{n} a_i \xi_i\right) = \sum_{i=1}^{n} a_i^2 \mathbb{V}(\xi_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_i a_j \mathbb{C}(\xi_i, \xi_j). \tag{4.60}$$

∘

Beweis. Mit $\mathbb{V}(\xi) = \mathbb{C}(\xi, \xi)$ und Theorem 4.18 gilt zunächst

$$\begin{aligned}
\mathbb{V}\left(a_0 + \sum_{i=1}^{n} a_i \xi_i\right) &= \mathbb{C}\left(a_0 + \sum_{i=1}^{n} a_i \xi_i, a_0 + \sum_{i=1}^{n} a_i \xi_i\right) \\
&= \mathbb{C}\left(a_0 + \sum_{i=1}^{n} \xi_i, a_0 + \sum_{j=1}^{n} a_j \xi_j\right) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \mathbb{C}(\xi_i, \xi_j) \\
&= \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j=i}}^{n} a_i a_j \mathbb{C}(\xi_i, \xi_j) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} a_i a_j \mathbb{C}(\xi_i, \xi_j) \\
&= \sum_{i=1}^{n} a_i a_i \mathbb{C}(\xi_i, \xi_i) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} a_i a_j \mathbb{C}(\xi_i, \xi_j) \\
&= \sum_{i=1}^{n} a_i^2 \mathbb{V}(\xi_i) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} a_i a_j \mathbb{C}(\xi_i, \xi_j) \\
&= \sum_{i=1}^{n} a_i^2 \mathbb{V}(\xi_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_i a_j \mathbb{C}(\xi_i, \xi_j),
\end{aligned} \tag{4.61}$$

Dabei wurde in der vierten Gleichung die Doppelsumme in solche Terme aufgespalten für die $i = j$ und für die $i \neq j$ und in siebten Gleichung ausgenutzt, dass

$$a_i a_j \mathbb{C}(\xi_i, \xi_j) = a_j a_i \mathbb{C}(\xi_j, \xi_i). \tag{4.62}$$

Wir verdeutlichen die darausfolgende Identität

$$\sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} a_i a_j \mathbb{C}(\xi_i, \xi_j) = 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_i a_j \mathbb{C}(\xi_i, \xi_j), \tag{4.63}$$

am Beispiel $n = 3$ untenstehend. Analog mag man sich vorstellen, über alle Elemente außer der Diagonalelemente der $i = 1, ..., n$ Zeilen und $j = 1, ..., n$ Spalten einer symmetrischen Matrix mit Einträgen $a_i a_j \mathbb{C}(\xi_i, \xi_j)$ zu summieren. Die linke Seite oberer Gleichung entspricht dann dem zeilenweisen Vorgehen der Summationsbildung für alle Spalteneinträge außer dem, der in der Spalte der jeweils betrachteten Zeile steht, d.h. dem Diagonaleintrag. Die rechte Seite der obigen Gleichung entspricht dann dem Vorgehen, die Symmetrie der Matrix auszunutzen, also die Tatsache zu berücksichtigen, dass die Einträge der Matrix rechts oberhalb und links unterhalb der Diagonalen identisch sind, so dass es genügt, die Einträge der oberen linke Hälfte aufzuaddieren und zu verdoppeln. Dabei werden für jede Zeile $i = 1, ..., n$ nur gerade die Spalten $j = i+1, ..., n$ betrachtet, die rechts von der Diagonale stehen und ihre Einträge aufsummiert. In der letzten Zeile steht dabei nur ein Diagonalelement, das nicht zur Summe gehört, weshalb der Index

der äußeren Summe nur bis $n-1$ läuft. Konkret ergibt sich für $n := 3$

$$
\begin{aligned}
&\sum_{i=1}^{3}\sum_{\substack{j=1\\j\neq i}}^{3} a_i a_j \mathbb{C}(\xi_i, \xi_j) \\
&= \sum_{\substack{j=1\\j\neq 1}}^{3} a_1 a_j \mathbb{C}(\xi_1, \xi_j) + \sum_{\substack{j=1\\j\neq 2}}^{3} a_2 a_j \mathbb{C}(\xi_2, \xi_j) + \sum_{\substack{j=1\\j\neq 3}}^{3} a_3 a_j \mathbb{C}(\xi_3, \xi_j) \\
&= a_1 a_2 \mathbb{C}(\xi_1, \xi_2) + a_1 a_3 \mathbb{C}(\xi_1, \xi_3) + a_2 a_1 \mathbb{C}(\xi_2, \xi_1) + a_2 a_3 \mathbb{C}(\xi_2, \xi_3) + a_3 a_1 \mathbb{C}(\xi_3, \xi_1) + a_3 a_2 \mathbb{C}(\xi_3, \xi_2) \\
&= a_1 a_2 \mathbb{C}(\xi_1, \xi_2) + a_2 a_1 \mathbb{C}(\xi_2, \xi_1) + a_1 a_3 \mathbb{C}(\xi_1, \xi_3) + a_3 1 a_1 \mathbb{C}(\xi_3, \xi_1) + a_2 a_3 \mathbb{C}(\xi_2, \xi_3) + a_3 a_2 \mathbb{C}(\xi_3, \xi_2) \\
&= 2\mathbb{C}(\xi_1, \xi_2) + 2\mathbb{C}(\xi_1, \xi_3) + 2\mathbb{C}(\xi_2, \xi_3) \\
&= 2a_1 a_2 \left(\mathbb{C}(\xi_1, \xi_2) + a_1 a_3 \mathbb{C}(\xi_1, \xi_3) + a_2 a_3 \mathbb{C}(\xi_2, \xi_3)\right) \\
&= 2\left(\sum_{j=2}^{3} a_1 a_j \mathbb{C}(\xi_1, \xi_j) + \sum_{j=3}^{3} a_2 a_j \mathbb{C}(\xi_2, \xi_j)\right) \\
&= 2\left(\sum_{j=1+1}^{3} a_1 a_j \mathbb{C}(\xi_1, \xi_j) + \sum_{j=2+1}^{3} a_2 a_j \mathbb{C}(\xi_2, \xi_j)\right) \\
&= 2\sum_{i=1}^{2}\sum_{j=i+1}^{3} a_i a_j \mathbb{C}(\xi_i, \xi_j).
\end{aligned}
\tag{4.64}
$$

□

Theorem 4.21 (Varianzen spezieller Linearkombinationen von Zufallsvariablen)**.**

(1) (Varianz bei Addition zweier Zufallsvariablen). Gegeben seien zwei Zufallsvariablen ξ und ζ. Dann gilt

$$\mathbb{V}(\xi + \zeta) = \mathbb{V}(\xi) + \mathbb{V}(\zeta) + 2\mathbb{C}(\xi, \zeta) \tag{4.65}$$

(2) (Varianz bei Subtraktion zweier Zufallsvariablen). Gegeben seien zwei Zufallsvariablen ξ und ζ. Dann gilt

$$\mathbb{V}(\xi - \zeta) = \mathbb{V}(\xi) + \mathbb{V}(\zeta) - 2\mathbb{C}(\xi, \zeta) \tag{4.66}$$

∘

Beweis. (1) Es seien $n := 2$, $\xi_1 := \xi$, $\xi_2 := \zeta$, $a_0 := 0$, $a_1 := 1$ und $a_2 := 1$. Dann gilt mit dem Theorem zur Varianz einer Linearkombination von Zufallsvariablen

$$
\begin{aligned}
\mathbb{V}(\xi + \zeta) &= \mathbb{V}(a_0 + a_1\xi_1 + a_2\xi_2) \\
&= \mathbb{V}\left(a_0 + \sum_{i=1}^{2} a_i \xi_i\right) \\
&= \sum_{i=1}^{2} a_i^2 \mathbb{V}(\xi_i) + 2\sum_{i=1}^{2-1}\sum_{j=i+1}^{2} a_i a_j \mathbb{C}(\xi_i, \xi_j) \\
&= \sum_{i=1}^{2} a_i^2 \mathbb{V}(\xi_i) + 2\sum_{i=1}^{1}\sum_{j=i+1}^{2} a_i a_j \mathbb{C}(\xi_i, \xi_j) \\
&= \sum_{i=1}^{2} a_i^2 \mathbb{V}(\xi_i) + 2\sum_{j=1+1}^{2} a_i a_j \mathbb{C}(\xi_1, \xi_j) \\
&= \sum_{i=1}^{2} a_i^2 \mathbb{V}(\xi_i) + 2\sum_{j=2}^{2} a_1 a_j \mathbb{C}(\xi_1, \xi_j) \\
&= a_1^2 \mathbb{V}(\xi_1) + a_2^2 \mathbb{V}(\xi_2) + 2a_1 a_2 \mathbb{C}(\xi_1, \xi_2) \\
&= 1^2 \cdot \mathbb{V}(\xi) + 1^2 \cdot \mathbb{V}(\zeta) + 2 \cdot 1 \cdot 1 \cdot \mathbb{C}(\xi, \zeta) \\
&= \mathbb{V}(\xi) + \mathbb{V}(\zeta) + 2\mathbb{C}(\xi, \zeta).
\end{aligned}
\tag{4.67}
$$

