

Bachelorarbeit

Bayesianische Analyse der Einflussfaktoren auf die Schlafqualität: Eine Fallstudie basierend auf Smart-Ring-Daten

Fakultät für Statistik
Ludwig-Maximilians-Universität München

Bianca Laetitia Zettler

München, 4. Dezember 2024



Eingereicht zur Erlangung des Bachelor of Science.
Betreut von Prof. Dr. Volker Schmid

Abstract

Die vorliegende Arbeit befasst sich mit der Untersuchung von Einflussfaktoren auf die Schlafqualität mittels einer bayesianischen Analyse. Dabei werden sowohl direkte als auch indirekte Mechanismen betrachtet. Grundlage der Analyse sind Daten eines Smart Rings, ergänzt durch Wetter- und Mondphasendaten, die eine umfassende Betrachtung der physiologischen und externen Einflussfaktoren ermöglichen.

Die Arbeit gliedert sich in mehrere Abschnitte. Zunächst wird die Datengrundlage detailliert beschrieben und eine umfassende deskriptive Analyse vorgenommen, um erste Erkenntnisse der Einflussfaktoren auf die Schlafqualität zu gewinnen. Im Hauptteil werden die Variablen sorgfältig selektiert und mittels bayesianischer Statistik modelliert. Im Rahmen der weiteren Untersuchung erfolgt eine Differenzierung in Bezug auf die Modellierung direkter und indirekter Einflussfaktoren auf die Schlafqualität.

Die Ergebnisse der Modellierung direkter Einflussfaktoren auf die Schlafqualität zeigen, dass die Atemfrequenz der stärkste direkte Prädiktor der Schlafqualität ist. Eine niedrigere Atemfrequenz korreliert mit einer höheren Schlafqualität. Eine weitere wichtige Einflussgröße ist die durchschnittliche Temperatur, die sich bei hohen Werten negativ auf den Schlaf auswirkt, sowie der Erholungsindex der eine deutliche positive Wirkung auf die Schlafqualität hat. Physiologische Parameter wie der Ruheherzfrequenzindex und der Körpertemperaturindex zeigten ebenfalls signifikante, aber weniger starke Effekte. Die Sonnenscheindauer, die Herzfrequenzvariabilität und das Datum hatten hingegen keine statistisch signifikante Wirkung.

Im zweiten Teil der Analyse erwies sich die Atemfrequenz als ein zentraler Faktor mit starker Korrelation zur Schlafqualität. Bei der Modellierung indirekter Einflussfaktoren auf die Schlafqualität wurde deutlich, dass Körpertemperaturschwankungen negativ und die durchschnittliche Tagestemperatur positiv auf die Atemfrequenz wirken, wobei eine niedrige Atemfrequenz im Sinne der Schlafqualität erwünscht ist. Diese Effekte fielen jedoch vergleichsweise gering aus. Andere untersuchte Prädiktoren erwiesen sich in diesem Kontext als nicht signifikant.

Diese Ergebnisse der bayesianischen Analysen werden präsentiert, interpretiert und evaluiert. Die Arbeit zeigt die Stärke bayesianischer Modellierungsansätze auf, die Unsicherheiten explizit quantifizieren und robuste Schätzungen der Modellparameter ermöglichen. Obwohl nichtlineare Modelle und Interaktionseffekte keine wesentlichen zusätzlichen Erkenntnisse lieferten, unterstützen die Ergebnisse die Einfachheit und Interpretierbarkeit linearer Modelle. Abschließend werden die Limitationen der Arbeit diskutiert und zusammengefasst. Limitationen der Arbeit, wie die spezifische Datengrundlage und eingeschränkte Generalisierbarkeit, verdeutlichen jedoch die Notwendigkeit zukünftiger Studien mit erweiterten Variablen und umfassenderen Daten.

Die Ergebnisse dieser Arbeit leisten einen wesentlichen Beitrag zum Verständnis der komplexen Mechanismen, die die Schlafqualität beeinflussen. Sie zeigen, dass sowohl physiologische als auch externe Umweltfaktoren eine wesentliche Rolle spielen und dass datengetriebene Ansätze wertvolle Erkenntnisse für die Verbesserung der individuellen Schlafqualität bieten können. Die in dieser Arbeit präsentierten Ansätze und Erkenntnisse können als Ausgangspunkt für weiterführende Studien dienen, die das Verständnis des Schlafs bereichern.

Inhaltsverzeichnis

1	Einleitung	1
2	Datengrundlage	3
2.1	Datenbeschreibung	3
2.1.1	Wichtige Modellvariablen	5
2.2	Datenvorbereitung	7
3	Deskriptive Datenanalyse	8
3.1	Verteilung des Schlafwerts	8
3.2	Zusammenhang zwischen Schlafwert und Einflussgrößen	9
4	Modellierung direkter Einflussfaktoren auf die Schlafqualität	12
4.1	Variablenselektion	12
4.1.1	Inhaltliche Überlegungen	12
4.1.2	Korrelationsanalyse	13
4.1.3	Prüfung auf Multikollinearität	14
4.1.4	Lasso-Regression	15
4.2	Bayesianisches lineares Modell mit Zielgröße Schlafwert	17
4.2.1	Modellformel	17
4.2.2	Priorverteilungen und Schätzung	17
4.2.3	Ergebnisse und Interpretation	18
4.2.4	Erweiterung des Modells um Interaktionen	19
4.3	Bayesianisches Modell mit Spline für die Atemfrequenz	20
4.3.1	Modellformel	20
4.3.2	Ergebnisse und Interpretation	21
4.4	Bayesianisches Modell mit Polynom für die Atemfrequenz	22
4.4.1	Modellformel	23
4.4.2	Ergebnisse und Interpretation	23
4.5	Evaluierung der Modelle	24
5	Modellierung indirekter Einflussfaktoren auf die Schlafqualität	28
5.1	Variablenselektion	28
5.1.1	Inhaltliche Überlegungen	28
5.1.2	Korrelationsanalyse	28
5.1.3	Prüfung auf Multikollinearität	30
5.1.4	Lasso-Regression	30
5.2	Bayesianisches lineares Modell mit Zielgröße Atemfrequenz	31
5.2.1	Erweiterung des Modells um Interaktionen	33
5.3	Nichtlineare bayesianische Modelle mit Zielgröße Atemfrequenz	33
5.4	Modellevaluierung	34

6	Ergebnisse und Diskussion	36
6.1	Direkte Einflussgrößen auf die Schlafqualität	36
6.2	Indirekte Einflussgrößen auf die Schlafqualität	37
6.3	Limitationen	38
7	Fazit	39
A	Anhang	41
B	Elektronischer Anhang	44
C	Abbildungsverzeichnis	45
D	Tabellenverzeichnis	46
E	Literaturverzeichnis	47
F	Eidesstattliche Erklärung	51

1 Einleitung

Schlaf ist eine der essenziellen Säulen menschlicher Gesundheit und Wohlbefindens. Er beeinflusst nicht nur die physische Regeneration und die Stabilisierung des Immunsystems, sondern ist auch entscheidend für die emotionale Ausgeglichenheit, die kognitive Leistungsfähigkeit und die langfristige Lebensqualität. In einer zunehmend hektischen und von digitalen Medien geprägten Gesellschaft stehen die Qualität und Quantität des Schlafs jedoch unter ständigem Druck. Ungünstige Schlafmuster und Schlafstörungen sind weit verbreitet und stellen eine erhebliche Herausforderung für die öffentliche Gesundheit dar. Vor diesem Hintergrund ist die Untersuchung der Faktoren, die die Schlafqualität beeinflussen, sowohl von wissenschaftlichem als auch von praktischem Interesse. Die Analyse der Schlafqualität gestaltet sich jedoch als komplex, da sie von einer Vielzahl interner und externer Faktoren beeinflusst wird. Physiologische Parameter, wie Herzfrequenz und Atemfrequenz, interagieren mit umweltbedingten Faktoren, wie der Umgebungstemperatur, und verhaltensbezogenen Aspekten, wie der körperlichen Aktivität oder den Schlafgewohnheiten. Diese Einflussfaktoren stehen oft in nichtlinearen und dynamischen Wechselbeziehungen, die eine präzise Modellierung und Analyse erfordern. Zudem können indirekte Mechanismen die Wirkung bestimmter Variablen auf die Schlafqualität verstärken oder abschwächen, was die Identifikation kausaler Zusammenhänge zusätzlich erschwert.

In der vorliegenden Arbeit wird die bayesianische Statistik als methodischer Ansatz verwendet, um die Einflussfaktoren auf die Schlafqualität systematisch zu untersuchen. Bayesianische Modelle bieten eine flexible Möglichkeit, Unsicherheiten in den Daten explizit zu quantifizieren und Vorwissen zu integrieren. Diese Ansätze sind besonders geeignet, um komplexe Datenstrukturen zu analysieren und sowohl lineare als auch nichtlineare Zusammenhänge zu berücksichtigen. Durch die probabilistische Natur bayesianischer Methoden können Ergebnisse nicht nur als Punktschätzungen, sondern auch als Wahrscheinlichkeitsverteilungen dargestellt werden, was eine differenziertere Interpretation der Zusammenhänge ermöglicht.

Ziel der Arbeit ist es, die zentralen Einflussfaktoren auf die Schlafqualität zu identifizieren und deren Zusammenhänge zu modellieren. Dabei wird ein zweistufiger Ansatz verfolgt, der sowohl direkte als auch indirekte Mechanismen berücksichtigt. Die Grundlage der Analyse bildet ein umfangreicher Datensatz, der über einen Zeitraum von mehr als 3.5 Jahren mit einem Smart Ring erhoben wurde. Ergänzend wurden Wetter- und Mondphasendaten integriert, um die Analyse um externe Umweltfaktoren zu erweitern. Diese Kombination erlaubt es, die Einflussfaktoren auf die Schlafqualität aus einer ganzheitlichen Perspektive zu betrachten und ein tieferes Verständnis der zugrunde liegenden Mechanismen zu entwickeln.

Die Arbeit ist wie folgt strukturiert: Im ersten Abschnitt wird die Datengrundlage detailliert beschrieben, einschließlich der Messmethoden und der erhobenen Variablen. Anschließend erfolgt eine deskriptive Analyse, um erste Einblicke in die Verteilung und Struktur der Daten zu gewinnen. Im Hauptteil werden die Daten mittels bayesianischer Statistik analysiert. Dabei wird zwischen der Modellierung direkter Einflussfaktoren auf die Schlafqualität und der Untersuchung indirekter Mechanismen unterschieden. Abschließend werden die Ergebnisse interpretiert und im Kontext bestehender Forschung diskutiert. Zudem werden Limitationen der Arbeit sowie Ansätze für zukünftige Forschung aufgezeigt.

Durch die Kombination innovativer statistischer Ansätze mit einer datengetriebenen Perspektive leistet diese Arbeit einen Beitrag zum Verständnis der Faktoren, die die Schlafqualität beeinflussen. Gleichzeitig werden methodische Ansätze bereitgestellt, die in zukünftigen Studien und in der Entwicklung praktischer Empfehlungen zur Förderung eines besseren Schlafs Anwendung finden können.

2 Datengrundlage

Die Datengrundlage dieser Arbeit besteht einerseits aus individuellen Gesundheitsdaten, die über einen Zeitraum von mehr als 3,5 Jahren (09.02.2021 bis 13.10.2024) mit einem Smart Ring gesammelt wurden. Dieser Smart Ring zeichnet verschiedene Schlaf-, Aktivitäts- und Vitalparameter auf, die detaillierte Einblicke in die Schlafqualität und den allgemeinen Gesundheitszustand ermöglichen. Die Grundlage dieser Arbeit basiert nicht ausschließlich auf den durch den Smart Ring erfassten Daten, sondern wurde durch externe Datenquellen ergänzt, um ein umfassenderes Bild der Einflussfaktoren auf die Schlafqualität zu ermöglichen. Diese Ergänzungen beinhalten insbesondere Informationen zu Mondphasen und Wetterbedingungen, die als potenzielle externe Einflussgrößen auf die Schlafqualität untersucht wurden.

2.1 Datenbeschreibung

Die Daten des Smart Rings enthalten 54 gemessene Größen, die in verschiedene Kategorien unterteilt werden können. Um eine detaillierte Übersicht zu geben, werden in Tabelle 1 die wichtigsten Variablen der Smart-Ring-Daten beschrieben. Hierbei sind die genannten “Scores” Metriken, die auf einer Skala von 0 bis 100 eine Bewertung über die jeweiligen Parameter abgeben.

Zusätzlich wurden Wetterdaten für die Region München, in der die Smart Ring-Daten gesammelt wurden, in die Analyse aufgenommen. Diese Daten wurden von der Plattform *Meteostat* bezogen und decken den Zeitraum von Februar 2021 bis Oktober 2024 ab. Der Datensatz umfasst elf Variablen, die in Tabelle 2 dargestellt sind. Die Wettervariablen umfassen unter anderem Temperaturen, Niederschlagsmengen, Windparameter und Sonnenscheindauer, die als potenziell relevante Umweltfaktoren für die Schlafqualität betrachtet werden können.

Die Mondphasendaten wurden mithilfe des R-Pakets *lunar* berechnet. Für jedes Datum im Datensatz wurde der jeweilige Mondphasenstatus festgehalten: zunehmend, abnehmend oder Vollmond. Ziel dieser Integration war es, mögliche Effekte des Mondes auf die Schlafqualität zu analysieren.

Die Hinzunahme dieser externen Datenquellen zielte darauf ab, zusätzliche erklärende Variablen zu integrieren, die nicht im ursprünglichen Datensatz des Smart Rings enthalten waren. Dadurch soll ein höherer Informationsgewinn aus den Modellen erzielt werden, insbesondere durch die Identifikation von Umwelteinflüssen, die die Schlafqualität direkt oder indirekt beeinflussen könnten.

Tabelle 1: Überblick über die Variablen der Smart-Ring-Daten

Kategorie	Variablen	Beschreibung
Schlafparameter	Sleep.Score, Total.Sleep.Score, REM.Sleep.Score, Deep.Sleep.Score, Restfulness.Score, Sleep.Efficiency.Score, Sleep.Latency.Score, Sleep.Timing.Score	Metriken zur Bewertung der Schlafqualität, Effizienz und Dauer in verschiedenen Schlafphasen
Vitalparameter	Average.Resting.Heart.Rate, Lowest.Resting.Heart.Rate, Average.HRV, Respiratory.Rate, Temperature.Deviation.C., Temperature.Trend.Deviation, Temperature.Score	Herzfrequenz- und Herzraten-Variabilitätsmessungen sowie die Atemfrequenz und Abweichungen der Körpertemperatur
Aktivitätsparameter	Activity.Score, Stay.Active.Score, Move.Every.Hour.Score, Meet.Daily.Targets.Score, Training.Frequency.Score, Training.Volume.Score, Activity.Burn, Total.Burn, Steps, Equivalent.Walking.Distance	Aktivitätsbezogene Kennzahlen, einschließlich Zielerreichung, Trainingshäufigkeit und Intensität sowie verbrannter Kalorien
Erholungsparameter	Readiness.Score, Recovery.Index.Score, Previous.Night.Score, Previous.Day.Activity.Score, Activity.Balance.Score, HRV.Balance.Score	Bewertung der Erholung nach Schlaf und Aktivität
Metabolische Parameter	Average.MET	Durchschnittlicher Stoffwechsel-Äquivalentwert

Tabelle 2: Beschreibung der Wetterdaten für die Region München

Variable	Beschreibung
date	Datum der Beobachtung
tavg	Durchschnittstemperatur des Tages (°C)
tmin	Minimale Temperatur des Tages (°C)
tmax	Maximale Temperatur des Tages (°C)
prcp	Gesamtniederschlagsmenge des Tages (mm)
snow	Schneehöhe am Tag (mm)
wdir	Durchschnittliche Windrichtung (Grad)
wspd	Durchschnittliche Windgeschwindigkeit (km/h)
wpgt	Spitzenböe des Tages (km/h)
pres	Luftdruck am Boden (hPa)
tsun	Sonnenscheindauer des Tages (Minuten)

2.1.1 Wichtige Modellvariablen

In den Modellen dieser Arbeit wurden zentrale Variablen aus den Smart-Ring-Daten sowie ergänzende Wetterdaten verwendet, um die Prädiktoren auf die Schlafqualität zu modellieren. Diese Auswahl berücksichtigt sowohl interne Faktoren, die durch Vitalparameter repräsentiert werden, als auch externe Umweltbedingungen, wie Wetterdaten, um eine möglichst umfassende Analyse zu gewährleisten. Der Selektionsprozess dieser Einflussfaktoren wird in den Abschnitten 4.1 und 5.1 genauer betrachtet. In Tabelle 3 sind diese wichtigsten Variablen zusammengefasst.

Tabelle 3: Beschreibung der in den Modellen verwendeten Variablen

Variable	Beschreibung
Schlafwert (Sleep.Score)	Bewertung der Schlafqualität (0–100), basierend auf Gesamtschlafdauer, Schlaffeffizienz, Schlafphasen und anderen Faktoren
Atemfrequenz (Respiratory.Rate)	Durchschnittliche Anzahl der Atemzüge pro Minute während des Schlafs (in bpm)
Ruheherzfrequenzindex (Resting.Heart.Rate.Score)	Durchschnittliche Herzfrequenz in Ruhe, bewertet auf einer Skala von 0 bis 100
Herzfrequenzvariabilitätsindex (HRV.Balance.Score)	Bewertung der Herzfrequenzvariabilität (0–100) in Relation zu individuellen Basiswerten
Erholungsindex (Recovery.Index.Score)	Bewertung der Erholung nach Belastung (0–100), ein höherer Wert deutet auf bessere Regeneration hin
Körpertemperaturindex (Temperature.Score)	Bewertung der Körpertemperaturabweichung vom Normalwert (0–100), ein höherer Score deutet auf eine geringere Abweichung hin
Vortagsaktivitätsindex (Previous.Day.Activity.Score)	Bewertung der gesamten Aktivität des vorherigen Tages (0–100)
Durchschnittstemperatur (tavg)	Durchschnittliche Tagestemperatur in Grad Celsius
Datum (date_numeric)	Numerische Kodierung des Datums zur Erfassung von zeitlichen Trends
Sonnenscheindauer (tsun)	Gesamtdauer des täglichen Sonnenscheins in Minuten
Luftdruck (pres)	Tagesdurchschnitt des atmosphärischen Drucks in Hektopascal (hPa)

Als primäre Zielgröße bietet der Schlafscore (Sleep.Score) eine umfassende Bewertung der Schlafqualität. Dieser Score kombiniert mehrere Parameter wie die Gesamtschlafdauer, Schlaffeffizienz, Verteilung der Schlafphasen (z. B. REM- und Tiefschlaf), den Restfulness-Score sowie das Timing des Schlafs. Auf einer Skala von 0 bis 100 bewertet er die

Qualität und Effizienz des Schlafs, wobei höhere Werte auf eine bessere Schlafqualität hinweisen. Diese Zielgröße integriert physiologische und verhaltensbezogene Dimensionen und ermöglicht eine detaillierte Analyse der Schlafarchitektur.