(2) Es seien $n \coloneqq 2$, $\xi_1 \coloneqq \xi$, $\xi_2 \coloneqq \zeta$, $a_0 \coloneqq 0$, $a_1 \coloneqq 1$ und $a_2 \coloneqq -1$. Dann gilt mit dem Theorem zur Varianz einer Linearkombination von Zufallsvariablen

$$\begin{aligned}
\mathbb{V}(\xi - \zeta) &= \mathbb{V}(a_0 + a_1\xi_1 + a_2\xi_2) \\
&= \mathbb{V}\left(a_0 + \sum_{i=1}^{2} a_i\xi_i\right) \\
&= \sum_{i=1}^{2} a_i^2\mathbb{V}(\xi_i) + 2\sum_{i=1}^{2-1}\sum_{j=i+1}^{2} a_i a_j\mathbb{C}(\xi_i, \xi_j) \\
&= \sum_{i=1}^{2} a_i^2\mathbb{V}(\xi_i) + 2\sum_{i=1}^{1}\sum_{j=i+1}^{2} a_i a_j\mathbb{C}(\xi_i, \xi_j) \\
&= \sum_{i=1}^{2} a_i^2\mathbb{V}(\xi_i) + 2\sum_{j=1+1}^{2} a_i a_j\mathbb{C}(\xi_1, \xi_j) \\
&= \sum_{i=1}^{2} a_i^2\mathbb{V}(\xi_i) + 2\sum_{j=2}^{2} a_1 a_j\mathbb{C}(\xi_1, \xi_j) \\
&= a_1^2\mathbb{V}(\xi_1) + a_2^2\mathbb{V}(\xi_2) + 2a_1a_2\mathbb{C}(\xi_1, \xi_2) \\
&= 1^2 \cdot \mathbb{V}(\xi) + (-1)^2 \cdot \mathbb{V}(\zeta) + 2 \cdot 1 \cdot (-1) \cdot \mathbb{C}(\xi, \zeta) \\
&= \mathbb{V}(\xi) + \mathbb{V}(\zeta) - 2\mathbb{C}(\xi, \zeta).
\end{aligned} \tag{4.68}$$

□

Wie den Erwartungswert und die Varianz kann man auch die Kovarianz in einer bedingten gemeinsamen Verteilung zweier Zufallsvariablen betrachten. Dies führt auf den Begriff der *bedingten Kovarianz*.

Definition 4.8 (Bedingte Kovarianz)**.** Gegeben sei ein Zufallsvektor $\xi \coloneqq (\xi_1, \xi_2, \xi_3)$ mit Ergebnisraum $\mathcal{X} \coloneqq \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$ und WMF oder WDF $p(x_1, x_2, x_3)$ und bedingter WMF oder WDF $p(x_1, x_2|x_3)$ für alle $x_3 \in \mathcal{X}_3$. Dann ist die bedingte Kovarianz von ξ_1 und ξ_2 gegeben $\xi_3 = x_3$ definiert als

$$\mathbb{C}(\xi_1, \xi_2|\xi_3 = x_3) = \mathbb{E}\left((\xi_1 - \mathbb{E}\left(\xi_1|\xi_3 = x_3\right))\left(\xi_2 - \mathbb{E}\left(\xi_2|\xi_3 = x_3\right)\right)|\xi_3 = x_3\right) \tag{4.69}$$

•

Die bedingte Kovarianz ist also im Sinne des bedingten Erwartungswerts des Zufallsvektors (ξ_1, ξ_2) definiert und ist, wie der bedingte Erwartungswert und die bedingte Varianz im Allgemeinen eine Zufallsvariable dar. Schließlich halten wir fest, dass auch für die bedingte Kovarianz der Kovarianzverschiebungssatz gilt.

Theorem 4.22 (Verschiebungssatz der bedingten Kovarianz)**.** *Gegeben sei ein Zufallsvektor $\xi \coloneqq (\xi_1, \xi_2, \xi_3)$. Dann gilt*

$$\mathbb{C}(\xi_1, \xi_2|\xi_3) = \mathbb{E}\left(\xi_1\xi_2|\xi_3\right) - \mathbb{E}\left(\xi_1|\xi_3\right)\mathbb{E}\left(\xi_2|\xi_3\right). \tag{4.70}$$

∘

Beweis. Ein Beweis ergibt sich durch die Ersetzung der entsprechenden Erwartungswerte im Beweis von Theorem 4.16 durch die entsprechenden bedingten Erwartungswerte. □

Das multivariate Analogon der Varianz einer Zufallsvariable ist die *Kovarianzmatrix eines Zufallsvektors*. Diese enkodiert neben den Varianzen der Komponenten des Zufallsvektors auch ihre paarweisen Kovarianzen und ist wie folgt definiert.

Definition 4.9 (Kovarianzmatrix eines Zufallsvektors)**.** ξ sei ein n-dimensionaler Zufallvektor. Dann ist die *Kovarianzmatrix* von ξ definiert als die $n \times n$ Matrix

$$\mathbb{C}(\xi) := \mathbb{E}\left((\xi - \mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))^T\right). \tag{4.71}$$

•

Die Kovarianzmatrix ist in Definition 4.9 formal analog zur Kovarianz zweier Zufallsvariablen definiert. Eine direkte Rückführung des Begriffs der Kovarianzmatrix eines Zufallsvektors auf den Begriff aus dem univariaten Kontext bekannten Begriff der Kovarianz zweier Zufallsvariablen erlaubt folgendesTheorem.

Theorem 4.23 (Eigenschaften der Kovarianzmatrix)**.** *ξ sei ein m-dimensionaler Zufallsvektor und $\mathbb{C}(\xi)$ sei seine Kovarianzmatrix. Dann gelten*

(1) (Elemente) Die Elemente von $\mathbb{C}(\xi)$ sind die Kovarianzen der Komponenten von ξ,

$$\mathbb{C}(\xi) = \left(\mathbb{C}(\xi_i, \xi_j)\right)_{1 \leq i,j \leq m}. \tag{4.72}$$

(2) (Kovarianzmatrixverschiebungssatz) Es gilt

$$\mathbb{C}(\xi) = \mathbb{E}\left(\xi\xi^T\right) - \mathbb{E}(\xi)\mathbb{E}(\xi)^T. \tag{4.73}$$

(3) (Linear-affine Transformation) Für $A \in \mathbb{R}^{n \times m}$ und $b \in \mathbb{R}^n$ gilt

$$\mathbb{C}(A\xi + b) = A\mathbb{C}(\xi)A^T. \tag{4.74}$$

(4) (Matrixeigenschaften) $\mathbb{C}(\xi)$ ist symmetrisch und positiv-semidefinit.