Einen weiteren zentralen Parameter stellt die Atemfrequenz (`Respiratory.Rate`) dar. Sie misst die durchschnittliche Anzahl der Atemzüge pro Minute während des Schlafs. Veränderungen in der Atemfrequenz können Rückschlüsse auf Stresslevel, Erholungsprozesse und potenzielle Schlafstörungen zulassen. Im Gegensatz zum Schlafscore handelt es sich hierbei um eine kontinuierliche Messgröße ohne normierte Bewertung.

Zur Bewertung der Ruheherzfrequenz wird der Ruheherzfrequenzindex (`Resting.Heart.Rate.Score`) herangezogen. Dieser Score gibt auf einer Skala von 0 bis 100 an, wie optimal die durchschnittliche Herzfrequenz in Ruhe ist. Ein höherer Score deutet auf eine niedrigere Ruheherzfrequenz hin, was allgemein mit besserer Fitness und effektiver Erholung assoziiert wird. Im Gegensatz dazu kann ein niedriger Score auf Stress, Überbelastung oder gesundheitliche Probleme hinweisen.

Ergänzend dazu liefert der Herzfrequenzvariabilitätsindex (`HRV.Balance.Score`) Informationen zur Herzfrequenzvariabilität. Auf einer Skala von 0 bis 100 bewertet dieser Index die Anpassungsfähigkeit des Körpers an Stresssituationen. Eine höhere Variabilität, die durch einen höheren Score ausgedrückt wird, ist ein Indikator für bessere Regeneration und ein effektiveres Stressmanagement. Diese Variable steht in enger Verbindung mit der Schlafqualität und der Fähigkeit des Körpers, sich an wechselnde Bedingungen anzupassen. Der Erholungsindex (`Recovery.Index.Score`) misst die Effizienz der Regenerationsprozesse nach physischen oder mentalen Belastungen. Werte nahe 100 deuten auf eine erfolgreiche Erholung hin, während niedrigere Werte auf unzureichende Regenerationsmechanismen hindeuten können. Dieser Score liefert wichtige Hinweise darauf, wie gut der Körper auf Belastungen reagiert und wie sich dies auf die Schlafqualität auswirkt.

Die Bewertung der Körpertemperatur erfolgt über den Temperaturindex (`Temperature.Score`). Dieser stellt Abweichungen von der individuellen Normaltemperatur auf einer Skala von 0 bis 100 dar. Höhere Werte zeigen geringere Abweichungen an, während niedrigere Scores auf potenziell störende Faktoren wie Krankheit oder Stress hinweisen können. Da die Körpertemperatur eine zentrale Rolle in der Thermoregulation spielt, sind ihre Auswirkungen auf die Schlafarchitektur von besonderem Interesse.

Zur Erfassung der Aktivität dient der Vortagsaktivitätsindex (`Previous.Day.Activity.Score`). Er bewertet die physische Aktivität des vorherigen Tages und gibt Hinweise darauf, wie diese die Schlafqualität beeinflussen könnte. Ein hoher Wert deutet darauf hin, dass der Vortag von angemessener körperlicher Aktivität geprägt war.

Neben den internen Parametern geben externe Umweltvariablen wie die durchschnittliche Tagestemperatur (`tavg`), die Sonnenscheindauer (`tsun`) und der Luftdruck (`pres`) Einblick in mögliche Einflüsse auf die Schlafqualität. Hohe Temperaturen können die Thermoregulation stören, während eine ausreichende Sonnenscheindauer den circadianen Rhythmus stabilisiert. Ein hoher Luftdruck kann hingegen physiologische Anpassungsprozesse des Körpers beeinflussen. Die numerische Kodierung des Datums (`date_numeric`) ermöglicht zudem die Analyse saisonaler Muster und zeitlicher Trends, die langfristige Veränderungen in Schlafqualität und Atemfrequenz aufdecken können.

Diese Kombination aus normierten Scores und gemessenen Parametern bildet eine solide Grundlage für die Untersuchung der Einflussfaktoren auf Schlafqualität. Sowohl interne

physiologische Prozesse als auch externe Umweltbedingungen werden berücksichtigt, um ein umfassendes Bild der zugrunde liegenden Mechanismen zu zeichnen.

2.2 Datenvorbereitung

Die Datenvorbereitung ist ein zentraler Schritt, um die Rohdaten in eine für die Analyse geeignete Form zu überführen und sicherzustellen, dass die statistischen Modelle robuste und valide Ergebnisse liefern.

Ein wichtiger Aspekt der Datenaufbereitung war der Umgang mit fehlenden Werten. Im Datensatz wurden fehlende Werte über mehrere Monate identifiziert, in denen der Smart-Ring nicht getragen wurde. An Tagen, an denen nicht alle Werte der einzelnen Prädiktoren vorliegen, wurden die fehlenden durch den Mittelwert der jeweiligen Variable imputiert. Letzteres Verfahren stellt sicher, dass die Verteilungseigenschaften der Variablen erhalten bleiben, ohne dass die Stichprobengröße erheblich reduziert wird. Dieses Vorgehen minimiert Verzerrungen, die durch systematisches Entfernen von Beobachtungen entstehen könnten, und gewährleistet eine größere Stabilität der Modelle. Dadurch blieben Daten über zwei Jahre und neun Monate übrig.

Zusätzlich wurden die Datentypen der Variablen überprüft und entsprechend ihrer inhaltlichen Bedeutung angepasst. Für die spätere Korrelationsanalyse wurden die Merkmale zu numerischen Datentypen konvertiert, dabei wurde das Datum numerisch kodiert in "date_numeric", wobei der Startzeitpunkt (09.02.2021) als 1 definiert wurde und jeden Tag ganzzahlig hochzählt.

Ein weiterer Schritt war die Berechnung eines zusätzlichen potenziellen Prädiktors. Aus der Datumsspalte wurde die Variable `is_weekend` abgeleitet, die angibt, ob eine Beobachtung an einem Wochenende oder einem Wochentag stattfand. Diese Variable dient der Untersuchung eines möglichen Wochenend-Effekts, da anzunehmen ist, dass sich das Schlafverhalten und die damit verbundenen physiologischen Parameter an Wochenenden im Vergleich zu Wochentagen systematisch unterscheiden könnten.

Durch diese Schritte der Datenvorbereitung wurde sichergestellt, dass der Datensatz konsistent, vollständig und strukturell geeignet für die geplanten Analysen ist. Dies bildet die Grundlage für valide und aussagekräftige Ergebnisse in den nachfolgenden Modellierungs- und Interpretationsschritten.

3 Deskriptive Datenanalyse

Die deskriptive Analyse dient dazu, die zentralen Eigenschaften der Variablen, die in der Modellierung verwendet werden, detailliert zu untersuchen. Dieser Abschnitt bietet einen Überblick über die Verteilungen, Wertebereiche und statistischen Kennzahlen der relevanten Größen und legt somit die Grundlage für die Interpretation der Zusammenhänge im Kontext der Schlafqualität. Im Folgenden wird explizit auf die ausgewählten potenziellen Prädiktoren für die Schlafqualität eingegangen.

3.1 Verteilung des Schlafwerts

Der Schlafwert ist die zentrale Zielgröße dieser Untersuchung und bildet die Basis für die Modellierung der Schlafqualität. Um die Verteilung des Schlafwerts genauer zu untersuchen, wurde ein Dichteplot erstellt, der sowohl ein Histogramm als auch die geschätzte Dichtefunktion des Schlafwerts zeigt. Dieser Plot ist in Abbildung 1 dargestellt.

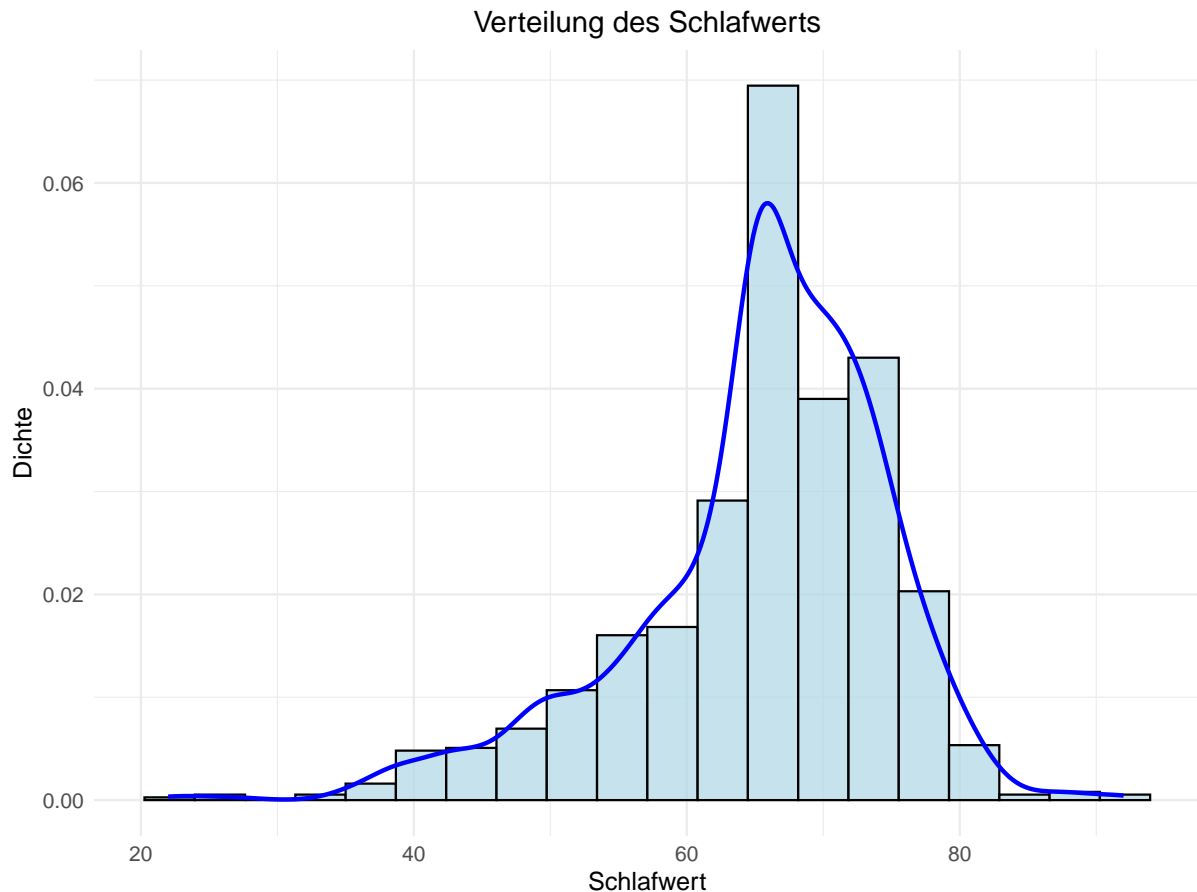


Abbildung 1: Dichteplot des Schlafwerts zur Überprüfung der Verteilung.

Die Verteilung des Schlafwerts zeigt eine unimodale Struktur mit einer leichten Schiefe nach links, was darauf hindeutet, dass die meisten Schlafwerte im Bereich zwischen 60

und 80 liegen. Dabei liegt der Median bei 66 und die Werte variieren zwischen 22 und 92. Der höchste Dichtewert befindet sich um die 70, was als häufigster Schlafwert identifiziert werden kann. Dies deutet darauf hin, dass die Mehrheit der untersuchten Nächte als durchschnittlich bis gut bewertet wurde. Die geringe Anzahl von Werten unterhalb von 40 und oberhalb von 85 zeigt, dass sowohl sehr niedrige als auch sehr hohe Schlafwerte selten sind.

Die Überlagerung der geschätzten Dichtefunktion mit dem Histogramm liefert eine nützliche Visualisierung, um die Glätte und das generelle Verteilungsmuster der Daten zu erkennen. Die Dichtekurve weist auf eine Annäherung an eine Normalverteilung hin, jedoch mit einem geringfügigen Überschuss an Datenpunkten im unteren Bereich, was auf leichte Abweichungen von der Symmetrie hindeutet.

Eine solche Analyse der Verteilung ist entscheidend, da sie grundlegende Annahmen für die spätere Modellierung betrifft. Die visuelle Inspektion legt nahe, dass der Schlafwert zumindest näherungsweise normalverteilt ist, was für die linearen Modellierungsansätze günstig ist. Dennoch weisen die beobachteten Abweichungen darauf hin, dass potenzielle Transformationen oder robuste Modellierungsansätze in Betracht gezogen werden könnten, um Verzerrungen zu minimieren.

3.2 Zusammenhang zwischen Schlafwert und Einflussgrößen

Die deskriptive Analyse der Variablen liefert einen umfassenden Überblick über die Verteilungen und Eigenschaften der in der Modellierung verwendeten Einflussgrößen. Tabelle 4 fasst die zentralen deskriptiven Statistiken der Variablen zusammen, darunter Minimum, erstes Quartil (Q1), Median, Mittelwert, drittes Quartil (Q3) und Maximum.

Tabelle 4: Deskriptive Statistiken der verwendeten Variablen

Variable	Min	Q1	Median	Mittelwert	Q3	Max
Atemfrequenz	13.4	14.4	14.7	14.7	15	18
Ruheherzfrequenzindex	1	77.8	88	81.5	99	100
Herzfrequenzvariabilitätsindex	1	72	81	74.7	88	100
Erholungsindex	1	68	87	79.5	100	100
Körpertemperaturindex	1	93.9	98	93.9	100	100
Vortagsaktivitätsindex	15	83	86.4	86.4	92	100
Durchschnittstemperatur	-1.29	-0.15	-0.03	-0.03	0.08	3.06
Sonnenscheindauer	0	100	200	150	300	500
Luftdruck	950	985	1010	1012	1020	1050

Atemfrequenzwerte bewegen sich in einem engen Bereich mit einem Median von 14,7 Atemzügen pro Minute und einem Maximum von 18. Diese geringe Streuung deutet auf eine homogene Verteilung dieser physiologischen Messgröße hin. Der Ruheherzfrequenzindex, der auf einer Skala von 0 bis 100 bewertet wird, weist einen Median von 88 auf. Der Herzfrequenzvariabilitätsindex hingegen zeigt eine größere Streuung mit einem Median von 81.

Die Verteilung des Erholungsindex hebt sich mit durchgehend hohen Werte hervor, wobei der Median bei 87 liegt und die Werte bis zu einer oberen Grenze von 100 reichen. Der

Körpertemperaturindex zeigt eine hohe Stabilität mit einem Median von 98 und einer engen Verteilung um den Mittelwert von 93,9.

Der Vortagsaktivitätsindex zeigt Werte bis zu 100, wobei der Median bei 86,4 liegt. Die Durchschnittstemperatur variiert stärker und zeigt eine nahezu symmetrische Verteilung mit einem Median von $-0,03^{\circ}\text{C}$. Sonnenscheindauer und Luftdruck decken ebenfalls eine breite Spannweite ab. Die Sonnenscheindauer reicht von 0 bis 500 Minuten, mit einem Median von 200 Minuten, während der Luftdruck zwischen 950 und 1050 hPa schwankt. Zur weiteren deskriptiven Analyse bieten die in Abbildung 2 dargestellten Scatterplots Einblicke in die Zusammenhänge zwischen den Einflussgrößen und dem Schlafwert. Die Werte der Einflussgrößen sind auf der x-Achse dargestellt, während die y-Achse den Schlafwert abbildet. Jede Variable wird in einer separaten Facette dargestellt, ergänzt durch eine lineare Trendlinie, die mögliche lineare Zusammenhänge visualisiert.

Die Atemfrequenz offenbart eine sichtbare starke Tendenz, dass höhere Atemfrequenzen mit niedrigeren Schlafwerten assoziiert sind. Als nächstes zeigt sich ein leicht positiver Zusammenhang mit zunehmendem Datum, wobei eine Lücke in den Daten auf fehlende Schlafwerte in diesem Zeitraum hinweist. Die Durchschnittstemperatur zeigt subtile Trends, die auf potenziell störende Effekte höherer Temperaturen auf die Schlafqualität hinweisen könnten. Der Erholungsindex hebt sich dagegen durch eine klare positive Beziehung zum Schlafwert hervor. Ähnlich deutet der Herzfrequenzvariabilitätsindex auf eine positive Korrelation mit der Schlafqualität hin.

Der Körpertemperaturindex zeigt einen positiven Trend in Bezug zum Schlafwert, wobei einige Ausreißer erkennbar sind. Der Ruheherzfrequenzindex zeigt eine leicht positive Tendenz, die darauf hinweist, dass niedrigere Ruheherzfrequenzen mit besseren Schlafwerten korrelieren könnten. Bei der Sonnenscheindauer wird ein schwacher negativer Zusammenhang sichtbar, der auf mögliche Nachteile einer intensiven Lichtexposition für den Schlaf hinweist.

Der Vortagsaktivitätsindex zeigt eine ähnliche Tendenz, wobei höhere Aktivitätslevel des Vortags tendenziell mit besseren Schlafwerten assoziiert sein könnten. Der Luftdruck scheint in diesem Zusammenhang eine geringere Rolle zu spielen, wobei sich die Werte breit streuen.

Zusammenfassend zeigt die deskriptive Analyse, dass sowohl physiologische als auch Umweltfaktoren in unterschiedlichem Maße mit der Schlafqualität in Verbindung stehen. Besonders die Atemfrequenz, der Erholungsindex, der Ruheherzfrequenzindex und der Herzfrequenzvariabilitätsindex erweisen sich als zentrale Prädiktoren, während andere Faktoren wie Luftdruck und das Datum womöglich keinen starken Einfluss haben könnten. Diese Ergebnisse bilden eine fundierte Grundlage für die weitere Modellierung und Analyse.