◦

Beweis. (1) Es gilt

$$\begin{aligned}
\mathbb{C}(\xi) &:= \mathbb{E}\left((\xi - \mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))^T\right) \\
&= \mathbb{E}\left(\left(\begin{pmatrix}\xi_1\\ \vdots \\ \xi_n\end{pmatrix} - \begin{pmatrix}\mathbb{E}(\xi_1)\\ \vdots \\ \mathbb{E}(\xi_n)\end{pmatrix}\right)\left(\begin{pmatrix}\xi_1\\ \vdots \\ \xi_n\end{pmatrix} - \begin{pmatrix}\mathbb{E}(\xi_1)\\ \vdots \\ \mathbb{E}(\xi_n)\end{pmatrix}\right)^T\right) \\
&= \mathbb{E}\left(\begin{pmatrix}\xi_1 - \mathbb{E}(\xi_1)\\ \vdots \\ \xi_n - \mathbb{E}(\xi_n)\end{pmatrix}\begin{pmatrix}\xi_1 - \mathbb{E}(\xi_1)\\ \vdots \\ \xi_n - \mathbb{E}(\xi_n)\end{pmatrix}^T\right) \\
&= \mathbb{E}\left(\begin{pmatrix}\xi_1 - \mathbb{E}(\xi_1)\\ \vdots \\ \xi_n - \mathbb{E}(\xi_n)\end{pmatrix}\begin{pmatrix}\xi_1 - \mathbb{E}(\xi_1) & \dots & \xi_n - \mathbb{E}(\xi_n)\end{pmatrix}\right) \\
&= \mathbb{E}\begin{pmatrix}(\xi_1 - \mathbb{E}(\xi_1))(\xi_1 - \mathbb{E}(\xi_1)) & \dots & (\xi_1 - \mathbb{E}(\xi_1))(\xi_n - \mathbb{E}(\xi_n)\\ \vdots & \ddots & \vdots \\ (\xi_n - \mathbb{E}(\xi_n))(\xi_1 - \mathbb{E}(\xi_1)) & \dots & (\xi_n - \mathbb{E}(\xi_n))(\xi_n - \mathbb{E}(\xi_n))\end{pmatrix} \\
&= \left(\mathbb{E}\left((\xi_i - \mathbb{E}(\xi_i))(\xi_j - \mathbb{E}(\xi_j))\right)\right)_{1 \leq i,j \leq n} \\
&= \left(\mathbb{C}(\xi_i, \xi_j)\right)_{1 \leq i,j \leq n}.
\end{aligned} \tag{4.75}$$

(2) Mit den Eigenschaften von Erwartungswerten gilt

$$\begin{aligned}\mathbb{C}(\xi) &= \mathbb{E}\left((\xi - \mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))^T\right) \\ &= \mathbb{E}\left(\xi\xi^T - \xi\mathbb{E}(\xi)^T - \mathbb{E}(\xi)\xi^T + \mathbb{E}(\xi)\mathbb{E}(\xi)^T\right) \\ &= \mathbb{E}\left(\xi\xi^T\right) - \mathbb{E}(\xi)\mathbb{E}(\xi)^T - \mathbb{E}(\xi)\mathbb{E}(\xi)^T + \mathbb{E}(\xi)\mathbb{E}(\xi)^T \\ &= \mathbb{E}\left(\xi\xi^T\right) - \mathbb{E}(\xi)\mathbb{E}(\xi)^T.\end{aligned} \tag{4.76}$$

(3) Mit den Eigenschaften von Erwartungswerten gilt

$$\begin{aligned}\mathbb{C}(A\xi + b) &= \mathbb{E}\left((A\xi + b - \mathbb{E}(A\xi + b))(A\xi + b - \mathbb{E}(A\xi + b))^T\right) \\ &= \mathbb{E}\left((A\xi + b - A\mathbb{E}(\xi) - b)(A\xi + b - A\mathbb{E}(\xi) - b)^T\right) \\ &= \mathbb{E}\left((A(\xi - \mathbb{E}(\xi)))(A(\xi - \mathbb{E}(\xi)))^T\right) \\ &= \mathbb{E}\left(A(\xi - \mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))^T A^T\right) \\ &= A\mathbb{E}\left((\xi - \mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))^T\right) A^T \\ &= A\mathbb{C}(\xi)A^T.\end{aligned} \tag{4.77}$$

(4) Die Symmetrie von $\mathbb{C}(\xi)$ folgt aus der Symmetrie der Kovarianz einer Zufallsvariable mit

$$\mathbb{C}(\xi_i, \xi_j) = \mathbb{C}(\xi_j, \xi_i) \text{ für alle } i = 1, ..., m, j = 1, ..., m. \tag{4.78}$$

Um die positive Semidefinitheit von $\mathbb{C}(\xi)$ nachzuweisen, ist zu zeigen, dass $a^T\mathbb{C}(\xi)a \geq 0$ für alle $a \in \mathbb{R}^m$ mit $a \neq 0_m$. Sei also $a \in \mathbb{R}^m$ mit $a \neq 0$. Dann gilt mit Aussage (3) für $A := a^T \in \mathbb{R}^{1\times m}$, dass

$$a^T\mathbb{C}(\xi)a = \mathbb{C}(a^T\xi). \tag{4.79}$$

Weiterhin gilt mit der Definition der Kovarianzmatrix aber, dass

$$\mathbb{C}(a^T\xi) = \mathbb{E}\left(\left(a^T\xi - \mathbb{E}(a^T\xi)\right)^2\right) = \mathbb{V}\left(a^T\xi\right). \tag{4.80}$$

Da mit den Eigenschaften der Varianz die Varianz der Zufallsvariable $a^T\xi$ aber immer nichtnegativ ist, folgt

$$a^T\mathbb{C}(\xi)a = \mathbb{V}(a^T\xi) \geq 0 \tag{4.81}$$

und damit die positive Semidefinitheit von $\mathbb{C}(\xi)$. □

Die Diagonalelemente von $\mathbb{C}(\xi)$ sind die Varianzen der Komponenten von ξ, da

$$\mathbb{V}(\xi_i) = \mathbb{C}(\xi_i, \xi_i) \text{ für } i = 1, ...m. \tag{4.82}$$

Eigenschaften (2) und (3) sind im Wesentlichen analog zu den Eigenschaften der Varianz.

Die Kovarianzmatrix eines Zufallsvektors ξ ist also die Matrix der Kovarianzen der Komponenten von ξ. Damit ist auch die Kovarianzmatrix direkt im Sinne des Begriffs der Kovarianz von Zufallsvektoren gegeben. Da die Kovarianz einer Zufallsvariable mit sich selbst bekanntlich ihre Varianz ist, enthält die Kovarianzmatrix auf ihrer Diagonalen die Varianzen der Komponenten von ξ.

Folgendes Theorem dokumentiert eine Schreibweise für die Kovarianzmatrix eines partitionierten Zufallsvektors im Sinne von Erwartungswerten von Zufallvektorprodukten an, die zum Beispiel im Rahmen der Kanonischen Korrelationsanalyse hilfreich ist.

Theorem 4.24 (Kovarianzmatrizen von Zufallsvektoren)**.** *Es seien*

$$\zeta = \begin{pmatrix}\xi \\ \upsilon\end{pmatrix} \text{ mit } \mathbb{E}(\zeta) := 0_m \tag{4.83}$$

ein $m_\xi + m_\upsilon$-dimensionaler Zufallsvektor und sein Erwartungswertvektor, respektive. Dann kann die $m \times m$ Kovarianzmatrix von ζ geschrieben werden als

$$\mathbb{C}(\zeta) = \begin{pmatrix} \Sigma_{\xi\xi} & \Sigma_{\xi_1\xi_2} \\ \Sigma_{\upsilon\xi} & \Sigma_{\upsilon\upsilon} \end{pmatrix} \in \mathbb{R}^{m \times m} \tag{4.84}$$

wobei

$$\begin{aligned} \Sigma_{\xi\xi} &:= \mathbb{E}\left(\xi\xi^T\right) \in \mathbb{R}^{m_\xi \times m_\xi} \\ \Sigma_{\xi_1\xi_2} &:= \mathbb{E}\left(\xi_1\xi_2^T\right) \in \mathbb{R}^{m_\xi \times m_\upsilon} \\ \Sigma_{\upsilon\xi} &:= \mathbb{E}\left(\upsilon\xi^T\right) \in \mathbb{R}^{m_\upsilon \times m_\xi} \\ \Sigma_{\upsilon\upsilon} &:= \mathbb{E}\left(\upsilon\upsilon^T\right) \in \mathbb{R}^{m_\xi \times m_\upsilon} \end{aligned} \tag{4.85}$$

∘

Beweis. Nach Definition der Kovarianzmatrix eines Zufallsvektors gilt

$$\begin{aligned} \mathbb{C}(z) &= \mathbb{E}\left((\zeta - \mathbb{E}(\zeta))(\zeta - \mathbb{E}(\zeta))^T\right) \\ &= \mathbb{E}\left((\zeta - 0_m)(\zeta - 0_m)^T\right) \\ &= \mathbb{E}\left(\zeta\zeta^T\right) \\ &= \mathbb{E}\left(\begin{pmatrix} \xi \\ \upsilon \end{pmatrix} \begin{pmatrix} \xi^T & \upsilon^T \end{pmatrix}\right) \\ &= \mathbb{E}\left(\begin{pmatrix} \xi\xi^T & \xi_1\xi_2^T \\ \upsilon\xi^T & \upsilon\upsilon^T \end{pmatrix}\right) \\ &= \begin{pmatrix} \mathbb{E}\left(\xi\xi^T\right) & \mathbb{E}\left(\xi_1\xi_2^T\right) \\ \mathbb{E}\left(\upsilon\xi^T\right) & \mathbb{E}\left(\upsilon\upsilon^T\right) \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_{\xi\xi} & \Sigma_{\xi_1\xi_2} \\ \Sigma_{\upsilon\xi} & \Sigma_{\upsilon\upsilon} \end{pmatrix} \end{aligned} \tag{4.86}$$

□

Schließlich ist man in manchen Anwendungen an einer normalisierten, maßstabsunabhängigen Repräsentation der Kovarianzen eines Zufallsvektors interessiert. Wie im univariaten Fall bietet sich hierfür die Normalisierung der Kovarianz zweier Zufallsvariablen mithilfe ihrer jeweiligen Varianzen im Sinne einer Korrelation an. Diese Überlegung führt auf den Begriff der *Korrelationsmatrix* eines Zufallsvektors.

Definition 4.10 (Korrelationsmatrix)**.** ξ sei ein n-dimensionaler Zufallsvektor. Dann ist die *Korrelationsmatrix* von ξ definiert als die $n \times n$ Matrix

$$\mathbb{R}(\xi) := \left(\rho_{ij}\right)_{1 \leq i,j \leq n} = \left(\frac{\mathbb{C}(\xi_i, \xi_j)}{\sqrt{\mathbb{V}(\xi_i)}\sqrt{\mathbb{V}(\xi_j)}} \right)_{1 \leq i,j \leq n}. \tag{4.87}$$

•

Da es sich bei den Varianzen der Komponenten von ξ um die Diagonalelement der Kovarianzmatrix von ξ handelt, ist die Korrelationsmatrix natürlich in der Kovarianzmatrix implizit. Weiterhin gelten, wie immer für Korrelationen, für die Einträge $\rho_{ij}, 1 \leq i, j \leq n$ der Korrelationsmatrix, dass

$$\rho_{ij} \in [-1, 1] \text{ für } 1 \leq i, j \in n \text{ und } \rho_{ii} = 1 \text{ für } 1 \leq i \leq n. \tag{4.88}$$

Die Begriffe des Stichprobenmittels, der Stichprobenvarianz und der Stichprobenkovarianz lassen sich auch auf den Fall multivariater Stichproben übertragen. Wir nutzen folgende Definition.

Definition 4.11 (Stichprobenmittel, -kovarianmatrix und -korrelationsmatrix)**.** $\upsilon_1, ..., \upsilon_n$ sei eine Menge von m-dimensionalen Zufallsvektoren, genannt *Stichprobe*.

- Das *Stichprobenmittel* der $\upsilon_1, ..., \upsilon_n$ ist definiert als der m-dimensionale Vektor

$$\bar{\upsilon} := \frac{1}{n} \sum_{i=1}^{n} \upsilon_i. \tag{4.89}$$

- Die *Stichprobenkovarianzmatrix* der $\upsilon_1, ..., \upsilon_n$ ist definiert als die $m \times m$ Matrix

$$C := \frac{1}{n-1} \sum_{i=1}^{n} (\upsilon_i - \bar{\upsilon})(\upsilon_i - \bar{\upsilon})^T. \tag{4.90}$$

- Die *Stichprobenkorrelationsmatrix* der $\upsilon_1, ..., \upsilon_n$ ist definiert als die $m \times m$ Matrix

$$D := \left(\frac{(C)_{ij}}{\sqrt{(C)_{ii}}\sqrt{(C)_{jj}}} \right)_{1 \leq i,j \leq m}. \tag{4.91}$$

•

Zur konkreten Berechnung von Stichprobenmittel, Stichprobenkovarianzmatrix und Stichprobenkorrrelationsmatrix basierend auf einem multivariaten Datensatz bieten sich die Aussagen des folgenden Theorems an.

Theorem 4.25 (Datenmatrix und Stichprobenstatistiken)**.**

Es sei

$$\upsilon := \begin{pmatrix} \upsilon_1 & \cdots & \upsilon_n \end{pmatrix} \tag{4.92}$$

eine $m \times n$ Datenmatrix, die durch die spaltenweise Konkatenation von n m-dimensionalen Zufallvektoren $\upsilon_1, ..., \upsilon_n$ gegeben sei. Dann ergeben sich

- *für das Stichprobenmittel*

$$\bar{\upsilon} = \frac{1}{n} \upsilon 1_n, \tag{4.93}$$

- *für die Stichprobenkovarianzmatrix*

$$C = \frac{1}{n-1} \left(\upsilon \left(I_n - \frac{1}{n} 1_{nn} \right) \upsilon^T \right), \tag{4.94}$$

- *und für Stichprobenkorrelationsmatrix mit*

$$D := diag\left(\sqrt{C_{y_{ii}}}^{-1}, i = 1, ..., m \right), \tag{4.95}$$

dass

$$R = DCD. \tag{4.96}$$

∘

Beweis. Die Darstellung des Stichprobenmittels ergibt sich aus

$$\begin{aligned}\bar{\upsilon} &:= \frac{1}{n}\sum_{i=1}^{n}\upsilon_i \\ &= \frac{1}{n}\begin{pmatrix}\sum_{i=1}^{n}\upsilon_{i1}\\ \vdots \\ \sum_{i=1}^{n}\upsilon_{im}\end{pmatrix} \\ &= \frac{1}{n}\left(\begin{pmatrix}\upsilon_{11} & \cdots & \upsilon_{n1}\\ \vdots & \ddots & \vdots \\ \upsilon_{1m} & \cdots & \upsilon_{nm}\end{pmatrix}\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}\right) \\ &= \frac{1}{n}\upsilon 1_n.\end{aligned} \tag{4.97}$$

Hinsichtlich der Darstellung der Stichprobenkovarianzmatrix halten wir zunächst fest, dass nach Definition gilt, dass

$$\begin{aligned}C &:= \frac{1}{n-1}\sum_{i=1}^{n}(\upsilon_i-\bar{\upsilon})(\upsilon_i-\bar{\upsilon})^T \\ &= \frac{1}{n-1}\sum_{i=1}^{n}(\upsilon_i\upsilon_i^T-\upsilon_i\bar{\upsilon}^T-\bar{\upsilon}\upsilon_i^T+\bar{\upsilon}\bar{\upsilon}^T) \\ &= \frac{1}{n-1}\left(\sum_{i=1}^{n}\upsilon_i\upsilon_i^T-\sum_{i=1}^{n}\upsilon_i\bar{\upsilon}^T-\sum_{i=1}^{n}\bar{\upsilon}\upsilon_i^T+\sum_{i=1}^{n}\bar{\upsilon}\bar{\upsilon}^T\right) \\ &= \frac{1}{n-1}\left(\sum_{i=1}^{n}\upsilon_i\upsilon_i^T-n\bar{\upsilon}\bar{\upsilon}^T-n\bar{\upsilon}\bar{\upsilon}^T+n\bar{\upsilon}\bar{\upsilon}^T\right) \\ &= \frac{1}{n-1}\left(\sum_{i=1}^{n}\upsilon_i\upsilon_i^T-n\bar{\upsilon}\bar{\upsilon}^T\right).\end{aligned} \tag{4.98}$$