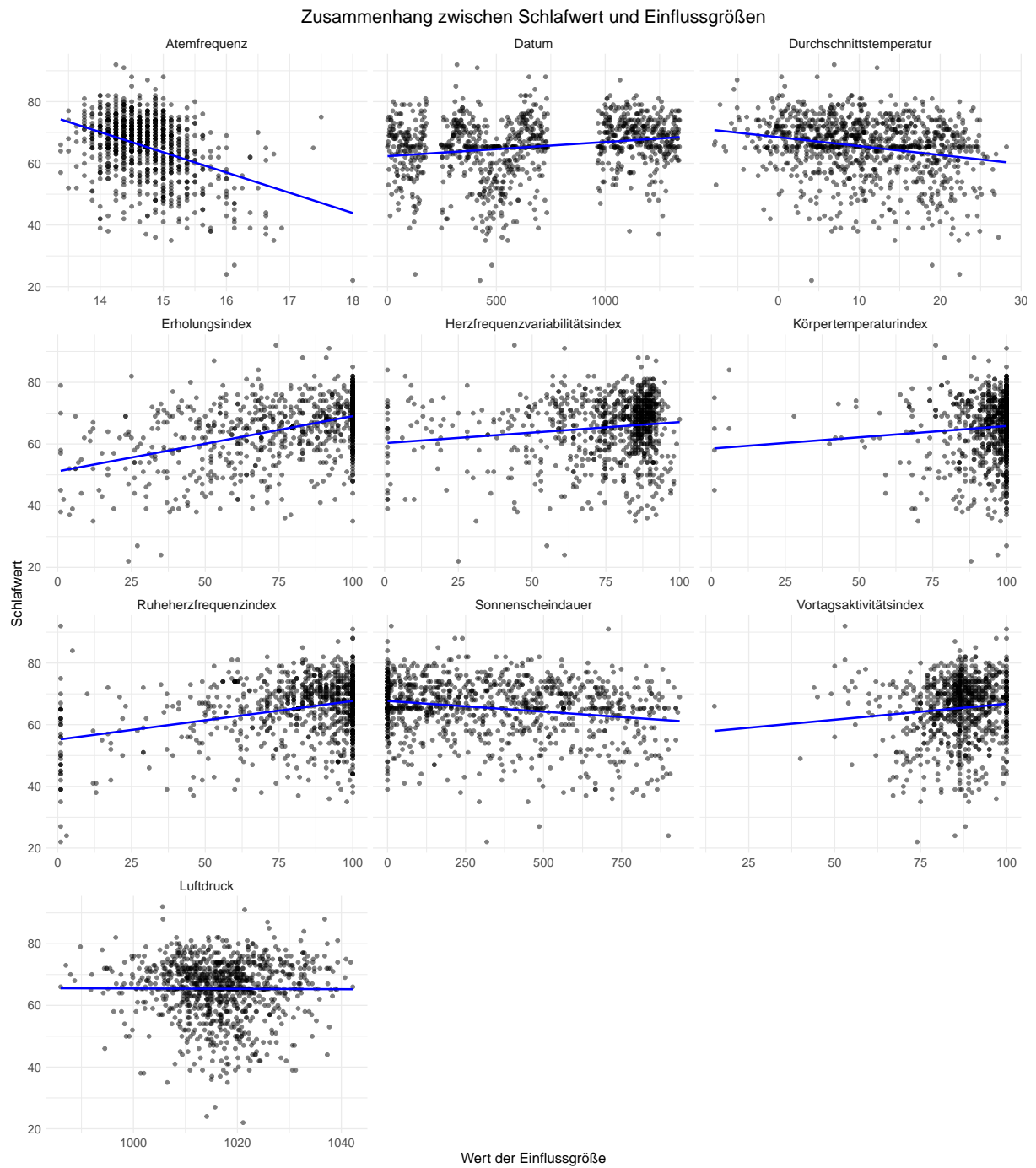


Abbildung 2: Zusammenhang zwischen Schlafwert und Einflussgrößen.

4 Modellierung direkter Einflussfaktoren auf die Schlafqualität

Die Modellierung in dieser Arbeit hatte zwei zentrale Ziele: Einerseits sollte die Schlafqualität basierend auf dem Schlafwert möglichst präzise vorhergesagt werden, andererseits sollten alternative Zielgrößen analysiert werden, um indirekte Einflüsse auf die Schlafqualität besser zu verstehen. Diese Vorgehensweise ermöglicht sowohl die direkte Modellierung der Schlafqualität als auch die Untersuchung physiologischer Mechanismen, die mit der Schlafqualität in Zusammenhang stehen, um indirekte Einflüsse zu verstehen. Zuerst wurde die Modellierung direkter Einflussfaktoren auf die Schlafqualität betrachtet. Im Zentrum stand dabei die systematische Auswahl relevanter Prädiktoren und die Modellierung potenzieller Zusammenhänge. Die Vorgehensweise gliederte sich in mehrere zentrale Schritte: Variablenselektion, die Bildung linearer und nichtlinearer Modelle und die Evaluierung dieser Modelle.

4.1 Variablenselektion

Die Variablenselektion ist ein zentraler Schritt in der Modellierung, da sie die Grundlage für die Auswahl relevanter Prädiktoren und die Vermeidung redundanter oder irrelevanter Variablen legt. Sie reduziert die Komplexität der Modelle und trägt zur besseren Interpretierbarkeit der Ergebnisse bei. In dieser Arbeit wurde ein systematischer Ansatz verfolgt, der sowohl statistische als auch inhaltliche Kriterien berücksichtigt.

4.1.1 Inhaltliche Überlegungen

Im ersten Schritt wurde der Schlafwert als Zielgröße festgelegt, da er eine aggregierte Metrik darstellt, die die allgemeine Schlafqualität umfassend beschreibt. Der Schlafwert des Smart Rings basiert auf einer gewichteten Kombination verschiedener Schlafmerkmale, darunter die Gesamtschlafdauer, die Effizienz des Schlafs, der Anteil an Tiefschlaf und REM-Schlaf sowie die Konsistenz des Schlafverhaltens. Diese Merkmale werden kombiniert, um einen ganzheitlichen Indikator der Schlafqualität zu liefern, der es ermöglicht, komplexe Zusammenhänge zwischen verschiedenen Faktoren zu erfassen.

Der Schlafwert weist somit hohe Korrelationen mit anderen spezifischen Schlafmetriken der Smart-Ring Daten auf, wie beispielsweise der Gesamtschlafdauer ($r = 0,812$), dem Gesamtschlafscore ($r = 0,831$) und dem Tiefschlafscore ($r = 0,634$). Diese hohen Korrelationen verdeutlichen, dass der Schlafwert als Zielgröße geeignet ist, die allgemeine Schlafqualität umfassend abzubilden.

Nach der Festlegung der Zielgröße wurden alle Variablen ausgeschlossen, die direkt den Schlaf erklären und somit keinen zusätzlichen Informationsgewinn für die Interpretation der Modelle liefern würden. Dazu zählen Variablen wie der Anteil an REM-Schlaf oder die Latenz, die bereits explizit in die Berechnung des Schlafwerts eingehen. So wurden Variablen ausgeschlossen, die eng mit der Zielgröße verknüpft sind oder aus ihr abgeleitet wurden.

Nach dem Ausschluss redundant erklärender Variablen wurde eine inhaltliche Bewertung der verbleibenden potenziellen Prädiktoren vorgenommen. Ziel war es, Variablen auszuwählen,

die mögliche neue Erkenntnisse liefern könnten, anstatt bereits etablierte Zusammenhänge zu reproduzieren. Beispielsweise wurden Variablen wie die Zeit, die im Bett verbracht wird, die Aufwachzeit und die Bettgezeit ausgeschlossen. Diese Variablen sind gut dokumentierte Einflussgrößen, deren Informationsgehalt keinen zusätzlichen Mehrwert für die Analyse bietet. Stattdessen wurde der Fokus auf weniger offensichtliche Einflussfaktoren gelegt, die neue wissenschaftliche Erkenntnisse fördern könnten.

4.1.2 Korrelationsanalyse

Nach der inhaltlichen Bewertung der Variablen wurde eine Korrelationsanalyse durchgeführt, um Variablen mit mittelstarken bis hohen Zusammenhängen mit dem Schlafwert zu identifizieren. Ziel dieser Analyse war es, jene Variablen zu selektieren, die sowohl einen statistisch relevanten Zusammenhang mit dem Schlafwert aufweisen als auch einen inhaltlichen Mehrwert bieten.

In Abbildung 8 werden die Korrelationen der ausgewählten Variablen mit dem Schlafwert dargestellt. Die dargestellten Variablen repräsentieren jene Prädiktoren, die nach der Selektion sowohl die höchsten Korrelationskoeffizienten mit dem Schlafwert aufweisen als auch durch ihren inhaltlichen Mehrwert als potenziell wichtige Einflussgrößen identifiziert wurden. Variablen mit sehr geringen oder insignifikanten Korrelationen wurden ausgeschlossen, da sie keinen substantiellen Beitrag zur Modellierung leisten würden. Beispiele hierfür sind die Mondphasen sowie bestimmte Wetterdaten wie der Niederschlag oder der Schneefall, die in vorläufigen Analysen keine relevanten Zusammenhänge mit der Schlafqualität gezeigt hatten.

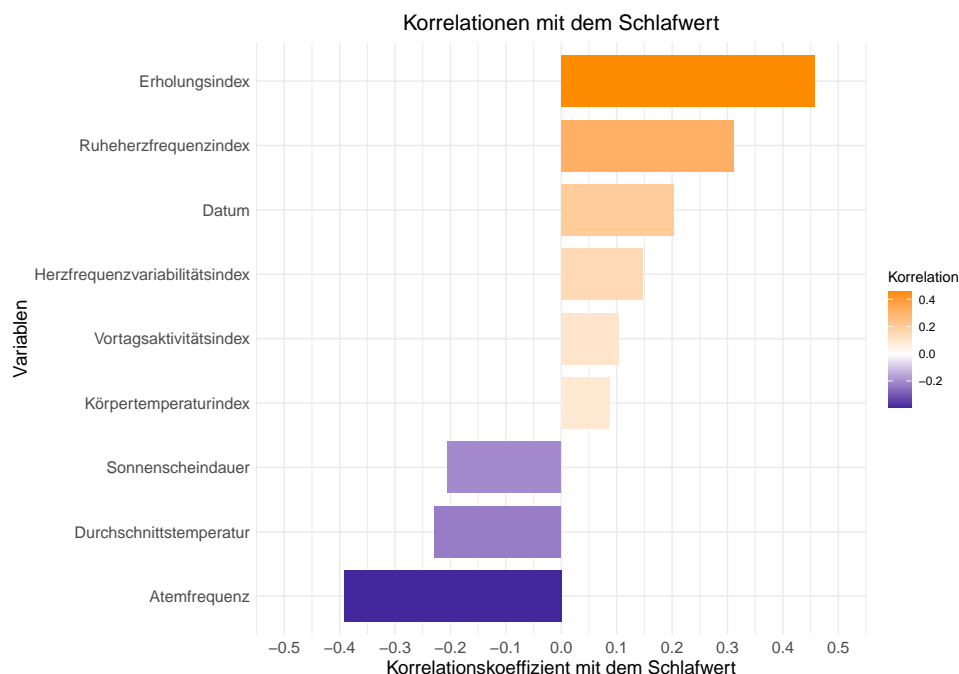


Abbildung 3: Korrelationen der ausgewählten Variablen mit dem Schlafwert

Die Abbildung zeigt die Korrelationskoeffizienten der ausgewählten Variablen mit dem Schlafwert. Die höchste positive Korrelation wurde für den Erholungsindex beobachtet,

dessen Korrelationskoeffizient bei etwa $r = 0.45$ liegt. Dieser Wert zeigt, dass eine schnelle Stabilisierung der Ruheherzfrequenz nach dem Einschlafen tendenziell mit einer höheren Schlafqualität einhergeht. Der Erholungsindex reflektiert die physiologische Anpassung an die Ruhephase und dient somit als zentraler Indikator für die nächtliche Regeneration. Seine Relevanz ergibt sich sowohl aus der hohen Korrelation mit dem Schlafwert als auch aus seiner Rolle als Maß für die nächtliche Erholungsfähigkeit.

Der Ruheherzfrequenzindex zeigt mit ca. $r \approx 0.30$ ebenfalls eine positive Korrelation mit dem Schlafwert. Eine stabile und niedrige Ruheherzfrequenz gilt als Zeichen für eine gute körperliche Erholung während der Nacht, was ihre inhaltliche Relevanz für die Analyse der Schlafqualität unterstreicht. Der Herzfrequenzvariabilitätsindex zeigt eine Korrelation von etwa $r = 0.15$ und wird aufgrund seiner Bedeutung als Maß für die Balance des autonomen Nervensystems in die Analyse aufgenommen. Die Fähigkeit, in der Nacht eine hohe Herzfrequenzvariabilität zu erreichen, wird häufig mit einem gesunden physiologischen Zustand assoziiert.

Der Vortagsaktivitätsindex weist eine Korrelation von etwa $r = 0.10$ auf, was darauf hindeutet, dass körperliche Betätigung am Tag vor dem Schlaf möglicherweise positive Effekte auf die Schlafqualität haben kann. Der Körpertemperaturindex zeigt eine geringe positive Korrelation mit knapp $r \approx -0.10$, was darauf hinweist, dass geringe Abweichungen in der Körpertemperatur von der Norm die Schlafqualität verbessern könnten.

Im negativen Bereich zeigt die nächtliche Atemfrequenz mit $r \approx -0.40$ die stärkste Korrelation. Eine erhöhte Atemfrequenz kann ein Indikator für Stress, gesundheitliche Probleme oder einen unruhigen Schlaf sein, was ihre starke negative Beziehung zur Schlafqualität erklärt. Ähnlich negativ korreliert die durchschnittliche Temperatur mit dem Schlafwert ($r \approx -0.30$), was darauf hinweist, dass höhere Temperaturen tendenziell mit einer geringeren Schlafqualität assoziiert sind. Die Sonnenscheindauer, mit einer Korrelation von etwa $r \approx -0.10$, reflektiert ebenfalls potenziell störende Umwelteinflüsse, die die nächtliche Regeneration beeinträchtigen könnten. Der Luftdruck zeigt eine geringe Korrelation mit $r \approx -0.05$, was auf einen minimalen Einfluss dieser Variablen auf die Schlafqualität schließen lässt.

Zusammenfassend zeigt die Abbildung, dass die ausgewählten Variablen sowohl statistisch als auch inhaltlich relevant sind. Die Korrelationsanalyse diente dazu, die Variablen mit den stärksten Beziehungen zum Schlafwert zu identifizieren, während Variablen mit geringen oder inkonsistenten Zusammenhängen ausgeschlossen wurden. Die verbleibenden Variablen bieten eine fundierte Grundlage für die weitere Modellierung und ermöglichen es, sowohl physiologische als auch externe Einflüsse auf die Schlafqualität zu analysieren. Diese Variablen spiegeln physiologische oder umweltbedingte Mechanismen wider, die einen zusätzlichen Erklärungswert liefern könnten und die Analyse der Schlafqualität erweitern.

4.1.3 Prüfung auf Multikollinearität

Um Multikollinearität zwischen den unabhängigen Variablen zu vermeiden, wurde der Variance-Inflation-Faktor (VIF) berechnet. Multikollinearität tritt auf, wenn zwei oder mehr unabhängige Variablen in einem Modell stark korrelieren, was die Stabilität der Schätzung der Regressionskoeffizienten negativ beeinflussen kann. Der VIF gibt an, wie stark die Varianz eines geschätzten Regressionskoeffizienten aufgrund von Multikollinearität im Vergleich zu einer Situation ohne Multikollinearität erhöht ist.

Mathematisch wird der VIF für eine unabhängige Variable X_j wie folgt definiert:

$$\text{VIF}(X_j) = \frac{1}{1 - R_j^2},$$

wobei R_j^2 der Determinationskoeffizient eines Hilfsmodells ist, in dem X_j als abhängige Variable und alle anderen Prädiktoren des ursprünglichen Modells als unabhängige Variablen verwendet werden. Ein VIF-Wert von 1 bedeutet, dass keine Multikollinearität vorliegt, während Werte über 5 oder 10 (je nach Konvention) als Hinweis auf problematische Multikollinearität gelten.

Für die Berechnung des VIFs wurde ein lineares Modell erstellt, in dem der Schlafwert als Zielgröße und die ausgewählten Prädiktoren als unabhängige Variablen verwendet wurden. Die berechneten VIF-Werte sind in Tabelle 5 dargestellt.

Tabelle 5: Variance-Inflation-Faktor (VIF) der ausgewählten Prädiktoren

Variable	VIF-Wert
Datum	1.38
Ruheherzfrequenzindex	1.91
Atemfrequenz	1.51
Herzfrequenzvariabilitätsindex	1.48
Erholungsindex	1.39
Körpertemperaturindex	1.27
Vortagsaktivitätsindex	1.10
Durchschnittstemperaturindex	1.57
Sonnenscheindauer	1.43

Die Ergebnisse zeigen, dass alle Variablen VIF-Werte unterhalb des Schwellenwerts von 5 aufweisen, was darauf hinweist, dass keine kritische Multikollinearität zwischen den Prädiktoren vorliegt. Daher wurden keine Variablen aufgrund hoher VIF-Werte aus dem Modell entfernt.

4.1.4 Lasso-Regression

Die Lasso-Regression (Least Absolute Shrinkage and Selection Operator) ist eine Methode der Regularisierung, die in linearen Modellen eingesetzt wird, um die Anzahl der Prädiktoren zu reduzieren und die Interpretierbarkeit zu erhöhen. Dabei wird ein zusätzlicher Strafterm in die Optimierungsfunktion eingefügt, der die Koeffizienten irrelevanter oder schwach einflussreicher Prädiktoren auf null setzen kann. Dies ermöglicht eine automatische Variablenselektion, was insbesondere bei hochdimensionalen Datensätzen von Vorteil ist. Die Optimierungsfunktion der Lasso-Regression lautet:

$$\min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^n \left(y_i - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\},$$

wobei n die Anzahl der Beobachtungen, p die Anzahl der Prädiktoren, y_i die Zielvariable, x_{ij} der Wert des j -ten Prädiktors und β_j der Koeffizient des j -ten Prädiktors sind. Der

Regularisierungsparameter λ kontrolliert die Stärke der Bestrafung. Ein hoher Wert von λ führt dazu, dass mehr Koeffizienten auf null gesetzt werden, während bei $\lambda = 0$ die Lasso-Regression einem gewöhnlichen linearen Modell entspricht.

Für die vorliegenden Daten wurde die Lasso-Regression mit der Zielgröße Schlafwert durchgeführt. Die Prädiktoren umfassten die Variablen Datum, Ruheherzfrequenzindex, Atemfrequenz, Herzfrequenzvariabilitätsindex, Erholungsindex, Körpertemperaturindex, Vortagsaktivitätsindex, Durchschnittstemperatur und Sonnenscheindauer. Fehlende Werte in den Prädiktoren wurden durch den Mittelwert der jeweiligen Spalte ersetzt, um die Analyse zu ermöglichen.

Mithilfe einer Kreuzvalidierung wurde der optimale Wert für den Regularisierungsparameter ermittelt, welcher $\lambda_{\min} = 0.0102$ beträgt. Mit diesem Wert wurden die Koeffizienten der Prädiktoren geschätzt, die in Tabelle 6 dargestellt sind.

Tabelle 6: Koeffizienten der Lasso-Regression mit optimalem λ

Variable	Koeffizient
Achsenabschnitt	92.47
Datum	0.0037
Ruheherzfrequenzindex	0.0523
Atemfrequenz	-2.6608
Herzfrequenzvariabilitätsindex	-0.0044
Erholungsindex	0.1310
Körpertemperaturindex	-0.0685
Vortagsaktivitätsindex	0.0511
Durchschnittstemperatur	-0.3109
Sonnenscheindauer	-0.0007

Die Ergebnisse zeigen, dass einige Prädiktoren, wie die Atemfrequenz (-2.6608) und die Durchschnittstemperatur (-0.3109), einen vergleichsweise hohen absoluten Einfluss auf den Schlafwert aufweisen, während andere, wie der Herzfrequenzvariabilitätsindex (-0.0044), einen geringen Effekt haben. Variablen mit einem Koeffizienten von null wurden in der Lasso-Analyse nicht ausgeschlossen, sondern zeigen nur sehr geringe Einflüsse.

Die Lasso-Regression identifizierte auch positive Zusammenhänge, wie etwa für den Erholungsindex (0.1310) und den Vortagsaktivitätsindex (0.0511). Dies deutet darauf hin, dass höhere Werte dieser Variablen mit einer besseren Schlafqualität assoziiert sind. Negative Koeffizienten, wie bei der Atemfrequenz oder der Durchschnittstemperatur, deuten hingegen auf eine Verschlechterung der Schlafqualität bei steigenden Werten hin, was physiologisch plausibel ist. Beispielsweise ist eine erhöhte Atemfrequenz oft mit einer physiologischen Belastung verbunden, während höhere Temperaturen die Schlafqualität beeinträchtigen können.

Die Ergebnisse der Lasso-Analyse dienen primär als Orientierungshilfe, um die relevanten Einflussgrößen auf die Schlafqualität zu identifizieren. Obwohl Variablen wie die Herzfrequenzvariabilität einen geringen Einfluss haben, wurden sie aufgrund ihrer potenziellen Relevanz für weiterführende Analysen dennoch in die Modelle integriert.

Die finale Auswahl der Prädiktoren umfasste somit die Variablen numerisches Datum, Ruheherzfrequenzindex, Atemfrequenz, Erholungsindex, Herzfrequenzvariabilitätsindex,

Körpertemperaturindex, Vortagsaktivitätsindex, Durchschnittstemperatur und Sonnenscheindauer. Diese Variablen wurden sorgfältig ausgewählt, um sowohl die Interpretierbarkeit der Modelle zu gewährleisten als auch neue wissenschaftliche Erkenntnisse zu ermöglichen. Der Ausschluss redundanter und offensichtlicher Variablen stellt sicher, dass die Modelle auf unabhängigen und informativen Prädiktoren basieren.

4.2 Bayesianisches lineares Modell mit Zielgröße Schlafwert

Für die Modellierung der Zielgröße Schlafwert wurde zunächst ein bayesianisches lineares Modell ohne Interaktionen zwischen den Prädiktoren erstellt. Dieses Modell ermöglicht eine klare und isolierte Interpretation der Haupteffekte und dient als Ausgangspunkt für die Analyse potenzieller Interaktionen und nichtlinearer Zusammenhänge. Durch die Verwendung bayesianischer Statistik wird eine probabilistische Sichtweise auf die Modellparameter ermöglicht, die Unsicherheiten explizit quantifiziert und damit eine umfassendere Interpretation der Ergebnisse erlaubt.

4.2.1 Modellformel

Das verwendete Modell basiert auf einer linearen Regressionsstruktur, die mathematisch wie folgt dargestellt werden kann:

$$\begin{aligned} \text{Schlafwert}_i = & \beta_0 + \beta_1 \cdot \text{Ruheherzfrequenzindex}_i + \beta_2 \cdot \text{Atemfrequenz}_i \\ & + \beta_3 \cdot \text{Herzfrequenzvariabilitätsindex}_i + \beta_4 \cdot \text{Erholungsindex}_i \\ & + \beta_5 \cdot \text{Vortagsaktivitätsindex}_i + \beta_6 \cdot \text{Durchschnittstemperatur}_i \\ & + \beta_7 \cdot \text{Datum}_i + \beta_8 \cdot \text{Sonnenscheindauer}_i \\ & + \beta_9 \cdot \text{Körpertemperaturindex}_i + \epsilon_i, \end{aligned}$$

wobei die Koeffizienten $\beta_0, \beta_1, \dots, \beta_9$ die Effekte der jeweiligen Prädiktoren auf den Schlafwert repräsentieren und ϵ_i den normalverteilten Fehlerterm beschreibt, der mit $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ modelliert wurde. Der Achsenabschnitt β_0 gibt den erwarteten Mittelwert des Schlafwerts an, wenn alle Prädiktoren gleich null sind. Die Prädiktoren wurden auf Basis der vorangegangenen Variablenselektion ausgewählt und umfassen sowohl physiologische als auch externe Einflussfaktoren.

4.2.2 Priorverteilungen und Schätzung

Für die Modellparameter wurden konservative Priorverteilungen definiert, um eine Überanpassung der Daten zu vermeiden, gleichzeitig jedoch ausreichend Flexibilität für die Schätzung starker Effekte zu gewährleisten. Die Koeffizienten der Prädiktoren wurden mit einer normalverteilten Prior von $\mathcal{N}(0, 1)$ belegt, was die Annahme widerspiegelt, dass die Effekte der meisten Variablen moderat ausfallen. Der Achsenabschnitt erhielt eine breitere Prior von $\mathcal{N}(0, 5)$, um Unsicherheiten in der Grundlinie des Schlafwerts zu berücksichtigen. Diese Wahl der Priorverteilungen erlaubt es, Effekte zu quantifizieren, ohne dass starke Vorannahmen die Ergebnisse dominieren.

Die Schätzung der Modellparameter erfolgte unter Verwendung der probabilistischen Programmiersprache Stan, implementiert über die R-Bibliothek brms. Der Hamiltonian Monte Carlo (HMC)-Sampler wurde eingesetzt, da er für hochdimensionale Probleme besonders effizient ist und zuverlässige Konvergenzergebnisse liefert. Zur Sicherstellung einer stabilen Anpassung wurde der Kontrollparameter `adapt_delta` auf 0,95 gesetzt. Die Modellanpassung erfolgte mit vier Markov-Ketten, die jeweils 2000 Iterationen umfassen, von denen die ersten 1000 als Burn-in ausgeschlossen wurden. Dies führte zu 4000 Nachbrenndurchgängen, die zur Schätzung der Posterior-Verteilungen verwendet wurden. Alle \hat{R} -Werte lagen bei 1,00, was eine vollständige Konvergenz der Markov-Ketten bestätigt. Die Posterior-Verteilungen der Parameter lieferten detaillierte Informationen über die Stärke und Richtung der Effekte sowie deren Unsicherheiten. Diese Modellstruktur stellt eine solide Grundlage dar, um die Zusammenhänge zwischen den Prädiktoren und der Zielgröße zu interpretieren und dient als Ausgangspunkt für weiterführende Analysen.

4.2.3 Ergebnisse und Interpretation

Das bayesianische lineare Modell zur Vorhersage der Zielgröße Schlafwert liefert detaillierte Schätzungen der Effekte der ausgewählten Prädiktoren. Die Ergebnisse der Modellschätzung, einschließlich der geschätzten Koeffizienten, zugehörigen Standardfehler sowie der 95%-Konfidenzintervalle, sind in Tabelle 7 dargestellt. Diese Parameter erlauben eine fundierte Analyse der Haupteffekte, die die Schlafqualität beeinflussen.

Tabelle 7: Ergebnisse des bayesianischen linearen Modells für den Schlafwert

Prädiktor	Schätzwert	Std. Fehler	95% KI (unten, oben)
Achsenabschnitt	85.32	8.12	(69.78, 101.02)
Ruheherzfrequenzindex	0.07	0.01	(0.04, 0.09)
Atemfrequenz	-2.08	0.46	(-3.01, -1.15)
Herzfrequenzvariabilitätsindex	-0.02	0.01	(-0.05, 0.01)
Erholungsindex	0.14	0.01	(0.12, 0.16)
Vortagsaktivitätsindex	0.03	0.03	(-0.01, 0.09)
Durchschnittstemperatur	-0.28	0.04	(-0.36, -0.20)
Datum	0.00	0.00	(0.00, 0.00)
Sonnenscheindauer	0.00	0.00	(-0.00, 0.00)
Körpertemperaturindex	-0.05	0.02	(-0.10, -0.01)

Die Schätzung des Achsenabschnitts zeigt einen Wert von 85.32 (95%-KI: 69.78, 101.02). Dies entspricht dem erwarteten Schlafwert, wenn alle Prädiktoren gleich null sind. Der Ruheherzfrequenzindex weist einen signifikant positiven Effekt auf den Schlafwert auf, mit einem geschätzten Koeffizienten von 0.07 (Standardfehler: 0.01, 95%-KI: 0.04, 0.09). Diese Ergebnisse verdeutlichen, dass höhere Werte des Ruheherzfrequenzindex, die eine stabile und niedrige Herzfrequenz während der Nacht anzeigen, mit einer verbesserten Schlafqualität assoziiert sind.

Die Atemfrequenz zeigt den stärksten negativen Einfluss auf den Schlafwert. Mit einem geschätzten Koeffizienten von -2.08 (Standardfehler: 0.46, 95%-KI: -3.01, -1.15) wird deutlich, dass eine erhöhte Atemfrequenz während des Schlafs einen negativen Effekt auf

die Schlafqualität hat. Dies lässt sich durch physiologische Belastungen oder mögliche Atemstörungen erklären, die mit einer erhöhten Atemfrequenz einhergehen.

Der Erholungsindex zeigt einen klaren positiven und signifikanten Effekt auf die Schlafqualität, mit einem geschätzten Koeffizienten von 0.14 (Standardfehler: 0.01, 95%-KI: 0.12, 0.16). Eine bessere nächtliche Regeneration, die durch höhere Werte im Erholungsindex angezeigt wird, wirkt sich positiv auf den Schlafwert aus. Die Ergebnisse unterstreichen die Relevanz dieser physiologischen Erholungsmetrik für die Analyse der Schlafqualität.

Der Herzfrequenzvariabilitätsindex zeigt mit einem geschätzten Wert von -0.02 (Standardfehler: 0.01, 95%-KI: -0.05, 0.01) keinen signifikanten Einfluss auf die Schlafqualität. Dies deutet darauf hin, dass die Herzfrequenzvariabilität in diesem Modell keine entscheidende Rolle für die Vorhersage des Schlafwerts spielt. Ähnlich zeigt auch der Vortagsaktivitätsindex keinen signifikanten Effekt, obwohl der geschätzte Koeffizient von 0.03 (Standardfehler: 0.03, 95%-KI: -0.01, 0.09) einen leichten positiven Trend vermuten lässt.

Die Durchschnittstemperatur weist mit einem Koeffizienten von -0.28 (Standardfehler: 0.04, 95%-KI: -0.36, -0.20) einen moderat negativen Einfluss auf die Schlafqualität auf. Diese Variable verdeutlicht, dass höhere Umgebungstemperaturen während der Nacht die Schlafqualität beeinträchtigen können, möglicherweise durch Störungen der Thermoregulation.

Die Sonnenscheindauer und der Körpertemperaturindex zeigen keine signifikanten Effekte auf den Schlafwert. Während die Sonnenscheindauer einen Koeffizienten von nahezu null aufweist, deutet der Körpertemperaturindex mit einem Koeffizienten von -0.05 (Standardfehler: 0.02, 95%-KI: -0.10, -0.01) auf einen schwachen negativen Effekt hin, was jedoch inhaltlich plausibel ist, da Temperaturabweichungen den Schlaf stören können.

Die geschätzte Standardabweichung des Fehlerterms (σ) liegt bei 7.64 und deutet auf eine moderate Streuung der Residuen hin. Diese Ergebnisse liefern wertvolle Einblicke in die Determinanten der Schlafqualität, wobei physiologische Prädiktoren wie der Erholungsindex und die Atemfrequenz besonders hervorgehoben werden. Externe Einflussfaktoren wie die Temperatur zeigen ebenfalls Relevanz und könnten in weiterführenden Modellen detaillierter untersucht werden.

4.2.4 Erweiterung des Modells um Interaktionen

Ein weiterer Schritt in der Modellierung bestand darin, mögliche Interaktionen zwischen den Prädiktoren zu untersuchen. Interaktionen können auftreten, wenn der Effekt eines Prädiktors auf die Zielgröße von einem anderen Prädiktor abhängt. Zur Identifikation relevanter Interaktionseffekte wurde in dieser Arbeit die Horseshoe-Regularisierung verwendet, ein Ansatz, der besonders geeignet ist, um in hochdimensionalen Daten wichtige Effekte von irrelevanten zu unterscheiden.

Die Horseshoe-Regularisierung ist eine bayesianische Methode, die sparsames Lernen ermöglicht, indem sie eine priore Verteilung verwendet, die extreme Werte für wichtige Parameter begünstigt, während irrelevante Parameter nahe null gesetzt werden. Diese Eigenschaft ist besonders vorteilhaft bei der Untersuchung von Interaktionen, da die Anzahl möglicher Kombinationen in einem Modell mit vielen Prädiktoren schnell ansteigt, was die Modellkomplexität erhöht und die Interpretierbarkeit erschwert.

In diesem Kontext wurde für jedes Prädiktorpaar eine Interaktion in das Modell aufgenommen und mittels Horseshoe-Regularisierung bewertet. Die Ergebnisse zeigten jedoch,

dass sämtliche geschätzten Interaktionsterme nahe null lagen, was darauf hinweist, dass keine signifikanten Interaktionen zwischen den Prädiktoren vorlagen. Die Horseshoe-Regularisierung hat somit geholfen, die Modellkomplexität zu reduzieren und gleichzeitig sicherzustellen, dass potenziell relevante Effekte nicht übersehen wurden.

Dieser Befund unterstützt die Annahme, dass die Hauptprädiktoren unabhängig voneinander zur Erklärung der Zielgröße Schlafwert beitragen und keine wechselseitige Verstärkung oder Abschwächung der Effekte vorliegt. Daher wurde beschlossen, in den weiteren Modellierungsansätzen keine Interaktionsterme zu berücksichtigen und sich stattdessen auf die Hauptprädiktoren sowie potenzielle nichtlineare Zusammenhänge zu konzentrieren.

4.3 Bayesianisches Modell mit Spline für die Atemfrequenz

Nach der Analyse des linearen Modells wurde untersucht, ob nichtlineare Beziehungen zwischen der Zielgröße Schlafwert und den Prädiktoren bestehen. Hierbei wurden zwei nichtlineare Ansätze betrachtet: Ein Modell, das einen Spline für die Atemfrequenz enthält, und ein Modell, das einen polynomiellen Term zweiter Ordnung für denselben Prädiktor integriert. Ziel dieser Ansätze war es, potenzielle nichtlineare Effekte genauer zu modellieren und ihre Bedeutung für die Schlafqualität zu evaluieren.

Die Entscheidung, spezifisch die Atemfrequenz in einem nichtlinearen Kontext zu modellieren, basiert auf mehreren evidenzbasierten Aspekten. Erstens zeigte die Atemfrequenz im linearen Modell den stärksten negativen Einfluss auf den Schlafwert mit einem Koeffizienten von -2.10 . Dieser deutliche Effekt verdeutlicht die zentrale Rolle der Atemfrequenz bei der Vorhersage der Schlafqualität. Zweitens wies die deskriptive Analyse auf eine komplexe Beziehung zwischen der Atemfrequenz und dem Schlafwert hin, die möglicherweise nicht vollständig durch ein einfaches lineares Modell erfasst werden kann. Insbesondere die Scatterplots in der deskriptiven Analyse deuteten auf einen potenziellen nichtlinearen Zusammenhang hin, da die Datenpunkte keinen klaren linearen Trend, sondern eine stärkere Streuung und Abweichungen in bestimmten Frequenzbereichen zeigten.

Durch die Verwendung eines Splines können subtile Übergänge und lokale Trends in der Beziehung zwischen der Atemfrequenz und dem Schlafwert modelliert werden. Der polynomielle Term zweiter Ordnung hingegen erlaubt die Erfassung symmetrischer nichtlinearer Muster, wie z. B. einer U- oder umgekehrten U-Form. Diese Ansätze ermöglichen eine differenzierte Betrachtung und erweitern die Analyse, indem sie den komplexen Zusammenhang zwischen der Atemfrequenz und der Schlafqualität genauer abbilden. Somit wird eine fundierte Basis geschaffen, um die Bedeutung der Atemfrequenz in der Modellierung der Schlafqualität sowohl statistisch als auch inhaltlich besser zu verstehen.

4.3.1 Modellformel

Ein Spline ist ein mathematisches Werkzeug, das eine Funktion in Stücke unterteilt und diese Segmente mit polynomalen Funktionen beschreibt, die an den sogenannten Knotenpunkten stetig miteinander verbunden sind. Diese Methode erlaubt es, den Wertebereich einer Variable flexibel zu modellieren, ohne die starre Form einer linearen Beziehung oder eines festen polynomiellen Terms vorzugeben. Der kubische Spline, der hier verwendet wurde, nutzt drittgradige Polynome innerhalb der Segmente, wodurch die Funktion sowohl glatt als auch flexibel bleibt. Der Hyperparameter des Splines regelt dabei die Glätte

der Kurve: Ein niedriger Wert begünstigt eine hohe Anpassung an lokale Datenmuster, während ein hoher Wert die Kurve glättet und damit Überanpassung verhindert. Diese Flexibilität macht Splines besonders geeignet für die Modellierung komplexer Beziehungen, wie sie bei der Atemfrequenz und deren Auswirkungen auf die Schlafqualität vermutet werden.