Mit $1_n 1_n^T = 1_{nn}$ ergibt sich dann weiterhin

$$\begin{aligned}\upsilon\left(I_n-\frac{1}{n}1_{nn}\right)\upsilon^T &= \left(\upsilon I_n-\frac{1}{n}\upsilon 1_{nn}\right)\upsilon^T \\ &= \upsilon\upsilon^T-\frac{1}{n}\upsilon 1_{nn}\upsilon^T \\ &= \begin{pmatrix}\upsilon_1 & \cdots & \upsilon_n\end{pmatrix}\begin{pmatrix}\upsilon_1^T\\ \vdots \\ \upsilon_n^T\end{pmatrix}-\frac{1}{n}\upsilon 1_n 1_n^T\upsilon^T \\ &= \sum_{i=1}^{n}\upsilon_i\upsilon_i^T-n\left(\frac{1}{n}\upsilon 1_n\right)\left(\frac{1}{n}1_n^T\upsilon^T\right) \\ &= \sum_{i=1}^{n}\upsilon_i\upsilon_i^T-n\bar{\upsilon}\bar{\upsilon}^T \\ &= \frac{1}{n-1}\sum_{i=1}^{n}(\upsilon_i-\bar{\upsilon})(\upsilon_i-\bar{\upsilon})^T \\ &= C.\end{aligned} \tag{4.99}$$

Hinsichtlich der Korrelationsmatrix ergibt sich nach Definition und für ein beliebiges Indexpaar i, j mit $1 \leq i, j \leq m$ schließlich, dass

$$\begin{aligned}R_{y_{ij}} &= \frac{(C)_{ij}}{\sqrt{(C)_{ii}}\sqrt{(C)_{jj}}} \\ &= \frac{1}{\sqrt{(C)_{ii}}}(C)_{ij}\frac{1}{\sqrt{(C)_{jj}}} \\ &= (DCD)_{ij}.\end{aligned} \tag{4.100}$$

□

4.4 Normalverteilungen

In diesem Abschnitt stellen wir einige grundlegende Definitionen und Theoreme zu multivariaten Normalverteilungen zusammen. Wir nutzen folgende grundlegende Definition.

Definition 4.12. ξ sei ein n-dimensionaler Zufallsvektor mit Ergebnisraum $\mathbb{R}^n$ und WDF

$$p : \mathbb{R}^n \to \mathbb{R}_{>0},\, x \mapsto p(x) := (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right). \tag{4.101}$$

Dann sagen wird, dass ξ einer *multivariaten (oder n-dimensionalen) Normalverteilung* mit *Erwartungswertparameter* $\mu \in \mathbb{R}^n$ und positiv-definitem *Kovarianzmatrixparameter* $\Sigma \in \mathbb{R}^{n\times n}$ unterliegt und nennen ξ einen *(multivariat) normalverteilten Zufallsvektor*. Wir kürzen dies mit $\xi \sim N(\mu, \Sigma)$ ab. Die WDF eines multivariat normalverteilten Zufallsvektors bezeichnen wir mit

$$N(x; \mu, \Sigma) := (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right). \tag{4.102}$$

•

Marginale, bedingte und gemeinsame Normalverteilungen

Multivariate Normalverteilungen haben die Eigenschaft, dass auch alle anderen assoziierten Verteilungen Normalverteilungen sind und deren Erwartungswert- und Kovarianzmatrixparameter aus den Parametern der jeweils komplementären Verteilung errechnet werden können. Insbesondere gilt zum einen, dass die uni- und multivariaten Marginalverteilungen multivariater Normalverteilungen wiederum Normalverteilungen sind. Zum anderen lassen sich wie alle multivariaten Verteilungen multivariate Normalverteilungen multiplikativ in eine marginale und eine bedingte Verteilung zerlegen. Insbesondere sind nun aber bei multivariaten Normalverteilungen diese Verteilungen wiederum (multivariate) Normalverteilungen, deren Parameter aus den Parametern der gemeinsame Verteilung errechnet werden können und umgekehrt. Wir fassen obige Erkenntnisse formal in den folgenden drei Theoremen zusammen.

Theorem 4.26 (Marginale Normalverteilungen)**.** *Es sei $n := k + l$ und $\xi = (\xi_1, ..., \xi_n)^T$ sei ein n-dimensionaler normalverteilter Zufallsvektor mit Erwartungswertparameter*

$$\mu = \begin{pmatrix} \mu_\upsilon \\ \mu_\zeta \end{pmatrix} \in \mathbb{R}^n, \tag{4.103}$$

mit $\mu_\upsilon \in \mathbb{R}^k$ and $\mu_\zeta \in \mathbb{R}^l$ und Kovarianzmatrixparameter

$$\Sigma = \begin{pmatrix} \Sigma_{\upsilon\upsilon} & \Sigma_{\upsilon\zeta} \\ \Sigma_{\zeta\upsilon} & \Sigma_{\zeta\zeta} \end{pmatrix} \in \mathbb{R}^{n\times n}, \tag{4.104}$$

mit $\Sigma_{\upsilon\upsilon} \in \mathbb{R}^{k\times k}$, $\Sigma_{\upsilon\zeta} \in \mathbb{R}^{k\times l}$, $\Sigma_{\zeta\upsilon} \in \mathbb{R}^{l\times k}$, und $\Sigma_{\zeta\zeta} \in \mathbb{R}^{l\times l}$. Dann sind $\upsilon := (\xi_1, ..., \xi_k)^T$ und $\zeta := (\xi_{k+1}, ..., \xi_n)^T$ k- und l-dimensionale normalverteilte Zufallsvektoren, respektive, und es gilt

$$\upsilon \sim N(\mu_\upsilon, \Sigma_{\upsilon\upsilon}) \text{ und } \zeta \sim N(\mu_\zeta, \Sigma_{\zeta\zeta}). \tag{4.105}$$

∘

Die Marginalverteilungen einer multivariaten Normalverteilung sind also auch Normalverteilungen und die Parameter der Marginalverteilungen ergeben sich aus den Parametern der gemeinsamen Verteilung. Für Beweise dieses Theorems verweisen wir auf Mardia et al. (1979) und Anderson (2003). Abbildung 4.1 visualisiert Theorem 4.26 für den Fall $n := 2, k := 1, l := 1$,

$$\mu := \begin{pmatrix} 1 \\ 2 \end{pmatrix} \in \mathbb{R}^2 \text{ und } \Sigma := \begin{pmatrix} 0.10 & 0.08 \\ 0.08 & 0.15 \end{pmatrix} \in \mathbb{R}^{2\times 2}. \tag{4.106}$$

Abbildung 4.1 A zeigt dabei die WDF des bivariaten Zufallsvektors ξ und Abbildung 4.1 B und C die WDFen der entsprechenden marginalen Zufallsvariablen υ und ζ.

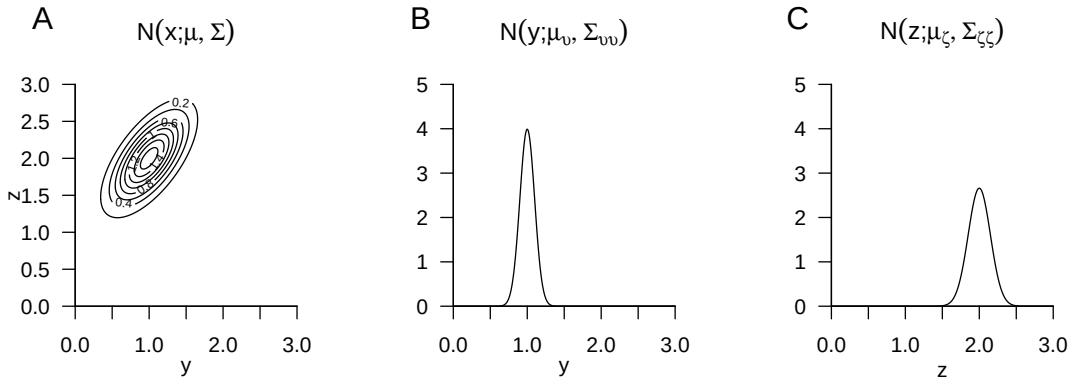

Abbildung 4.1 Marginale Verteilungen eines bivariaten normalverteilten Zufallsvektor.

Mithilfe einer marginalen und einer bedingten multivariaten Normalverteilung lässt sich eine gemeinsame multivariate Normalverteilung konstruieren, deren Parameter sich aus den Parametern der marginalen und bedingten Verteilung ergeben. Dies ist die zentrale Aussage folgenden Theorems.