Im vorliegenden Fall wurde ein kubischer Spline für die Variable Atemfrequenz verwendet, um deren potenziell komplexe Beziehung mit dem Schlafwert zu erfassen. Die vollständige Modellformel lautet:

$$\begin{aligned} \text{Schlafwert}_i = & \beta_0 + f(\text{Atemfrequenz}_i, k = 5) + \beta_1 \cdot \text{Ruheherzfrequenzindex}_i \\ & + \beta_2 \cdot \text{Herzfrequenzvariabilitätsindex}_i + \beta_3 \cdot \text{Erholungsindex}_i \\ & + \beta_4 \cdot \text{Vortagsaktivitätsindex}_i + \beta_5 \cdot \text{Durchschnittstemperatur}_i \\ & + \beta_6 \cdot \text{Datum}_i + \beta_7 \cdot \text{Sonnenscheindauer}_i \\ & + \beta_8 \cdot \text{Körpertemperaturindex}_i + \epsilon_i, \end{aligned}$$

wobei $f(\text{Atemfrequenz}_i)$ die durch den Spline modellierte Funktion darstellt und $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ den normalverteilten Fehlerterm beschreibt.

Die Modellanpassung erfolgte wie für das lineare Modell mit der probabilistischen Programmiersprache Stan, implementiert über die R-Bibliothek **brms**. Der Hamiltonian Monte Carlo-Sampler wurde erneut verwendet, da er für hochdimensionale Probleme effizient und zuverlässig ist. Die gleichen Kontrollparameter wie im bayesianischen linearen Modell wurden verwendet, um Konvergenzprobleme zu vermeiden.

4.3.2 Ergebnisse und Interpretation

Die Ergebnisse des bayesianischen Modells mit einem Spline-Term für die Atemfrequenz sind in Tabelle 8 dargestellt. Der Intercept beträgt 56.27 (95%-KI: 49.91, 62.55) und repräsentiert den geschätzten durchschnittlichen Schlafwert, wenn alle Prädiktoren auf ihrem Referenzwert liegen. Die Ruheherzfrequenz zeigt einen positiven Effekt (0.04, 95%-KI: 0.02, 0.07) mit einem Standardfehler von 0.01, was darauf hinweist, dass eine Erhöhung des Ruheherzfrequenzindex mit einer Verbesserung des Schlafwerts assoziiert ist. Dies könnte darauf zurückzuführen sein, dass höhere Werte des Index auf eine bessere Regeneration und Anpassung an den Schlaf hinweisen.

Die Herzfrequenzvariabilität hat einen minimal negativen Effekt (−0.01, 95%-KI: −0.03, 0.02) mit einem Standardfehler von 0.01, der jedoch statistisch nicht signifikant ist. Der Erholungsindex weist hingegen mit 0.13 (95%-KI: 0.11, 0.15) und einem Standardfehler von 0.01 einen starken und signifikanten positiven Einfluss auf die Schlafqualität auf. Dieser Befund unterstreicht, dass eine effiziente nächtliche Erholung maßgeblich zur Verbesserung der Schlafqualität beiträgt.

Die Vortagsaktivität zeigt einen positiven, aber statistisch nicht signifikanten Effekt (0.05, 95%-KI: −0.01, 0.10) mit einem Standardfehler von 0.03. Dies deutet darauf hin, dass die Aktivität des Vortages in diesem Modell keine starke Rolle spielt. Die durchschnittliche Temperatur hat mit −0.29 (95%-KI: −0.37, −0.22) und einem Standardfehler von 0.04 einen signifikant negativen Einfluss auf den Schlafwert, was darauf hinweist, dass höhere Temperaturen die Schlafqualität beeinträchtigen können.

Die Zeitvariable Datum zeigt einen sehr kleinen Effekt (0.00, 95%-KI: 0.00, 0.00) mit einem Standardfehler von 0.00, der zwar statistisch signifikant ist, jedoch praktisch kaum relevant erscheint. Ähnlich verhält es sich bei der Sonnenscheindauer, deren Effekt ebenfalls statistisch unbedeutend ist (−0.00, 95%-KI: −0.00, 0.00) mit einem Standardfehler von 0.00. Der Körpertemperaturindex zeigt mit −0.08 (95%-KI: −0.13, −0.03) und einem Standardfehler von 0.02 einen signifikanten negativen Effekt, was nahelegt, dass größere Abweichungen der Körpertemperatur vom Normwert mit einer schlechteren Schlafqualität einhergehen können.

Der Spline-Term für die Atemfrequenz liefert eine flexible Schätzung der Beziehung zwischen Atemfrequenz und Schlafwert. Der geschätzte Hyperparameter des Splines beträgt 16.08 (95%-KI: 7.46, 34.01), was auf eine hohe Anpassungsfähigkeit des Modells hindeutet. Die Ergebnisse des Splines (−0.08, 95%-KI: −2.05, 1.89) mit einem Standardfehler von 1.03 deuten darauf hin, dass die Beziehung zwischen Atemfrequenz und Schlafwert potenziell nicht linear ist, jedoch keine starken Hinweise auf signifikante nichtlineare Effekte vorliegen.

Tabelle 8: Ergebnisse des bayesianischen Modells mit Spline für Atemfrequenz

Prädiktor	Schätzwert	Std. Fehler	95% KI (unten, oben)
Intercept	56.27	3.25	(49.91, 62.55)
Ruheherzfrequenzindex	0.04	0.01	(0.02, 0.07)
Herzfrequenzvariabilitätsindex	-0.01	0.01	(-0.03, 0.02)
Erholungsindex	0.13	0.01	(0.11, 0.15)
Vortagsaktivitätsindex	0.05	0.03	(-0.01, 0.10)
Durchschnittstemperatur	-0.29	0.04	(-0.37, -0.22)
Datum	0.00	0.00	(0.00, 0.00)
Sonnenscheindauer	-0.00	0.00	(-0.00, 0.00)
Körpertemperaturindex	-0.08	0.02	(-0.13, -0.03)
Spline für Atemfrequenz	-0.08	1.03	(-2.05, 1.89)

Die Konvergenz der Markov-Ketten wurde durch \hat{R} -Werte von 1.00 für alle Parameter bestätigt, was auf eine zuverlässige Schätzung der Posterior-Verteilungen hinweist. Die geschätzte Standardabweichung des Fehlerterms ($\sigma = 7.60$, 95%-KI: 7.29, 7.93) zeigt, dass das Modell die Streuung der Daten gut abbildet.

Insgesamt bestätigen die Ergebnisse des Modells, dass einige Prädiktoren wie der Erholungsindex und die Ruheherzfrequenz signifikant mit der Schlafqualität zusammenhängen, während andere Variablen wie die Herzfrequenzvariabilität und die Sonnenscheindauer keine starken Effekte aufzeigen. Die Flexibilität des Splines erlaubt es, nichtlineare Muster zu identifizieren, wobei die Atemfrequenz als wichtiger, potenziell komplexer Einflussfaktor auf die Schlafqualität identifiziert wurde.

4.4 Bayesianisches Modell mit Polynom für die Atemfrequenz

Nach der Analyse des linearen Modells und des Modells mit Spline für die Atemfrequenz wurde ein bayesianisches Modell entwickelt, das nichtlineare Effekte der Atemfrequenz durch die Integration eines quadratischen Terms berücksichtigt. Der polynomiale Ansatz ist eine etablierte Methode, um nichtlineare Zusammenhänge zu modellieren. Anders als

ein Spline, der hohe Flexibilität bietet und komplexe Muster abbilden kann, eignet sich ein polynomieller Ansatz insbesondere dann, wenn eine einfache, global glatte Beziehung zwischen der Zielvariablen und dem Prädiktor vermutet wird. Durch die Hinzufügung des quadratischen Terms kann das Modell sowohl monotone als auch U-förmige oder umgekehrte U-förmige Zusammenhänge abbilden.

Der Vorteil eines polynomiellen Ansatzes liegt in seiner mathematischen Einfachheit und der leichten Interpretierbarkeit. Für die Atemfrequenz könnte beispielsweise erwartet werden, dass sowohl sehr niedrige als auch sehr hohe Werte nachteilig für die Schlafqualität sind, was in einer umgekehrten U-förmigen Beziehung resultieren würde. Um diese Hypothese zu testen, wurde ein bayesianisches Modell mit quadratischem Term für die Atemfrequenz erstellt.

4.4.1 Modellformel

Das bayesianische Modell mit polynominalem Ansatz erweitert die Modellierung, indem die Atemfrequenz als Prädiktor in einer zweiten Potenz (Quadrat) berücksichtigt wird. Diese Erweiterung erlaubt es, nichtlineare Effekte der Atemfrequenz auf den Schlafwert zu erfassen, ohne auf einen Spline zurückzugreifen. Die mathematische Modellformel lautet wie folgt:

$$\begin{aligned} \text{Schlafwert}_i = & \beta_0 + \beta_1 \cdot \text{Ruheherzfrequenzindex}_i + \beta_2 \cdot \text{Herzfrequenzvariabilitätsindex}_i \\ & + \beta_3 \cdot \text{Erholungsindex}_i + \beta_4 \cdot \text{Vortagsaktivitätsindex}_i \\ & + \beta_5 \cdot \text{Durchschnittstemperatur}_i + \beta_6 \cdot \text{Datum}_i \\ & + \beta_7 \cdot \text{Sonnenscheindauer}_i + \beta_8 \cdot \text{Körpertemperaturindex}_i \\ & + \beta_9 \cdot (\text{Atemfrequenz}_i) + \beta_{10} \cdot (\text{Atemfrequenz}_i^2) + \epsilon_i, \end{aligned}$$

wobei die Koeffizienten β_9 und β_{10} die linearen und quadratischen Effekte der Atemfrequenz darstellen. Der Fehlerterm ϵ_i wird als normalverteilt angenommen, mit $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

Die Priorverteilungen und Modellannahmen entsprechen denjenigen des linearen Modells. Die Koeffizienten β_1 bis β_{10} wurden mit einer normalverteilten Prior $\mathcal{N}(0, 1)$ belegt, während der Achsenabschnitt β_0 eine breitere Prior von $\mathcal{N}(0, 5)$ erhielt. Die Anpassung erfolgte unter Verwendung der probabilistischen Programmiersprache Stan und des Hamiltonian Monte Carlo-Samplers. Auch hier wurden alle Markov-Ketten validiert, und die Konvergenz wurde durch \hat{R} -Werte von 1.00 bestätigt.

4.4.2 Ergebnisse und Interpretation

Die Schätzergebnisse für die Regressionskoeffizienten des bayesianischen Modells mit einem quadratischen Term für die Atemfrequenz sind in Tabelle 9 zusammengefasst. Der Intercept beträgt 50.72 (Standardfehler: 3.17, 95%-KI: 44.28, 56.85) und repräsentiert den durchschnittlichen Schlafwert, wenn alle Prädiktoren auf ihrem Referenzwert liegen.

Die Ruheherzfrequenz zeigt einen signifikanten positiven Effekt auf den Schlafwert (0.07, Standardfehler: 0.01, 95%-KI: 0.04, 0.10). Dies deutet darauf hin, dass eine höhere Ruheherzfrequenz mit einem Anstieg des Schlafwerts assoziiert ist, was auf eine Verbindung

zwischen der Herzgesundheit und der Schlafqualität hinweist. Der Herzfrequenzvariabilitätsindex weist hingegen einen minimal negativen Effekt auf (-0.02 , Standardfehler: 0.01 , 95%-KI: $-0.04, 0.01$), der jedoch statistisch nicht signifikant ist.

Ein klarer positiver Einfluss wird beim Erholungsindex beobachtet (0.15 , Standardfehler: 0.01 , 95%-KI: $0.12, 0.17$), was die Bedeutung der nächtlichen Regeneration für die Schlafqualität unterstreicht. Die Vortagsaktivität zeigt ebenfalls einen positiven Effekt (0.04 , Standardfehler: 0.03 , 95%-KI: $-0.01, 0.10$), der aber nicht signifikant ist.

Die durchschnittliche Temperatur hat einen signifikanten negativen Einfluss auf den Schlafwert (-0.33 , Standardfehler: 0.04 , 95%-KI: $-0.40, -0.25$), was darauf hindeutet, dass höhere Temperaturen den Schlafwert negativ beeinflussen können. Der Effekt der Zeitvariable Datum (-0.00 , Standardfehler: 0.00 , 95%-KI: $-0.00, 0.00$) ist null und praktisch irrelevant.

Das polynomiale Modell untersucht die potenziellen nichtlinearen Effekte der Atemfrequenz auf den Schlafwert. Die beiden quadratischen Terme der Atemfrequenz zeigen dabei keine signifikanten Effekte (-0.53 , Standardfehler: 1.00 , 95%-KI: $-2.50, 1.41$ für den linearen Term und -0.35 , Standardfehler: 1.01 , 95%-KI: $-2.33, 1.62$ für den quadratischen Term). Dies deutet darauf hin, dass die Atemfrequenz in dieser Modellierung keine wesentliche Rolle für die Vorhersage des Schlafwerts spielt.

Trotz der fehlenden Signifikanz der quadratischen Terme zeigt das Modell, dass andere Prädiktoren, wie der Erholungsindex und die durchschnittliche Temperatur, weiterhin einen wesentlichen Einfluss auf den Schlafwert ausüben.

Tabelle 9: Ergebnisse des bayesianischen Modells mit Polynom für die Atemfrequenz

Prädiktor	Schätzwert	Std. Fehler	95% KI (unten, oben)
Intercept	50.72	3.17	(44.28, 56.85)
Ruheherzfrequenzindex	0.07	0.01	(0.04, 0.10)
Atemfrequenz (linear)	-0.53	1.00	(-2.50, 1.41)
Atemfrequenz (quadratisch)	-0.35	1.01	(-2.33, 1.62)
Herzfrequenzvariabilitätsindex	-0.02	0.01	(-0.04, 0.01)
Erholungsindex	0.15	0.01	(0.12, 0.17)
Vortagsaktivitätsindex	0.04	0.03	(-0.01, 0.09)
Durchschnittstemperatur	-0.33	0.04	(-0.40, -0.25)
Datum	0.00	0.00	(0.00, 0.01)
Sonnenscheindauer	-0.00	0.00	(-0.00, 0.00)
Körpertemperaturindex	-0.05	0.02	(-0.09, -0.00)

4.5 Evaluierung der Modelle

Die Modellgüte der linearen und nichtlinearen Modelle wurde unter Verwendung von Posterior Predictive Checks (PPC), dem Leave-One-Out Cross-Validation (LOO-CV)-Kriterium und Residuals-vs-Fitted-Plots überprüft. Ziel dieser Analyse war es, die Übereinstimmung zwischen den Modellvorhersagen und den tatsächlichen Daten zu bewerten, die Modellgüte visuell zu beurteilen sowie die relative Generalisierungsfähigkeit der Modelle zu vergleichen.

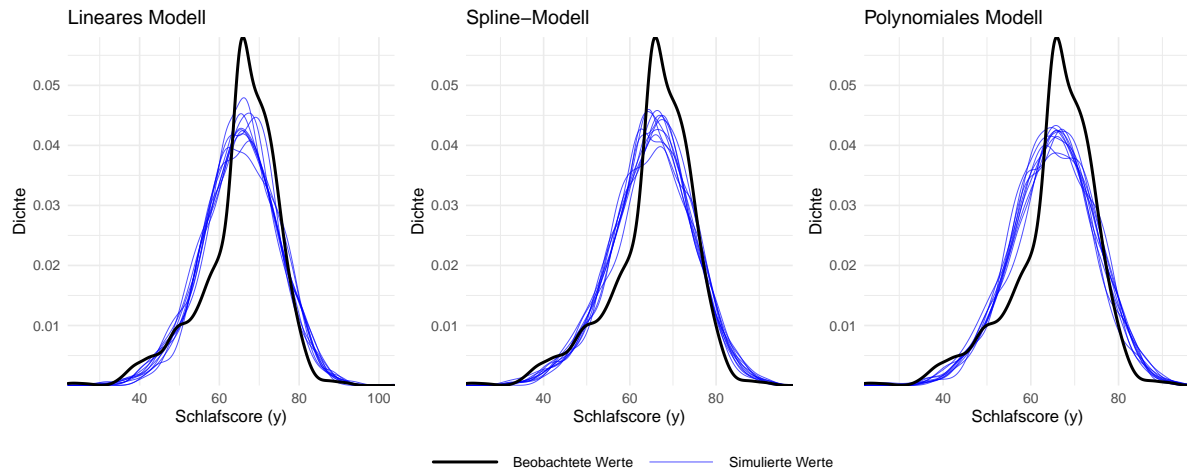


Abbildung 4: Posterior Predictive Checks der Modelle

Der Posterior Predictive Check (PPC) ist eine zentrale Methode zur Beurteilung der Modellgüte in der bayesianischen Statistik. Hierbei werden simulierte Daten y_{rep} , die aus dem Posterior der Modelle generiert wurden, mit den beobachteten Daten y verglichen. Eine gute Übereinstimmung zwischen beiden Verteilungen spricht dafür, dass das Modell die Daten adäquat beschreibt. In Abbildung 4 sind die Ergebnisse der Posterior Predictive Checks für die drei Modelle mit der Zielgröße Schlafwert dargestellt.

Das einfache lineare Modell zeigt in der linken Grafik eine gute Übereinstimmung zwischen den simulierten und beobachteten Werten, insbesondere im Bereich der zentralen Verteilung. Leichte Abweichungen in den Randbereichen deuten jedoch darauf hin, dass das lineare Modell bestimmte Extremwerte nicht vollständig abbilden konnte. Im Vergleich dazu weist das Modell mit polynomialen Termen (mittlere Grafik) eine geringere Übereinstimmung in den Randbereichen auf, was auf eine potenziell schlechtere Anpassung hinweist. Das Modell mit Splines (rechte Grafik) zeigt hingegen eine insgesamt bessere Anpassung an die Daten. Besonders auffällig ist, dass dieses Modell die Randbereiche der Verteilung genauer beschreibt, was die höhere Flexibilität von Splines bei der Modellierung komplexer Zusammenhänge unterstreicht.

Zusätzlich zu den Posterior Predictive Checks wurde das Leave-One-Out Cross-Validation (LOO-CV)-Kriterium herangezogen, um die Modelle quantitativ zu vergleichen. Dieses Verfahren schätzt die Generalisierungsfähigkeit eines Modells, indem iterativ jeweils ein Datenpunkt aus dem Datensatz entfernt und auf Grundlage der verbleibenden Daten eine Vorhersage für den entfernten Punkt berechnet wird. Die Modellgüte wird dann durch den erwarteten log-punktweisen Vorhersagewert (*expected log pointwise predictive density*, *elpd*) bewertet, wobei höhere *elpd*-Werte auf eine bessere Modellgüte hinweisen.