Theorem 4.27 (Gemeinsame Normalverteilungen)**.** *ξ sei ein m-dimensionaler normalverteilter Zufallsvektor mit WDF*

$$p : \mathbb{R}^m \to \mathbb{R}_{>0},\ x \mapsto p(x) := N(x; \mu_\xi, \Sigma_{\xi\xi}) \text{ mit } \mu_\xi \in \mathbb{R}^m, \Sigma_{\xi\xi} \in \mathbb{R}^{m\times m}, \tag{4.107}$$

$A \in \mathbb{R}^{n\times m}$ sei eine Matrix, $b \in \mathbb{R}^n$ sei ein Vektor und υ sei ein n-dimensionaler bedingt normalverteilter Zufallsvektor mit bedingter WDF

$$p : \mathbb{R}^n \to \mathbb{R}_{>0},\ y \mapsto p(y|x) := N(y; A\xi + b, \Sigma_{\upsilon\upsilon}) \text{ mit } \Sigma_{\upsilon\upsilon} \in \mathbb{R}^{n\times n}. \tag{4.108}$$

Dann ist der $m+n$-dimensionale Zufallsvektor $(\xi, \upsilon)^T$ normalverteilt mit (gemeinsamer) WDF

$$p : \mathbb{R}^{m+n} \to \mathbb{R}_{>0},\ \begin{pmatrix} x \\ y \end{pmatrix} \mapsto p\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = N\left(\begin{pmatrix} x \\ y \end{pmatrix}; \mu_{\xi,\upsilon}, \Sigma_{\xi,\upsilon}\right), \tag{4.109}$$

mit $\mu_{\xi,\upsilon} \in \mathbb{R}^{m+n}$ und $\Sigma_{\xi,\upsilon} \in \mathbb{R}^{m+n\times m+n}$ und insbesondere

$$\mu_{\xi,\upsilon} = \begin{pmatrix} \mu_\xi \\ A\mu_\xi + b \end{pmatrix} \text{ und } \Sigma_{\xi,\upsilon} = \begin{pmatrix} \Sigma_{\xi\xi} & \Sigma_{\xi\xi}A^T \\ A\Sigma_{\xi\xi} & \Sigma_{\upsilon\upsilon} + A\Sigma_{\xi\xi}A^T \end{pmatrix}. \tag{4.110}$$

∘

Insbesondere ergeben sich die Parameter der gemeinsamen Verteilung also als linear-affine Transformation der Parameter der induzierenden Verteilungen. Abbildung 4.2 visualisiert Theorem 4.27 für den Fall $m := 1, n := 1, \mu_\xi := 1, \Sigma_{\xi\xi} := 0.2, A := 1, b := 1$ und $\Sigma_{\upsilon\upsilon} := 0.1$. Abbildung 4.2 A zeigt dabei die WDF der Zufallsvariable ξ, Abbildung 4.2 B zeigt die WDF der bedingten Verteilung der Zufallsvariable υ gegeben ξ und Abbildung 4.1 C schließlich zeigt die WDFen des induzierten bivariaten Zufallsvektors (ξ, υ).

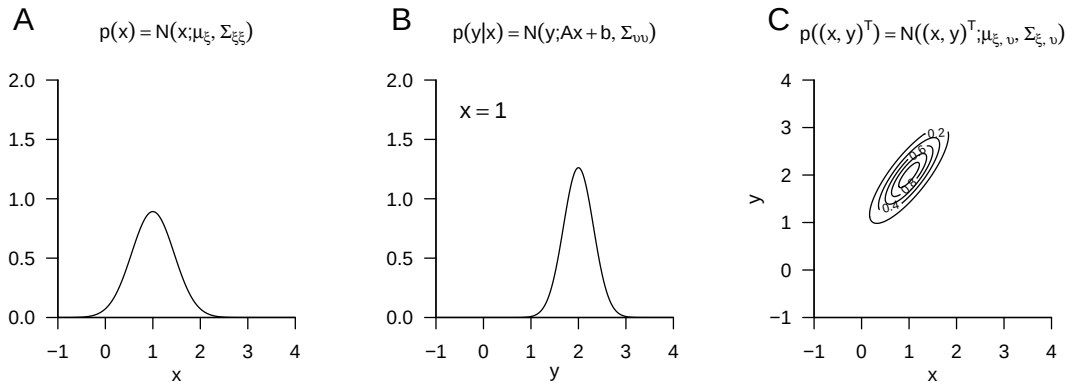

Abbildung 4.2 Gemeinsame Verteilungen einer marginalen und einer auf dieser bedingten normalverteilten Zufallsvariable.

Die Definition einer multivariaten Normalverteilung erlaubt es weiterhin, die bedingten Verteilungen aller Komponenten des entsprechenden Zufallsvektors direkt mithilfe der Parameter der multivariaten Normalverteilung zu bestimmen. Dies ist die zentrale Aussage folgenden Theorems.

Theorem 4.28 (Bedingte Normalverteilungen)**.** (ξ, υ) *sei ein* $m + n$*-dimensionaler normalverteilter Zufallsvektor mit WDF*

$$p : \mathbb{R}^{m+n} \to \mathbb{R}_{>0}, \begin{pmatrix} x \\ y \end{pmatrix} \mapsto p\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) := N\left(\begin{pmatrix} x \\ y \end{pmatrix}; \mu_{\xi,\upsilon}, \Sigma_{\xi,\upsilon}\right), \tag{4.111}$$

mit

$$\mu_{\xi,\upsilon} = \begin{pmatrix} \mu_\xi \\ \mu_\upsilon \end{pmatrix}, \Sigma_{\xi,\upsilon} = \begin{pmatrix} \Sigma_{\xi\xi} & \Sigma_{\xi\upsilon} \\ \Sigma_{\upsilon\xi} & \Sigma_{\upsilon\upsilon} \end{pmatrix}, \tag{4.112}$$

mit $x, \mu_\xi \in \mathbb{R}^m, y, \mu_\upsilon \in \mathbb{R}^n$ *und* $\Sigma_{\xi\xi} \in \mathbb{R}^{m\times m}, \Sigma_{\xi\upsilon} \in \mathbb{R}^{m\times n}, \Sigma_{\upsilon\upsilon} \in \mathbb{R}^{n\times n}$*. Dann ist die bedingte Verteilung von* ξ *gegeben* υ *eine* m*-dimensionale Normalverteilung mit bedingter WDF*

$$p : \mathbb{R}^m \to \mathbb{R}_{>0}, x \mapsto p(x|y) := N(x; \mu_{\xi|\upsilon}, \Sigma_{\xi|\upsilon}) \tag{4.113}$$

mit

$$\mu_{\xi|\upsilon} = \mu_\xi + \Sigma_{\xi\upsilon}\Sigma_{\upsilon\upsilon}^{-1}(y - \mu_\upsilon) \in \mathbb{R}^m \tag{4.114}$$

und

$$\Sigma_{\xi|\upsilon} = \Sigma_{\xi\xi} - \Sigma_{\xi\upsilon}\Sigma_{\upsilon\upsilon}^{-1}\Sigma_{\upsilon\xi} \in \mathbb{R}^{m\times m}. \tag{4.115}$$

∘

Im Zusammenspiel mit Theorem 4.27 und Theorem 4.26 können die Parameter bedingter und marginaler Normalverteilungen also aus den Parametern der komplementären bedingten und marginalen Normalverteilungen bestimmt werden. Abbildung 4.3 visualisiert Theorem 4.28 für den Fall $m := 2, n := 1$,

$$\mu := \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ und } \Sigma := \begin{pmatrix} 0.12 & 0.09 \\ 0.09 & 0.12 \end{pmatrix} \tag{4.116}$$

Dabei zeigt Abbildung 4.3 A die WDF des bivariaten Zufallsvektors $(\xi, \upsilon)^T$ und Abbildung 4.3 B und C zeigen die WDF der bedingten Verteilung der Zufallsvariable ξ gegeben $\upsilon = 1.5$ und $\upsilon = 2.8$, respektive.