Tabelle 10: Ergebnisse des Leave-One-Out Cross-Validation-Kriteriums

Modell	<i>elpd-diff</i>	<i>se-diff</i>
Spline-Modell	0.0	0.0
Lineares Modell	-6.1	4.0
Polynomiales Modell	-14.3	7.4

In Ergänzung zu den PPCs und LOO-CV-Ergebnissen wurden die Residuen der Modelle im Verhältnis zu den angepassten Werten (Fitted Values) visualisiert. Diese Analyse ermöglicht es, systematische Abweichungen, Heteroskedastizität oder andere Modellierungsprobleme zu identifizieren.

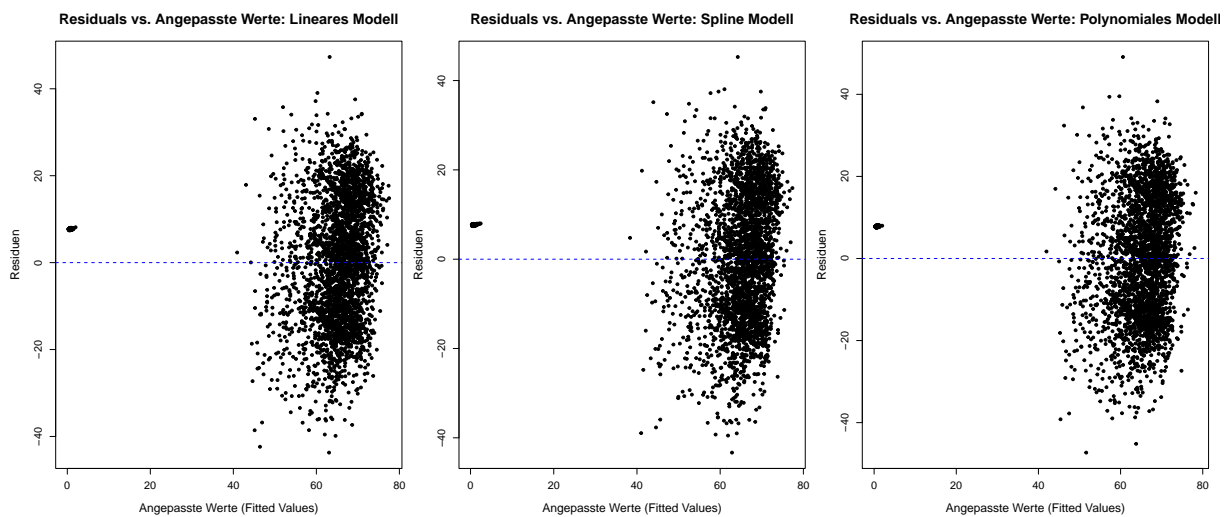


Abbildung 5: Residuen vs. Angepasste Werte für die drei Modelle

Abbildung 5 zeigt die Residuenplots für die drei Modelle. Im linearen Modell (linke Grafik) lassen sich klare Muster in den Residuen erkennen, insbesondere bei extremen Werten der angepassten Werte. Dies deutet darauf hin, dass das lineare Modell leichte Schwierigkeiten hat, nichtlineare Zusammenhänge oder Extremwerte adäquat zu modellieren. Beim Spline-Modell (mittlere Grafik) zeigen die Residuen eine homogenere Verteilung, ohne klare systematische Muster, was auf eine bessere Modellanpassung hinweist. Die Flexibilität des Splines ermöglicht es, komplexe nichtlineare Beziehungen in den Daten zu erfassen. Im polynomiale Modell (rechte Grafik) sind leichte Muster in den Residuen sichtbar, insbesondere bei mittleren angepassten Werten, was darauf hindeutet, dass dieses Modell ebenfalls nicht die optimale Anpassung an die Daten bietet.

Zusammenfassend zeigen die Residuals-vs-Fitted-Plots, dass das Spline-Modell die beste Anpassung an die Daten bietet, gefolgt vom linearen Modell und dem polynomialen Modell. Diese Ergebnisse unterstützen die Schlussfolgerungen aus den PPCs und dem LOO-CV-Kriterium.

Zusammenfassend bestätigen die verschiedenen Gütemaße und Visualisierungen, dass das Spline-Modell die präzisesten Vorhersagen liefert und die beste Wahl für die Mo-

dellierung der Zielgröße Schlafwert darstellt. Trotzdem ist zu beachten, dass das lineare Modell eine gute Modellgüte aufzeigt und dabei weniger Komplexität und eine leichtere Interpretierbarkeit mit sich bringt.

5 Modellierung indirekter Einflussfaktoren auf die Schlafqualität

Die Wahl der Zielgröße Atemfrequenz wurde sowohl durch inhaltliche als auch durch statistische Überlegungen motiviert. In den vorangegangenen Modellen mit der Zielgröße Schlafwert zeigte sich die Atemfrequenz als ein zentraler Einflussfaktor auf die Schlafqualität, insbesondere in der Lasso-Regularisierung sowie den Ergebnissen des linearen bayesianischen Modells. Darüber hinaus wies die Atemfrequenz eine vergleichsweise hohe Korrelation mit dem Schlafwert auf ($r = -0.39$), was ihre Bedeutung als potenziell erklärende Variable unterstreicht.

Im Vergleich zu alternativen Variablen wie der durchschnittlichen Außentemperatur und dem Erholungsindex bietet die Atemfrequenz den Vorteil, dass sie ein direkter physiologischer Indikator ist, der potenziell beeinflusst werden kann. Während die durchschnittliche Tagestemperatur für ein Modell informativ sein kann, ist sie inhaltlich als Zielgröße ungeeignet, da sie von externen Umweltbedingungen abhängt und nicht individuell veränderbar ist. Ebenso ist der Erholungsindex, obwohl er einen signifikanten Zusammenhang mit dem Schlafwert aufwies, eine aggregierte Metrik, die verschiedene Einflussfaktoren zusammenfasst und daher weniger spezifisch ist. Dies macht ihn ungeeignet, um als primäre Zielgröße herangezogen zu werden.

Die Wahl der Zielgröße Atemfrequenz ist somit durch ihren direkten physiologischen Bezug, ihre Relevanz in den vorherigen Analysen und ihre praktische Anwendbarkeit für weiterführende Interpretationen begründet. Ziel der Modellierung ist es, die Mechanismen zu untersuchen, die die nächtliche Atemfrequenz beeinflussen, um ein besseres Verständnis für die zugrunde liegenden physiologischen Prozesse und deren potenzielle Beeinflussbarkeit zu gewinnen.

5.1 Variablenselektion

Die Variablenselektion für die Modelle mit der Zielgröße Atemfrequenz folgte dem selben systematischen Ansatz wie bei den Modellen mit dem Schlafwert als Zielvariable. Ziel war es, die Analyse auf Prädiktoren zu fokussieren, die sowohl inhaltlich als auch statistisch relevant für die Erklärung der Zielgröße sind.

5.1.1 Inhaltliche Überlegungen

Zunächst wurden dieselben Variablen wie im vorherigen Selektionsprozess ausgeschlossen, die entweder redundante oder wenig informative Informationen lieferten oder direkt mit Schlafmetriken verknüpft. Dieser Prozess stellte sicher, dass die verbleibenden Prädiktoren unabhängige Erklärungsgrößen darstellen und das Modell sowohl interpretierbar als auch präzise bleibt.

5.1.2 Korrelationsanalyse

Die Korrelationsanalyse wurde erneut durchgeführt, um die linearen Zusammenhänge zwischen den verbleibenden Variablen und der Zielgröße Atemfrequenz zu bewerten. Ziel war es, Variablen mit den stärksten Korrelationen zu identifizieren, die für die weitere Analyse

berücksichtigt werden sollen. Abbildung 6 zeigt die Korrelationsanalyse für die zuvor ausgewählten Variablen. Außerdem wurden hier ebenfalls nach den höchsten Korrelationen mit der Atemfrequenz selektiert.

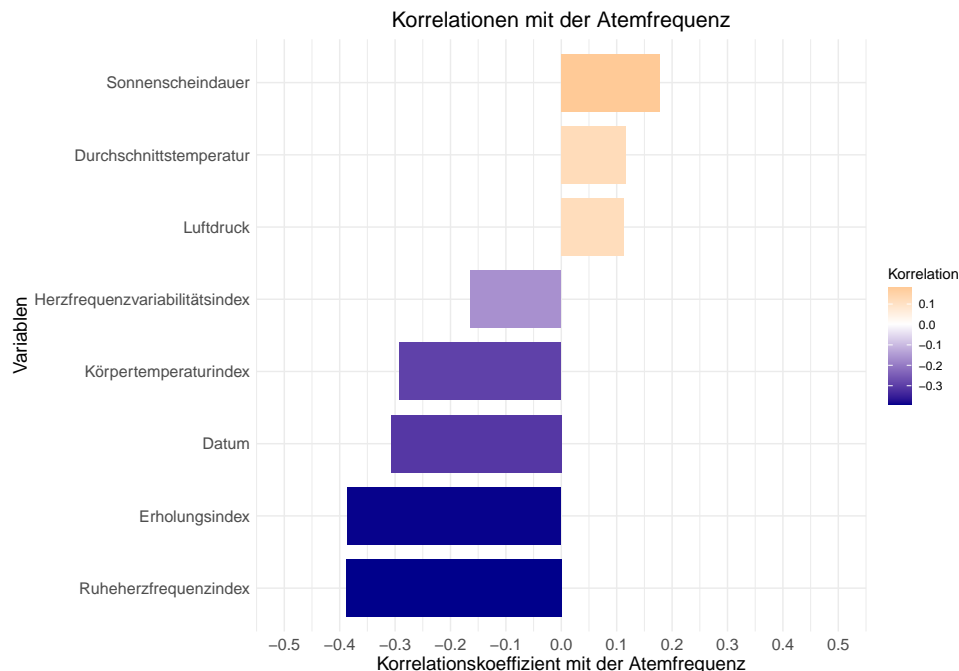


Abbildung 6: Korrelationen der ausgewählten Variablen mit der Atemfrequenz

Die Abbildung zeigt, dass der Ruheherzfrequenzindex mit einer Korrelation von $r = -0.4$ den stärksten negativen Zusammenhang mit der Atemfrequenz aufweist, somit ist eine niedrige Ruheherzfrequenz mit einer niedrigen Atemfrequenz assoziiert. Ebenfalls stark negativ korreliert ist der Erholungsindex mit $r = -0.3$, was darauf hindeutet, dass eine schnellere nächtliche Erholung der Herzfrequenz mit einer niedrigeren Atemfrequenz verbunden ist. Der Körpertemperaturindex zeigt eine moderate negative Korrelation von $r = -0.2$, was darauf hindeutet, dass niedrigere Körpertemperaturen mit einer stabileren und geringeren Atemfrequenz assoziiert sind.

Die Herzfrequenzvariabilität zeigt eine schwach negative Korrelation von $r = -0.1$, was die Relevanz der autonomen Regulation für die Atemfrequenz betont. Unter den meteorologischen Variablen zeigen sich ebenfalls interessante Zusammenhänge: Der Luftdruck weist eine positive Korrelation von $r = 0.1$ auf, was auf potenzielle Einflüsse durch atmosphärische Bedingungen hinweist, während die durchschnittliche Tagestemperatur ($r = 0.1$) und die tägliche Sonnenscheindauer ($r = 0.2$) ebenfalls moderate Zusammenhänge mit der Atemfrequenz zeigen. Diese Ergebnisse deuten darauf hin, dass externe Umweltbedingungen die nächtliche Atmung indirekt beeinflussen könnten.

Die Zeitvariable Datum zeigt mit $r = -0.3$ einen negativen Zusammenhang mit der Atemfrequenz, was auf langfristige Trends oder saisonale Effekte hinweisen könnte. Diese Ergebnisse bestätigen, dass physiologische und externe Faktoren die Atemfrequenz während des Schlafs beeinflussen können.

Basierend auf diesen Ergebnissen umfassen die final ausgewählten Variablen den Ruhe-

herzfrequenzindex, den Herzfrequenzvariabilitätsindex, den Körpertemperaturindex, den Erholungsindex, die durchschnittliche Temperatur, die tägliche Sonnenscheindauer, den Luftdruck und die Zeitvariable Datum. Diese Auswahl stellt sicher, dass sowohl physiologische als auch umweltbedingte Einflussfaktoren berücksichtigt werden, die für die Vorhersage der nächtlichen Atemfrequenz relevant sind. Die Kombination dieser Prädiktoren erlaubt es, die zugrunde liegenden Mechanismen der Atemfrequenz differenziert zu analysieren und potenzielle Interventionsmöglichkeiten zu identifizieren.

5.1.3 Prüfung auf Multikollinearität

Nach der Auswahl der relevanten Prädiktoren wurde wie für die Modelle mit Zielvariable Schlafwert eine Multikollinearitätsprüfung durchgeführt, um sicherzustellen, dass die unabhängigen Variablen nicht stark miteinander korrelieren und somit keine Stabilitätsprobleme bei der Schätzung der Regressionskoeffizienten auftreten. Die Bewertung der Multikollinearität erfolgte erneut mithilfe des Variance-Inflation-Faktors (VIF). Alle ausgewählten Variablen wiesen VIF-Werte unterhalb der gängigen Schwellenwerte von 5 auf, was darauf hinweist, dass keine kritische Multikollinearität in den Daten vorliegt. Die Ergebnisse der VIF-Berechnung sind in Tabelle 11 dargestellt.

Tabelle 11: Variance-Inflation-Faktor (VIF) der ausgewählten Prädiktoren

Variable	VIF-Wert
Ruheherzfrequenzindex	1.77
Herzfrequenzvariabilitätsindex	1.46
Körpertemperaturindex	1.22
Erholungsindex	1.30
Datum	1.22
Sonnenscheindauer	1.56
Durchschnittstemperatur	1.73
Luftdruck	1.22

Alle Variablen konnten daher uneingeschränkt in den weiteren Modellierungsprozess integriert werden. Dies unterstreicht die Robustheit der Variablenselektion und gewährleistet, dass die Schätzungen der Regressionskoeffizienten zuverlässig und stabil bleiben.

5.1.4 Lasso-Regression

Im nächsten Schritt wurde wieder eine Lasso-Regularisierung durchgeführt, um ein erstes Bild der wichtigsten Prädiktoren für die Zielgröße Atemfrequenz zu erhalten.

Tabelle 12: Koeffizienten der Lasso-Regression mit optimalem λ

Variable	Koeffizient
Achsenabschnitt	10.65
Ruheherzfrequenz	-0.0065
Herzfrequenzvariabilität	0.0034
Körpertemperatur	-0.0081
Erholungsindex	-0.0052
Datum	-0.0004
Sonnenscheindauer	0.00004
Durchschnittstemperatur	0.0139
Luftdruck	0.0055

Die Ergebnisse der Lasso-Regression zeigen, dass einige Variablen eine leichte Assoziation mit der Zielgröße Atemfrequenz aufweisen, während andere nahezu auf null gesetzt wurden. Der Achsenabschnitt wurde auf 10.65 geschätzt, was den Basiswert der Atemfrequenz repräsentiert, wenn alle Prädiktoren gleich null sind. Die Ruheherzfrequenz hatte einen leicht negativen Koeffizienten von -0.0065, was darauf hindeutet, dass eine Erhöhung der Ruheherzfrequenz mit einer minimalen Reduktion der Atemfrequenz einhergeht. Im Gegensatz dazu zeigte die Herzfrequenzvariabilität einen positiven Koeffizienten von 0.0034, was darauf hinweist, dass eine höhere Variabilität mit einer geringfügigen Zunahme der Atemfrequenz verbunden ist.

Die Körpertemperatur und der Erholungsindex wiesen negative Koeffizienten von -0.0081 beziehungsweise -0.0052 auf, was darauf schließen lässt, dass physiologische Faktoren, die auf Erholung und Temperaturregulation hinweisen, die Atemfrequenz senken könnten. Interessanterweise zeigten meteorologische Variablen wie die Durchschnittstemperatur und der Luftdruck positive Koeffizienten von 0.0139 und 0.0055. Dies deutet darauf hin, dass externe Umweltfaktoren eine direkte Rolle bei der Variabilität der Atemfrequenz spielen könnten.

Die Variablen Datum und Sonnenscheindauer hatten nahezu vernachlässigbare Koeffizienten von -0.0004 und 0.00004, was darauf hinweist, dass sie in diesem Modell nur eine sehr geringe Bedeutung haben. Der Koeffizient für die Aktivität betrug -0.0001, was ebenfalls auf eine minimale Rolle dieser Variable in der Modellierung hindeutet.

Die Lasso-Regularisierung liefert hierbei wertvolle Hinweise auf die wichtigsten Einflussgrößen, die in weiteren Modellierungsansätzen detaillierter untersucht werden können. Um möglichst informative Ergebnissen in den weiteren Modellen zu erhalten wurden keine Variablen durch die Lasso-Regression ausgeschlossen.

5.2 Bayesianisches lineares Modell mit Zielgröße Atemfrequenz

Für die Analyse der Zielgröße Atemfrequenz wurde ein bayesianisches lineares Regressionsmodell entwickelt, um die linearen Zusammenhänge zwischen der Zielgröße und den ausgewählten Prädiktoren zu untersuchen. Die Modellformel des bayesianischen linearen Modells kann wie folgt mathematisch dargestellt werden:

$$\begin{aligned} \text{Atemfrequenz}_i = & \beta_0 + \beta_1 \cdot \text{Ruheherzfrequenzindex}_i + \beta_2 \cdot \text{Herzfrequenzvariabilitätsindex}_i \\ & + \beta_3 \cdot \text{Körpertemperaturindex}_i + \beta_4 \cdot \text{Erholungsindex}_i \\ & + \beta_5 \cdot \text{Datum}_i + \beta_6 \cdot \text{Sonnenscheindauer}_i \\ & + \beta_7 \cdot \text{Durchschnittstemperatur}_i + \beta_8 \cdot \text{Luftdruck}_i + \epsilon_i, \end{aligned}$$

wobei $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ den normalverteilten Fehlerterm beschreibt. Die Prädiktoren wurden auf Grundlage der vorhergehenden Variablenselektion ausgewählt und umfassen sowohl physiologische als auch meteorologische Einflussgrößen.

Die Priorverteilungen und die Modellanpassung erfolgten analog zum bayesianischen linearen Modell mit der Zielgröße Schlafwert (siehe Abschnitt 4.2.2). Die Wahl informativer Priors und der Einsatz des Hamiltonian Monte Carlo (HMC) mit dem No-U-Turn Sampler (NUTS) garantierten eine effiziente und zuverlässige Schätzung der Parameter. Die Konvergenz der Posterior-Verteilungen wurde durch \hat{R} -Werte von 1.00 bestätigt.

Die Ergebnisse des bayesianischen linearen Modells werden in Tabelle 13 dargestellt.