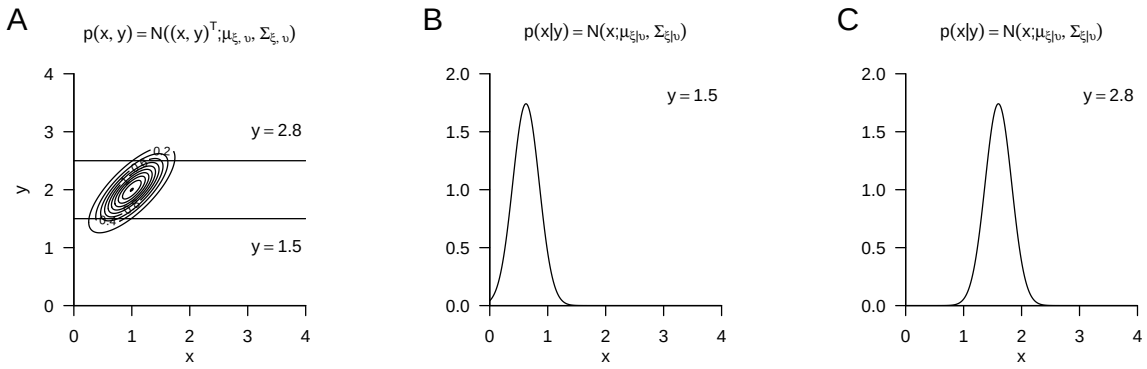

Abbildung 4.3 Bedingte Normalverteilungen

Lineare Transformationen

In diesem Abschnitt stellen wir einige Resultate zu den Verteilungen transformierter normalverteilter Zufallsvektoren zusammen. Wir verzichten dabei auf Beweise.

Theorem 4.29 (Invertierbare lineare Transformation eines normalverteilten Zufallsvektors)**.** *$\xi \sim N(\mu_\xi, \Sigma_\xi)$ sei ein normalverteilter n-dimensionaler Zufallsvektor und es sei $\zeta := A\xi$ mit einer invertierbaren Matrix $A \in \mathbb{R}^{n \times n}$. Dann gilt*

$$\zeta \sim N\left(\mu_\zeta, \Sigma_\zeta\right) \text{ mit } \mu_\zeta = A\mu_\xi \text{ und } \Sigma_\zeta = A\Sigma_\xi A^T. \tag{4.117}$$

∘

Nach Theorem 4.29 ergibt die invertierbare lineare Transformation eines multivariat normalverteilten Zuallsvektors also wiederum einen multivariat normalverteilten Zufallsvektor und die Parameter der Verteilung dieses normalverteilten Zufallsvektors ergeben sich aus den Parametern der Verteilung des ursprünglichen Zufallsvektors und der Transformationsmatrix.

Beispiel

Als Beispiel betrachten wir die invertierbare lineare Transformation eines bivariaten normalverteilten Zufallsvektors ξ. Es seien

$$\mu_\xi := \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ und } \Sigma_\xi := \begin{pmatrix} 0.20 & 0.15 \\ 0.15 & 0.20 \end{pmatrix} \tag{4.118}$$

der Erwartungswert- und Kovarianzmatrixparameter von ξ, respektive, und es sei

$$A := \begin{pmatrix} -2 & 1 \\ -1 & 2 \end{pmatrix} \tag{4.119}$$

die Transformationsmatrix. Da $|A| = -3 \neq 0$ ist A invertierbar und es gilt nach Theorem 4.29, dass

$$\zeta \sim N(\mu_\zeta, \Sigma_\zeta) \text{ mit } \mu_\zeta = A\mu_\xi = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \text{ und } A\Sigma_\xi A^T = \begin{pmatrix} 0.40 & 0.05 \\ 0.05 & 0.40 \end{pmatrix}. \tag{4.120}$$

Abbildung 4.4 A zeigt Isokonturen der WDF von ξ und Realisierungen $x^{(i)} \in \mathbb{R}^2$ von ξ für $i = 1, ..., 50$. Abbildung 4.4 B zeigt die transfomierten Realisierungen $z^{(i)} = Ax^{(i)} \in \mathbb{R}^2$ von ζ sowie die Isokonturen der WDF von ζ nach Theorem 4.29.

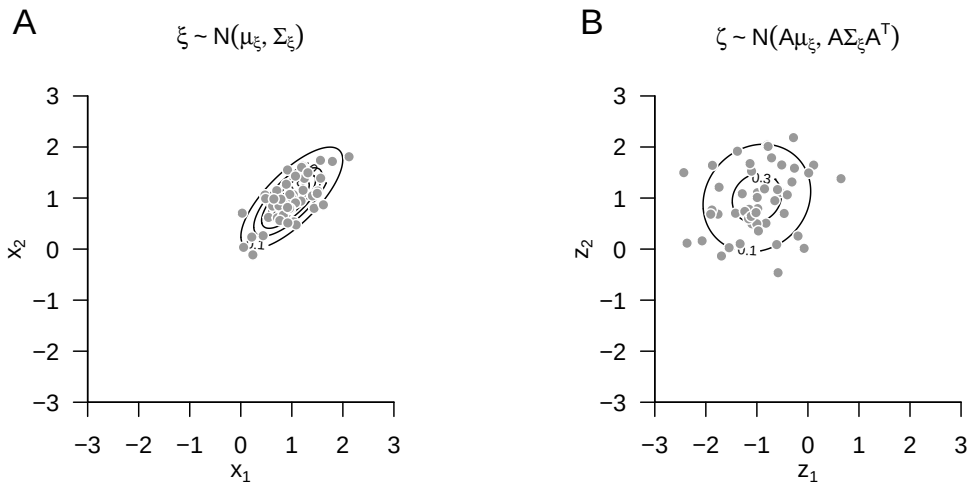

Abbildung 4.4 Invertierbare lineare Transformation eines normalverteilten Zufallsvektors

Die Tatsache, dass ein linear transfomierter normalverteilter Zufallsvektor wiederum normalverteilt ist und dass sich die Parameter der Verteilung des transformierten Zufallsvektors aus den Parametern der Verteilung des ursprünglichen Zufallsvektors sowie den Transformationsparametern bestimmen lassen, bleibt auch im Falle einer nicht notwendigerweise invertierbaren linearen Transformation und auch im Falle einer nicht notwendigerweise invertierbaren linear-affinen Transformation wahr. Dies ist die Aussage folgenden zentralen Theorems. Für einen Beweis verweisen wir auf Anderson (2003).

Theorem 4.30 (Linear-affine Transformation eines normalverteilten Zufallsvektors)**.** $\xi \sim N(\mu_\xi, \Sigma_\xi)$ *sei ein normalverteilter n-dimensionaler Zufallsvektor und es sei*

$$\zeta := Ax + b \text{ mit } A \in \mathbb{R}^{m \times n} \text{ und } b \in \mathbb{R}^m. \tag{4.121}$$

Dann gilt

$$\zeta \sim N(\mu_\zeta, \Sigma_\zeta) \text{ mit } \mu_\zeta = A\mu + b \in \mathbb{R}^m \text{ und } \Sigma_\zeta = A\Sigma A^T \in \mathbb{R}^{m \times m}. \tag{4.122}$$

∘

Referenzen

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis* (3rd ed). Wiley-Interscience.

Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. F. (1996). Comparison of Beck Depression Inventories-IA and-II in Psychiatric Outpatients. *Journal of Personality Assessment*, *67*(3), 588–597. https://doi.org/10.1207/s15327752jpa6703_13

Bohus, M., Kleindienst, N., Limberger, M. F., Stieglitz, R.-D., Domsalla, M., Chapman, A. L., Steil, R., Philipsen, A., & Wolf, M. (2009). The Short Version of the Borderline Symptom List (BSL-23): Development and Initial Data on Psychometric Properties. *Psychopathology*, *42*(1), 32–39. https://doi.org/10.1159/000173701

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. Wiley New York.

Borsboom, D. (2009). *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*. Cambridge University Press.

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The Concept of Validity. *Psychological Review*, *111*(4), 1061–1071. https://doi.org/10.1037/0033-295X.111.4.1061

Bühner, M. (2010). *Einführung in die Test- und Fragebogenkonstruktion* (2., aktualisierte und erw. Aufl., [Nachdr.]). Pearson Studium.

Cattell, R. B. (1957). *Personality and Motivation - Structure and Measurement*.

Cronbach, L. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, *16*(3), 297–334.