Tabelle 13: Ergebnisse des bayesianischen linearen Modells für Atemfrequenz

Prädiktor	Schätzwert	Std. Fehler	95% KI (unten, oben)
Achsenabschnitt	10.51	1.83	(6.93, 14.04)
Ruheherzfrequenzindex	-0.01	0.00	(-0.01, -0.00)
Herzfrequenzvariabilitätsindex	0.00	0.00	(0.00, 0.01)
Körpertemperaturindex	-0.01	0.00	(-0.01, -0.01)
Erholungsindex	-0.01	0.00	(-0.01, -0.00)
Datum	-0.00	0.00	(-0.00, -0.00)
Sonnenscheindauer	0.00	0.00	(-0.00, 0.00)
Durchschnittstemperatur	0.01	0.00	(0.01, 0.02)
Luftdruck	0.01	0.00	(0.00, 0.01)

Die Ergebnisse zeigen, dass die Ruheherzfrequenz einen signifikanten geringen negativen Einfluss auf die Atemfrequenz hat (-0.01 , 95%-KI: $-0.01, -0.00$). Dieser Zusammenhang deutet darauf hin, dass eine Erhöhung der Ruheherzfrequenz mit einer minimalen Reduktion der Atemfrequenz verbunden ist.

Der Erholungsindex zeigt ebenfalls einen kleinen, aber signifikanten negativen Einfluss (-0.01 , 95%-KI: $-0.01, -0.00$). Dies deutet darauf hin, dass eine effektive nächtliche Regeneration mit einer niedrigeren Atemfrequenz assoziiert ist. Auch die Körpertemperatur weist mit -0.01 (95%-KI: $-0.01, -0.01$) einen negativen Zusammenhang auf, was zeigt, dass eine Abweichung von der normalen Körpertemperatur die Atemfrequenz leicht senken kann.

Unter den meteorologischen Prädiktoren zeigt die durchschnittliche Temperatur einen positiven und signifikanten niedrigen Einfluss (0.01 , 95%-KI: $0.01, 0.02$), was darauf hinweist, dass höhere Temperaturen mit einer leicht erhöhten Atemfrequenz verbunden sind. Der Luftdruck hat ebenfalls einen positiven Effekt (0.01 , 95%-KI: $0.00, 0.01$), der jedoch gering ausfällt.

Andere Variablen wie die Herzfrequenzvariabilität (0.00, 95%-KI: 0.00, 0.01), die Sonnenscheindauer (0.00, 95%-KI: -0.00, 0.00) und die Zeitvariable Datum (-0.00, 95%-KI: -0.00, -0.00) zeigen keine signifikanten Effekte, was auf deren begrenzte Relevanz für die Vorhersage der Atemfrequenz in diesem Modell hinweist.

Insgesamt unterstreichen die Ergebnisse, dass insbesondere physiologische Faktoren wie die Ruheherzfrequenz, der Erholungsindex und die Körpertemperatur sowie meteorologische Einflüsse wie die durchschnittliche Temperatur und der Luftdruck signifikante Prädiktoren für die Atemfrequenz darstellen, jedoch in geringem Maße.

5.2.1 Erweiterung des Modells um Interaktionen

Nach der Spezifikation des bayesianischen linearen Modells für die Zielgröße Atemfrequenz wurde untersucht, ob die Integration von Interaktionseffekten zwischen den Prädiktoren zu einer Verbesserung der Modellerklärung beitragen könnte. Interaktionen sind besonders dann sinnvoll, wenn angenommen wird, dass der Einfluss eines Prädiktors auf die Zielgröße durch einen anderen Prädiktor moderiert wird. Um mögliche Interaktionen systematisch zu identifizieren, wurde ein Modell mit allen potenziellen Interaktionen der ausgewählten Prädiktoren spezifiziert. Hierbei wurden sämtliche Kombinationen von Variablen berücksichtigt. Für die Schätzung der Interaktionen wurde erneut die Horseshoe-Regularisierung verwendet, da diese Methode sich besonders für hochdimensionale Modelle eignet und in der Lage ist, unwichtige Koeffizienten stark abzuschwächen, während wichtige Effekte erhalten bleiben.

Die Ergebnisse der Analyse zeigten jedoch, dass sämtliche Interaktionen keinen signifikanten Einfluss auf die Atemfrequenz hatten. Alle geschätzten Effekte der Interaktionen lagen nahe null, und die entsprechenden 95%-Konfidenzintervalle umfassten stets den Wert null. Dies deutet darauf hin, dass die Einflüsse der betrachteten Interaktionen auf die Zielgröße nicht substantiell waren. Daher wurde entschieden, keine Interaktionseffekte in das finale Modell zu integrieren.

Die Tatsache, dass die Interaktionen keinen signifikanten Beitrag zur Modellgüte leisteten, könnte darauf hindeuten, dass die zugrunde liegenden Zusammenhänge zwischen den Prädiktoren und der Atemfrequenz unabhängig voneinander sind. Dies ist insbesondere bei physiologischen Variablen plausibel, da diese oft isolierte Mechanismen widerspiegeln, die nicht notwendigerweise miteinander in Wechselwirkung stehen. Die Entscheidung, keine Interaktionen in das Modell aufzunehmen, vereinfacht zudem die Interpretation der Ergebnisse, da sich die geschätzten Haupteffekte auf die direkten Einflüsse der Prädiktoren auf die Atemfrequenz beschränken.

5.3 Nichtlineare bayesianische Modelle mit Zielgröße Atemfrequenz

Die Überprüfung der Residuenplots für die Zielgröße Atemfrequenz im linearen Modell ergab keine Hinweise auf systematische Muster oder deutliche Abweichungen. Die Residuen der stetigen Variablen verteilten sich zufällig um die Null-Linie, was darauf hinweist, dass die Annahme eines linearen Zusammenhangs zwischen den Prädiktoren und der Zielgröße erfüllt ist. Diese Beobachtung deutet darauf hin, dass das lineare Modell die

Daten angemessen beschreibt.

Um die Modellstruktur weiter zu validieren, wurden dennoch nichtlineare Modelle getestet, indem Splines und quadratische Terme für die erklärenden Variablen eingeführt wurden. Dieser Ansatz ermöglichte es, potenzielle nichtlineare Zusammenhänge zu berücksichtigen, selbst wenn sie nicht sofort aus den Residuenplots ersichtlich waren.

Die Ergebnisse der nichtlinearen Modelle ergaben jedoch keine signifikanten Effekte für die eingeführten nichtlinearen Terme. Die geschätzten Koeffizienten lagen nahe null, und die entsprechenden 95%-Kreditintervalle umschlossen ebenfalls null. Darüber hinaus zeigte ein Vergleich der Modellgüte zwischen den linearen und nichtlinearen Modellen, dass die Einführung nichtlinearer Terme keine Verbesserung der Modellanpassung bewirkte. Zusammenfassend bestätigen sowohl die Überprüfung der Residuenplots als auch die Ergebnisse der nichtlinearen Modelle, dass ein lineares Modell für die Zielgröße Atemfrequenz ausreichend ist. Daher wurde auf die weitere Verwendung nichtlinearer Modelle verzichtet.

5.4 Modellevaluierung

Die Modellgüte des linearen Modells mit der Zielgröße Atemfrequenz wurde evaluiert, da die Modelle mit Splines und polynomialen Termen keine signifikanten Einflüsse zeigten und daher nicht weiter betrachtet wurden. Zur Evaluation wurden ebenfalls ein Posterior Predictive Check (PPC) und die Rhat-Werte verwendet, um die Anpassung des Modells an die beobachteten Daten und die Konvergenz der MCMC-Sampling-Prozesse zu bewerten. Der PPC überprüft, inwiefern die simulierten Daten (y_{rep}) des Modells mit den beobachteten Daten (y) übereinstimmen. Abbildung 7 zeigt die Dichteverteilungen der beobachteten Werte (y) in Schwarz und der simulierten Werte (y_{rep}) in Blau. Eine gute Übereinstimmung zwischen den beiden Verteilungen deutet darauf hin, dass das Modell die tatsächlichen Daten gut repräsentiert.

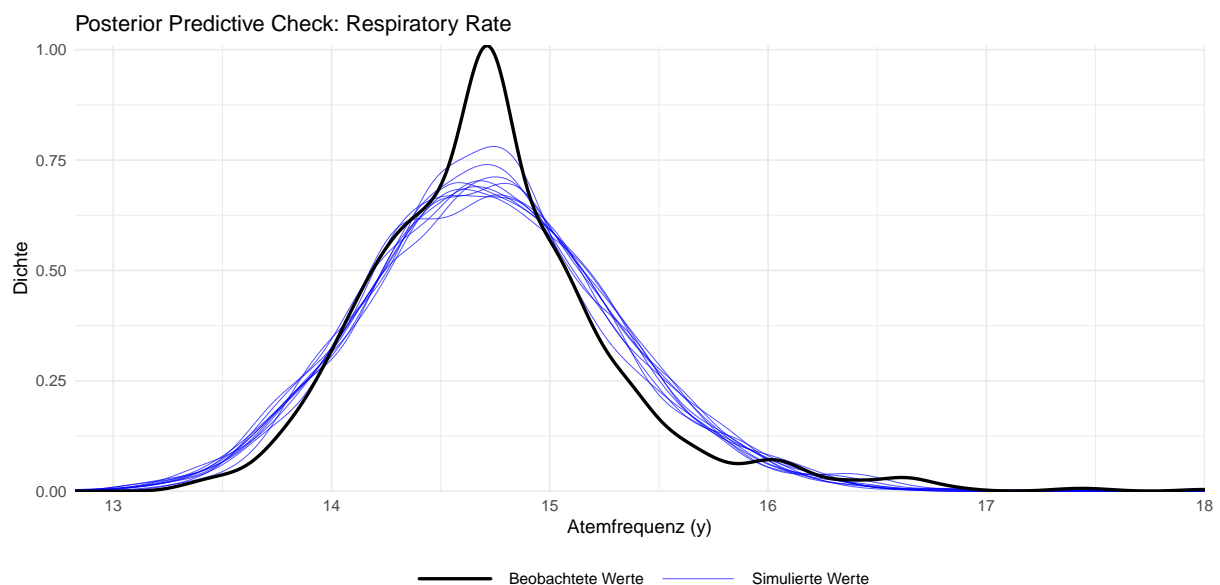


Abbildung 7: Posterior Predictive Check des linearen Modells

Die Abbildung zeigt, dass das lineare Modell die zentrale Tendenz der Daten gut abbildet. Die beobachtete Verteilung und die simulierten Verteilungen stimmen insbesondere im zentralen Bereich um den Mittelwert der Atemfrequenz gut überein. Dennoch lassen sich leichte Abweichungen in den Randbereichen feststellen, was darauf hinweist, dass das Modell die Extrema der Verteilung weniger genau erfasst. Diese Abweichungen könnten auf eine begrenzte Flexibilität des linearen Modells hinweisen, komplexere Muster oder nichtlineare Zusammenhänge zu erfassen.

Neben der Bewertung der Modellanpassung wurden die Rhat-Werte überprüft, um die Konvergenz des MCMC-Samplings zu bewerten. Für alle Parameter des linearen Modells lagen die Rhat-Werte bei exakt 1.00, was auf eine vollständige Konvergenz der Ketten hinweist. Dies bedeutet, dass die Ergebnisse des MCMC-Samplings stabil sind und keine Hinweise auf Sampling-Probleme vorliegen.

Die Ergebnisse des PPC und der Rhat-Werte zeigen, dass das lineare Modell mit der Zielgröße Atemfrequenz eine insgesamt gute Anpassung an die Daten aufweist. Während die zentrale Tendenz der Daten gut modelliert wird, könnten leichte Abweichungen in den Randbereichen auf eine mögliche Verbesserung durch die Berücksichtigung zusätzlicher erklärender Variablen hinweisen. Aufgrund der guten Konvergenz der MCMC-Sampling-Prozesse können die geschätzten Parameter des Modells als robust betrachtet werden.

6 Ergebnisse und Diskussion

6.1 Direkte Einflussgrößen auf die Schlafqualität

Bereits die Ergebnisse der Lasso-Regression konnten erste Einblicke in die wichtigsten direkten Einflussfaktoren auf die Schlafqualität geben, dazu zählt in erster Linie die Atemfrequenz. Des Weiteren zeigten vor allem die Durchschnittstemperatur und der Erholungsindex eine hohe Wichtigkeit. Daraus folgend ermöglichen die Modelle mit der Zielgröße Schlafwert tiefgehende Einblicke in die zentralen Variablen, die die Schlafqualität direkt beeinflussen.

Der Ruheherzfrequenzindex weist in allen Modellen einen signifikant leichten positiven Effekt auf die Schlafqualität auf. Der geschätzte Koeffizient im linearen Modell beträgt $\beta = 0.07$ (95% KI: 0.04, 0.09), im Spline-Modell $\beta = 0.04$ (95% KI: 0.02, 0.07) und im polynomialen Modell $\beta = 0.07$ (95% KI: 0.04, 0.10) was darauf hinweist, dass eine niedrigere Ruheherzfrequenz, die häufig mit einem entspannten physiologischen Zustand und effektiver Erholung assoziiert wird, die Schlafqualität verbessert.

Der Erholungsindex zeigt in jedem Modell einen signifikanten positiven Effekt auf die Schlafqualität mit einem geschätzten Koeffizienten von $\beta = 0.14$ (95% KI: 0.12, 0.16) im linearen Modell, im Spline-Modell $\beta = 0.13$ (95% KI: 0.11, 0.15) und im polynomialen Modell $\beta = 0.15$ (95% KI: 0.12, 0.17). Dieser Effekt reflektiert die Bedeutung der nächtlichen physiologischen Regeneration, insbesondere die Erholung nach Aktivität. Die Ergebnisse deuten darauf hin, der Erholungsfaktor des Körpers mit einer höheren Schlafqualität verbunden ist.

Der Körpertemperaturindex zeigt ebenfalls einen kleinen, jedoch signifikanten negativen Effekt ($\beta = -0.05$, 95% KI: -0.10, -0.01) im linearen Modell, im Spline-Modell $\beta = -0.08$ (95% KI: -0.13, -0.03) und im Polynomialen Modell $\beta = -0.05$ (95% KI: -0.09, 0.00), der darauf hinweist, dass niedrigere Abweichungen der Körpertemperatur von der Norm mit einer besseren Schlafqualität korrelieren. Dies unterstützt die Annahme, dass eine stabile Körpertemperatur entscheidend für eine optimale Schlafarchitektur ist.

Unter den externen Faktoren hat die durchschnittliche Temperatur einen signifikant negativen Einfluss auf die Schlafqualität. Der geschätzte Koeffizient beträgt $\beta = -0.28$ (95% KI: -0.36, -0.20) im linearen Modell, im Spline-Modell $\beta = -0.29$ (95% KI: -0.37, -0.22) und im polynomialen Modell $\beta = -0.33$ (95% KI: -0.40, -0.25), was darauf hindeutet, dass höhere Außentemperaturen mit einer Verschlechterung der Schlafqualität assoziiert sind.

Die Atemfrequenz zeigte im linearen Modell einen stark signifikant negativen Effekt auf die Schlafqualität mit einem geschätzten Koeffizienten von $\beta = -2.08$ (95% KI: -3.01, -1.15), wobei die nichtlinearen Modelle hier keine signifikanten Ergebnisse zeigen. Eine erhöhte Atemfrequenz während des Schlafs könnte auf physiologische Belastungen, Stress oder gesundheitliche Probleme hinweisen, die die Schlafqualität negativ beeinflussen. Diese Ergebnisse verdeutlichen die hohe Bedeutung der Atemfrequenz als indirekten Indikator für die Schlafqualität und zeigen, dass eine lineare Modellierung am sinnvollsten erscheint. Andere Variablen wie das Datum und die Sonnenscheindauer hatten in allen drei Modellen keinen signifikanten Einfluss auf die Schlafqualität, da ihre Konfidenzintervalle den Wert null umfassten. Das Fehlen eines signifikanten Zusammenhangs deutet darauf hin, dass sie im Rahmen dieser Analyse keine zentrale Rolle spielt. Ähnlich zeigte die

Herzfrequenzvariabilität nur minimale Effekte auf die Schlafqualität.

Zusammenfassend liefern die Ergebnisse wertvolle Einblicke in die Determinanten der Schlafqualität. Physiologische Variablen wie die Atemfrequenz, der Ruheherzfrequenzindex und der Erholungsindex erweisen sich als besonders relevante Prädiktoren, während externe Faktoren wie die durchschnittliche Temperatur ebenfalls eine wichtige Rolle spielen.

6.2 Indirekte Einflussgrößen auf die Schlafqualität

Die Modelle mit der Zielgröße Atemfrequenz bieten tiefgehende Einblicke in Variablen, die indirekte Einflüsse auf die Schlafqualität ausüben könnten. Die Atemfrequenz, die in den vorangegangenen Modellen als eine der signifikantesten Variablen zur Erklärung der Schlafqualität identifiziert wurde, zeigt eine vergleichsweise starke Korrelation mit dem Schlafwert ($r = -0.40$). Diese Assoziation, kombiniert mit ihrer Eigenschaft als direkt messbarer physiologischer Indikator, rechtfertigt ihre Wahl als Zielgröße zur Analyse indirekter Einflussfaktoren.

Die Ergebnisse des bayesianischen linearen Modells mit der Zielgröße Atemfrequenz zeigen, dass sowohl physiologische als auch externe Umweltfaktoren leichte Effekte auf die nächtliche Atemfrequenz haben. Unter den physiologischen Prädiktoren erweist sich der Ruheherzfrequenzindex als geringer Einflussfaktor. Der geschätzte Koeffizient beträgt $\beta = -0.01$ (95% KI: -0.01, -0.00), was auf einen negativen Zusammenhang zwischen einer niedrigeren Ruheherzfrequenz und einer reduzierten Atemfrequenz hinweist, wobei die hohe Unsicherheit zu beachten ist.

Ein weiterer Prädiktor ist der Erholungsindex mit einem geschätzten Koeffizienten von $\beta = -0.01$ (95% KI: -0.01, -0.00). Dieser Zusammenhang deutet darauf hin, dass eine schnellere Stabilisierung der Herzfrequenz nach dem Einschlafen mit einer niedrigeren Atemfrequenz verbunden sein könnte, jedoch ohne Signifikanz.

Die Körpertemperatur weist einen kleinen negativen Effekt auf die Atemfrequenz auf ($\beta = -0.01$, 95% KI: -0.01, -0.01), hier ist statistische Signifikanz vorhanden, was aufzeigt, dass sich Körpertemperaturschwankungen negativ auf die Atemfrequenz auswirken.