De Beurs, E., Boehnke, J. R., & Fried, E. I. (2022). Common Measures or Common Metrics? A Plea to Harmonize Measurement Results. *Clinical Psychology & Psychotherapy*, *29*(5), 1755–1767. https://doi.org/10.1002/cpp.2742

Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). *Mathematics for Machine Learning* (1. Aufl.). Cambridge University Press. https://doi.org/10.1017/9781108679930

Derogatis, L. R., & Melisaratos, N. (1983). The Brief Symptom Inventory: An Introductory Report. *Psychological Medicine*, *13*(3), 595–605. https://doi.org/10.1017/S0033291700048017

Feldt, L. S. (1965). The Approximate Sampling Distribution of Kuder-Richardson Reliability Coefficient Twenty. *Psychometrika*, *30*(3), 357–370. https://doi.org/10.1007/BF02289499

Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical Inference for Coefficient Alpha. *Applied Psychological Measurement*, *11*(1), 93–103. https://doi.org/10.1177/014662168701100107

Ghojogh, B., Ghodsi, A., Karray, F., & Crowley, M. (2022). *Factor Analysis, Probabilistic Principal Component Analysis, Variational Inference, and Variational Autoencoder: Tutorial and Survey* (arXiv:2101.00734). arXiv. https://doi.org/10.48550/arXiv.2101.00734

Gulliksen, H. (1950). *Theory of Mental Tests*. John Wiley & Sons Inc. https://doi.org/10.1037/13240-000

Hautzinger, M., Keller, F., & Kühner, C. (2006). *BDI-II Beck Depressions-Inventar*. Pearson.

Horst, P. (1965). *Factor Analysis of Data Matrices*. Holt, Rinehart and Winston, Inc.
Hotelling, H. (1933). Analysis of Complex Variables into Principal Components. *Journal of Educational Psychology*, *24*, 417-441 and 498-520.
Hyvärinen, A., Khemakhem, I., & Monti, R. (2024). Identifiability of Latent-Variable and Structural-Equation Models: From Linear to Nonlinear. *Annals of the Institute of Statistical Mathematics*, *76*(1), 1–33. https://doi.org/10.1007/s10463-023-00884-4
Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous Univariate Distributions, Volume 2* (2nd ed). Wiley.
Jöreskog, K. G. (1967). *Some Contributions to Maximum Likelihood Factor Analysis*.
Jöreskog, K. G. (1969). *A General Approach to Confirmatory Maximum Likelihood Factor Analysis*.
Kaiser, H. F. (1958). The Varimax Criterion for Analytic Rotation in Factor Analysis. *Psychometrika*, *23*(3), 187–200. https://doi.org/10.1007/BF02289233
Kaiser, H. F. (1970). A Second Generation Little Jiffy. *Psychometrika*, *35*(4), 401–415. https://doi.org/10.1007/BF02291817
Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark Iv. *Educational and Psychological Measurement*, *34*(1), 111–117. https://doi.org/10.1177/001316447403400115
Keller, F., Hautzinger, M., & Kühner, C. (2008). Zur faktoriellen Struktur des deutschsprachigen BDI-II. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *37*(4), 245–254. https://doi.org/10.1026/1616-3443.37.4.245
Krauth, J. (1995). *Testkonstruktion und Testtheorie*. Beltz, Psychologie Verl.-Union.
Kristof, W. (1963). The Statistical Theory of Stepped-up Reliability Coefficients When a Test Has Been Divided into Several Equivalent Parts. *Psychometrika*, *28*(3), 221–2238. https://doi.org/10.1007/BF02289571
Lawley, D. (1940). The Estimation of Factor Loadings by the Method of Maximum Likelihood. *Proceedings of the Royal Society of Edinburgh. Section B: Biological Sciences*.
Lawley, D., & Maxwell, A. (1971). *Factor Analysis as a Statistical Method* (2. Aufl.). London Butterworths.
Lazarsfeld, P. (1959). Latent Structure Analysis. In *Psychology: A Study of a Science, Vol. 3*. McGraw-Hill.
Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Auflage). Beltz, Psychologie Verlags Union.
Lord, F. M. (1980). *Applications of Item Response Theory To Practical Testing Problems* (0. Aufl.). Routledge. https://doi.org/10.4324/9780203056615
Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores* (Nachdr. der Ausg. Reading, Mass. [u.a.], 1968). Information Age Publ.
Magnus, J. R., & Neudecker, H. (1989). Matrix Differential Calculus with Applications in Statistics and Econometrics. *Journal of the American Statistical Association*, *84*(408), 1103. https://doi.org/10.2307/2290109
Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press.
McCall, W. A. (1922). *How to Measure in Education*. MacMillan Co. https://doi.org/10.1037/13551-000
Meintrup, D., & Schäffler, S. (2005). *Stochastik: Theorie und Anwendungen*. Springer.
Moosbrugger, H., & Kelava, A. (Hrsg.). (2012). *Testtheorie und Fragebogenkonstruktion: mit 66 Abbildungen und 41 Tabellen* (2., aktualisierte und überarbeitete Auflage). Springer.
Neudecker, H. (1969). Some Theorems on Matrix Differentiation with Special Reference to Kronecker Matrix Products. *Journal of the American Statistical Association*, *64*.
Novick, M. R. (1966). The Axioms and Principal Results of Classical Test Theory. *Journal*

of Mathematical Psychology, *3*(1), 1–18. https://doi.org/10.1016/0022-2496(66)90002-2

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests* (Expanded ed). University of Chicago Press.

Reise, S. P., & Waller, N. G. (2009). Item Response Theory and Clinical Measurement. *Annual Review of Clinical Psychology*, *5*(1), 27–48. https://doi.org/10.1146/annurev.clinpsy.032408.153553

Rencher, A. C., & Christensen, W. F. (2012). *Methods of Multivariate Analysis* (Third Edition). Wiley.

Revelle, W. (2024). *Psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University.

Rohe, K., & Zeng, M. (2023). Vintage Factor Analysis with Varimax Performs Statistical Inference. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *85*(4), 1037–1060. https://doi.org/10.1093/jrsssb/qkad029

Rosseel, Y. (2012). **Lavaan** : An *R* Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.18637/jss.v048.i02

Rosseel, Y. (2021). Evaluating the Observed Log-Likelihood Function in Two-Level Structural Equation Modeling with Missing Data: From Formulas to R Code. *Psych*, *3*(2), 197–232. https://doi.org/10.3390/psych3020017

Rost, J. (2004). *Lehrbuch Testtheorie, Testkonstruktion* (2., vollständig überarbeitete und erweiterte Aufl). H. Huber.

Roweis, S., & Ghahramani, Z. (1999). A Unifying Review of Linear Gaussian Models. *Neural Computation*, *11*(2), 305–345. https://doi.org/10.1162/089976699300016674

Rubin, D. B., & Thayer, D. T. (1982). EM Algorithms for ML Factor Analysis. *Psychometrika*, *47*(1), 69–76. https://doi.org/10.1007/BF02293851

Schmidt, K. D. (2009). *Maß und Wahrscheinlichkeit*. Springer.

Seashore, H. G. (1955). Methods of Expressing Test Scores. *Test Service Bulletin*, *48*, 7–10.

Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, *15*(2), 201. https://doi.org/10.2307/1412107

Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science, New Series*, *103*(2684), 677–680. https://www.jstor.org/stable/1671815

Thurstone, L. (1936). The Factorial Isolation of Primary Abilities. *Psychometrika*, *1*(3), 175–182. https://doi.org/10.1007/BF02288363

Thurstone, L. (1937). Current Misuse of the Factorial Methods. *Psychometrika*, *2*(2), 73–76. https://doi.org/10.1007/BF02288060

Thurstone, L. (1947). Multiple Factor Analysis. *University of Chicago Press.*

Van Zyl, J. M., Neudecker, H., & Nel, D. G. (2000). On the Distribution of the Maximum Likelihood Estimator of Cronbach's Alpha. *Psychometrika*, *65*(3), 271–280. https://doi.org/10.1007/BF02296146

Wishart, J. (1928). The Generalised Product Moment Distribution in Samples from a Normal Multivariate Population. *Biometrika*, *20A*(1/2), 32. https://doi.org/10.2307/2331939

Yanai, H., & Ichikawa, M. (2007). Factor Analysis. In *Handbook of Statistics, Vol 26* (1st ed). Elsevier.

Yuan, K.-H., & Bentler, P. M. (2002). On Robusiness of the Normal-Theory Based Asymptotic Distributions of Three Reliability Coefficient Estimates. *Psychometrika*, *67*(2), 251–259. https://doi.org/10.1007/BF02294845