Zu den externen Einflussfaktoren gehört die durchschnittliche Tagestemperatur, die einen minimalen positiven Einfluss auf die nächtliche Atemfrequenz zeigt. Der geschätzte Koeffizient beträgt $\beta = 0.01$ (95% KI: 0.01, 0.02) und ist somit in diesem Modell signifikant. Ein ähnlicher positiver Effekt wird für den Luftdruck beobachtet ($\beta = 0.01$, 95% KI: 0.00, 0.01), jedoch ohne statistische Signifikanz.

Andere externe Variablen wie die Sonnenscheindauer hatten keinen signifikanten Effekt auf die Atemfrequenz ($\beta = 0.00$, 95% KI: -0.00, 0.00). Ebenso verhält es sich für die Zeitvariable Datum ($\beta = -0.00$, 95% KI: -0.00, -0.00). Diese Variablen besitzen somit keine praktische Relevanz im Kontext dieses Modells.

Die nichtlinearen Modelle mit Splines und quadratischen Termen zeigten keine zusätzlichen signifikanten Effekte über die im linearen Modell gefundenen hinaus. Die geschätzten Koeffizienten der nichtlinearen Terme lagen nahe null, und ihre 95%-Konfidenzintervalle schlossen null ein. Dies deutet darauf hin, dass die Beziehungen zwischen den Prädiktoren und der Atemfrequenz weitgehend linear sind. Residuenplots bestätigten diese Annahme, da keine systematischen Muster oder Abweichungen erkennbar waren.

6.3 Limitationen

Obwohl die Ergebnisse dieser Arbeit wertvolle Einblicke in die direkten und indirekten Einflussfaktoren auf die Schlafqualität liefern, müssen einige Limitationen berücksichtigt werden, die die Generalisierbarkeit und Interpretation der Ergebnisse beeinflussen könnten. Eine zentrale Limitation liegt in der Datengrundlage. Die verwendeten Daten stammen unter anderem von einem Smart Ring, dessen Messmethoden zwar etabliert, aber dennoch Einschränkungen unterworfen sind. Insbesondere können die Algorithmen zur Berechnung des Schlafwerts und der physiologischen Parameter wie Atemfrequenz, Herzfrequenzvariabilität und Erholungsindex Fehler in den Messungen aufweisen, die die Validität der Ergebnisse beeinflussen könnten. Diese Algorithmen sind herstellersistpezifisch und basieren auf nicht transparenten Berechnungen, was bedeutet, dass eine externe Validierung der Messgrößen fehlt. Solche Einschränkungen könnten dazu führen, dass bestimmte Zusammenhänge über- oder unterschätzt werden.

Darüber hinaus basiert die Analyse auf einer retrospektiven Betrachtung von Daten, die über einen längeren Zeitraum hinweg gesammelt wurden. Obwohl dies die Untersuchung zeitlicher Trends und Saisonalität ermöglicht, ist nicht auszuschließen, dass externe Faktoren, die nicht erfasst wurden, die Ergebnisse beeinflusst haben. Zum Beispiel könnten Ereignisse wie Krankheiten, Stress oder außergewöhnliche Umwelteinflüsse die physiologischen Parameter und damit auch die Schlafqualität beeinflusst haben, ohne dass diese Faktoren direkt in die Modelle integriert wurden. Diese unkontrollierten Einflussfaktoren könnten die Präzision der Modellschätzungen beeinträchtigen.

Des Weiteren handelt es sich in dieser Arbeit um eine Fallstudie, daher fehlt die Generalisierbarkeit der Ergebnisse. Die Ergebnisse dieser Arbeit spiegeln daher primär die spezifischen Eigenschaften der analysierten Stichprobe wider und sollten mit Vorsicht auf andere Populationen übertragen werden. Zudem ist die Stichprobengröße mit knapp 3 Jahren ausbauungsfähig.

Ein weiterer Punkt betrifft die Modellierung selbst. Bayesianische Modelle in brms (oder anderen Frameworks wie Stan) verwenden probabilistische Sampling-Methoden wie den Hamiltonian Monte Carlo (HMC)-Algorithmus. Diese Verfahren erzeugen Stichproben aus der Posterior-Verteilung, und obwohl sie bei korrekter Konvergenz die wahre Verteilung gut approximieren, gibt es natürliche stochastische Schwankungen zwischen den einzelnen Durchläufen. Selbst bei ausreichender Konvergenz der Ketten (Chains) können kleine Unterschiede in den Mittelwerten und Standardabweichungen der Posterior-Schätzungen auftreten.

Zusammenfassend weisen die Ergebnisse dieser Arbeit trotz sorgfältiger Methodik mehrere Limitationen auf, die bei der Interpretation und Generalisierung berücksichtigt werden sollten. Künftige Studien könnten diese Einschränkungen adressieren, indem sie umfassendere Datensätze, zusätzliche potenzielle Einflussgrößen und alternative Modellierungsansätze einbeziehen. Dies würde nicht nur die Validität und Aussagekraft der Ergebnisse erhöhen, sondern auch ein tieferes Verständnis der Mechanismen hinter der Schlafqualität und ihrer Determinanten ermöglichen.

7 Fazit

Die vorliegende Arbeit hatte das Ziel, die Einflussfaktoren auf die Schlafqualität durch eine bayesianische Analyse systematisch zu untersuchen. Dabei wurde zuerst die Datengrundlage betrachtet und eine ausführliche deskriptive Analyse zur Gewinnung erster Eiblicke in die Daten vorgenommen. Die Modellierung verfolgte einen zweistufigen Ansatz, der sowohl direkte als auch indirekte Einflussgrößen auf die Schlafqualität beleuchtete. Diese Methodik ermöglichte es, ein umfassendes Verständnis für die determinierenden Faktoren der Schlafqualität zu gewinnen und deren potenzielle Interventionsmöglichkeiten zu identifizieren.

Zunächst erwies sich im ersten Teil der Arbeit vor allem die Atemfrequenz sowohl in der deskriptiven Analyse als auch in den Modellen als zentraler direkter Einflussfaktor auf die Schlafqualität, wobei eine niedrigere Atemfrequenz höhere Schlafqualität bedeutet. Gefolgt davon zeigte sich die durchschnittliche Temperatur als weiterer wichtiger Prädiktor für guten Schlaf, hier wirkt sich eine hohe Außentemperatur negativ aus. Drittens schildert der Erholungsindex eine deutliche positive Wirkung auf den Schlafwert. Des weiteren konnten die Modelle aufzeigen, dass physiologische Parameter wie der Ruheherzfrequenzindex und der Körpertemperaturindex signifikante Prädiktoren für die Schlafqualität darstellen, wenn auch in geringerem Maße. Zuletzt zeigten die Sonnenscheindauer, die Herfrequenzvariabilität und das Datum nahezu keine statistisch signifikanten Auswirkungen auf den Schlaf.

Darüber hinaus wurde im zweiten Teil der Arbeit die Atemfrequenz als indirekte Zielgröße modelliert, um die Mechanismen zu untersuchen, die der Schlafqualität zugrunde liegen könnten. Die Ergebnisse zeigen, dass die Atemfrequenz stark mit der Schlafqualität korreliert und von inhaltlicher Relevanz ist. Jedoch erwiesen sich in den Modellen nur die Körpertemperaturschwankung und die durchschnittliche Tagestemperatur als geringe signifikant Einflussfaktoren, wobei sich eine schwankende Körpertemperatur negativ und die Durchschnittstemperatur des Tages positiv auf die Atemfrequenz auswirkt. Bei der Interpretation ist hierbei zu beachten, dass eine niedrige Atemfrequenz im Sinne der Schlafqualität erwünscht ist. Alle anderen Prädiktoren erwiesen sich in diesem Kontext als nicht signifikant.

Die Verwendung bayesianischer Modellierungsansätze hat sich in dieser Arbeit als besonders vorteilhaft erwiesen, da sie eine explizite Quantifizierung von Unsicherheiten und eine robuste Schätzung der Modellparameter ermöglichte. Durch die systematische Integration von Vorwissen und die Berücksichtigung potenzieller Nichtlinearitäten konnte die Analyse vertieft und die Aussagekraft der Ergebnisse erhöht werden. Gleichzeitig zeigten die Ergebnisse, dass die zugrunde liegenden Beziehungen zwischen den Prädiktoren und den Zielgrößen weitgehend linear sind, was die Einfachheit und Interpretierbarkeit der linearen Modelle unterstützt. Nichtlineare Modelle und Interaktionseffekte lieferten hingegen keine wesentlichen zusätzlichen Erkenntnisse, was die Robustheit der linearen Modelle weiter unterstreicht.

Die Limitationen der Arbeit, insbesondere die Abhängigkeit von einer spezifischen Datengrundlage, die begrenzte Berücksichtigung potenzieller Einflussgrößen und die eingeschränkte Generalisierbarkeit der Ergebnisse, müssen bei der Interpretation der Befunde berücksichtigt werden. Dennoch bieten die Ergebnisse eine solide Grundlage für zukünftige

Forschungsvorhaben. Künftige Studien könnten die hier identifizierten Einflussfaktoren weiter vertiefen, indem sie umfassendere Datensätze mit zusätzlichen Variablen einbeziehen, wie etwa psychologische Faktoren oder Ernährungsmuster. Zudem könnten dynamische Modellierungsansätze und zeitabhängige Analysen die kausalen Zusammenhänge zwischen den Prädiktoren und der Schlafqualität genauer erfassen.

Zusammenfassend leistet die vorliegende Arbeit einen wesentlichen Beitrag zum Verständnis der maßgeblichen Einflussfaktoren auf die Schlafqualität. Die Ergebnisse zeigen, dass sowohl physiologische als auch externe Umweltfaktoren eine bedeutende Rolle spielen. Damit bietet die Arbeit sowohl wissenschaftliche Erkenntnisse als auch praxisrelevante Implikationen, die zur Verbesserung der individuellen Schlafqualität und des allgemeinen Wohlbefindens beitragen können.

A Anhang

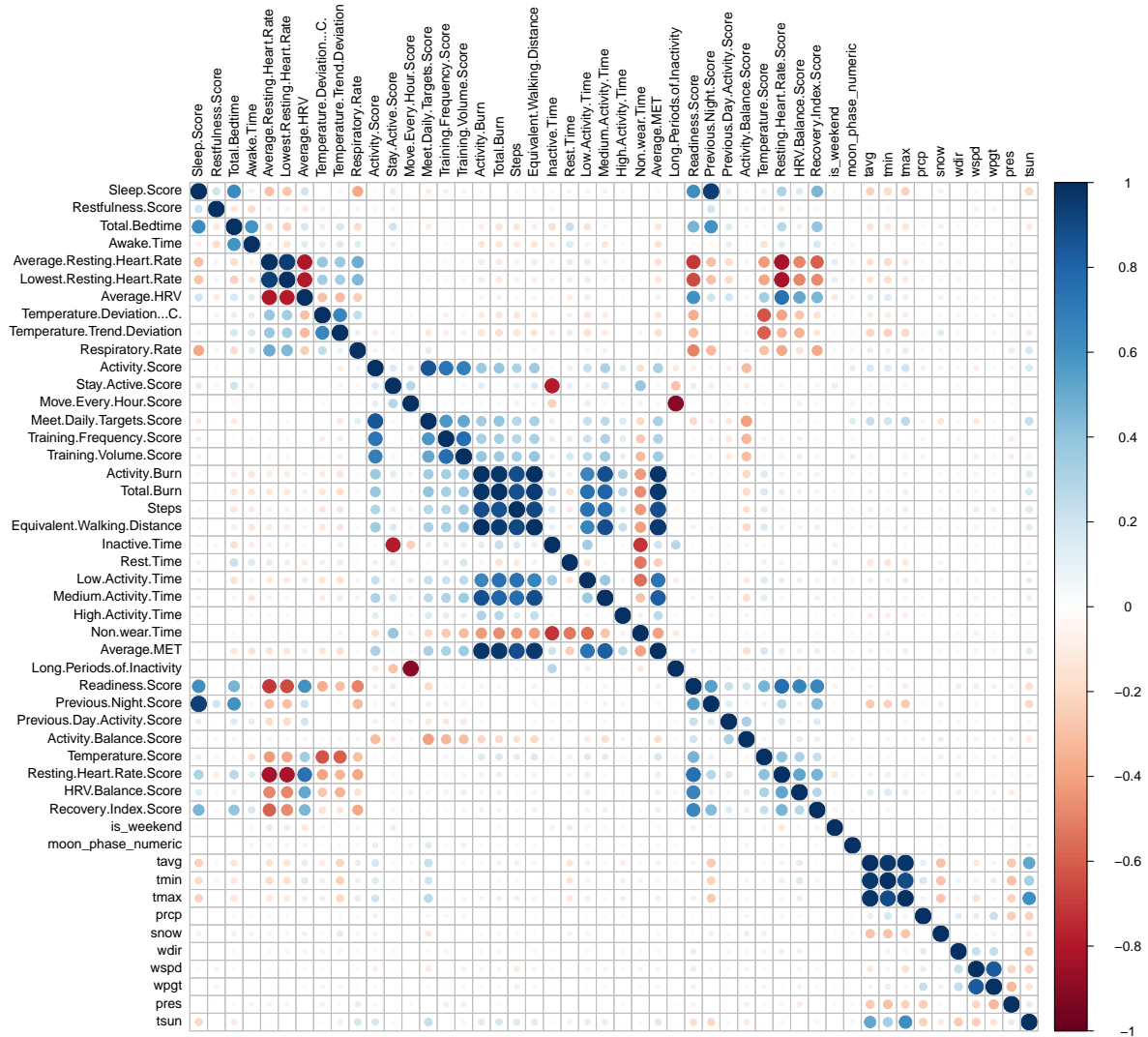


Abbildung 8: Korrelationsmatrix der numerischen Variablen. Die Farben und Kreise zeigen die Stärke und Richtung der Korrelationen an.

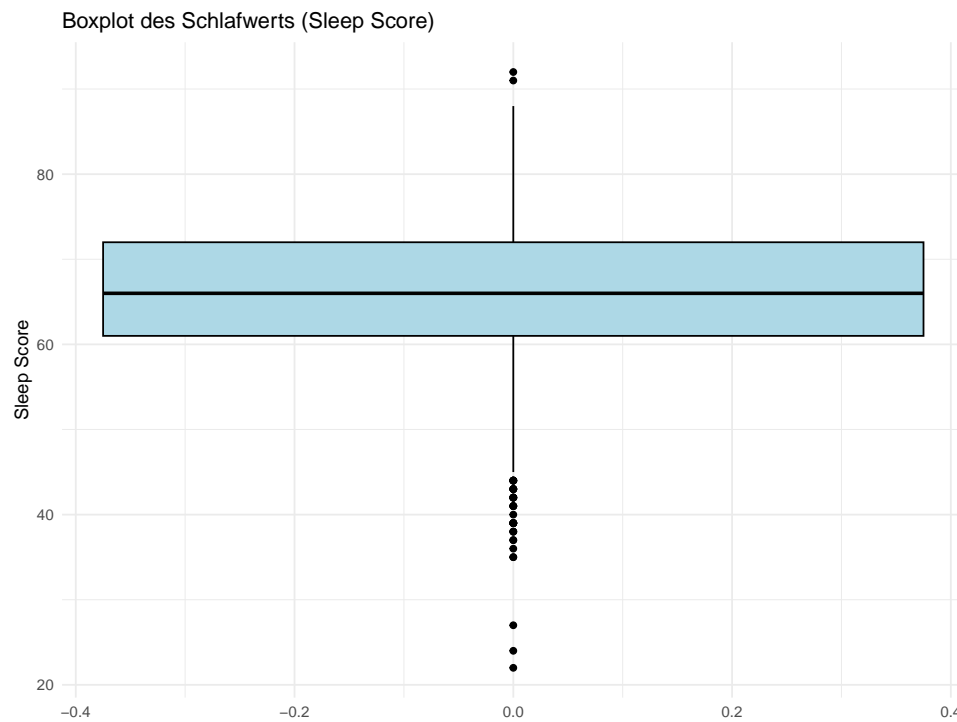


Abbildung 9: Boxplot des Schlafwerts (Sleep Score). Der Plot zeigt die Verteilung der Schlafqualität basierend auf dem Schlafwert, einschließlich möglicher Ausreißer.

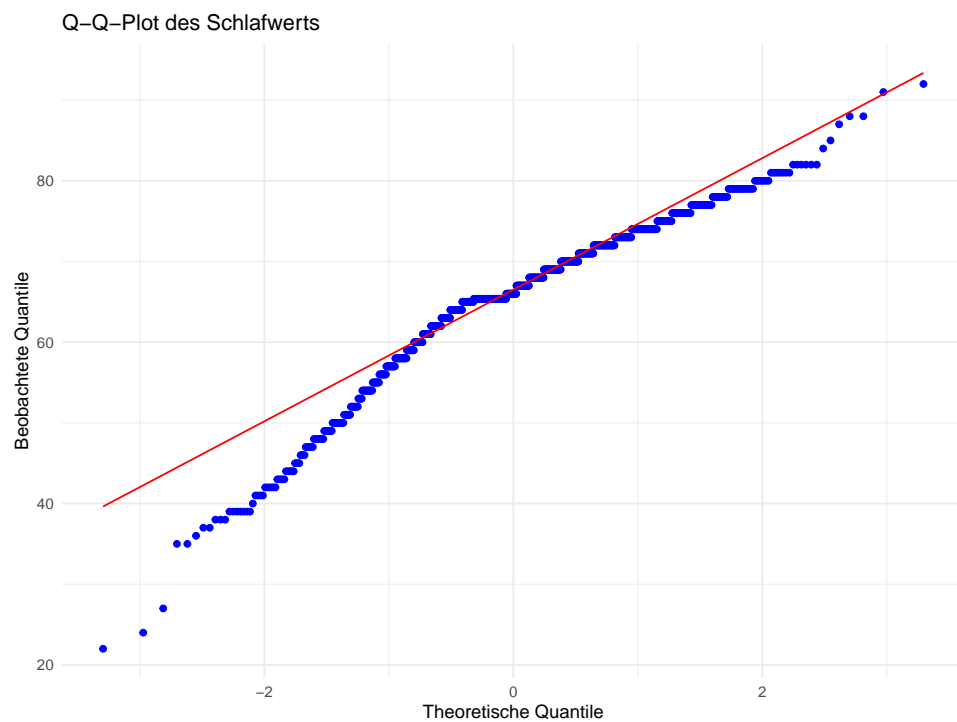


Abbildung 10: Q-Q-Plot des Schlafwerts (Sleep Score). Der Plot zeigt, wie gut die Verteilung des Schlafwerts einer theoretischen Normalverteilung entspricht.

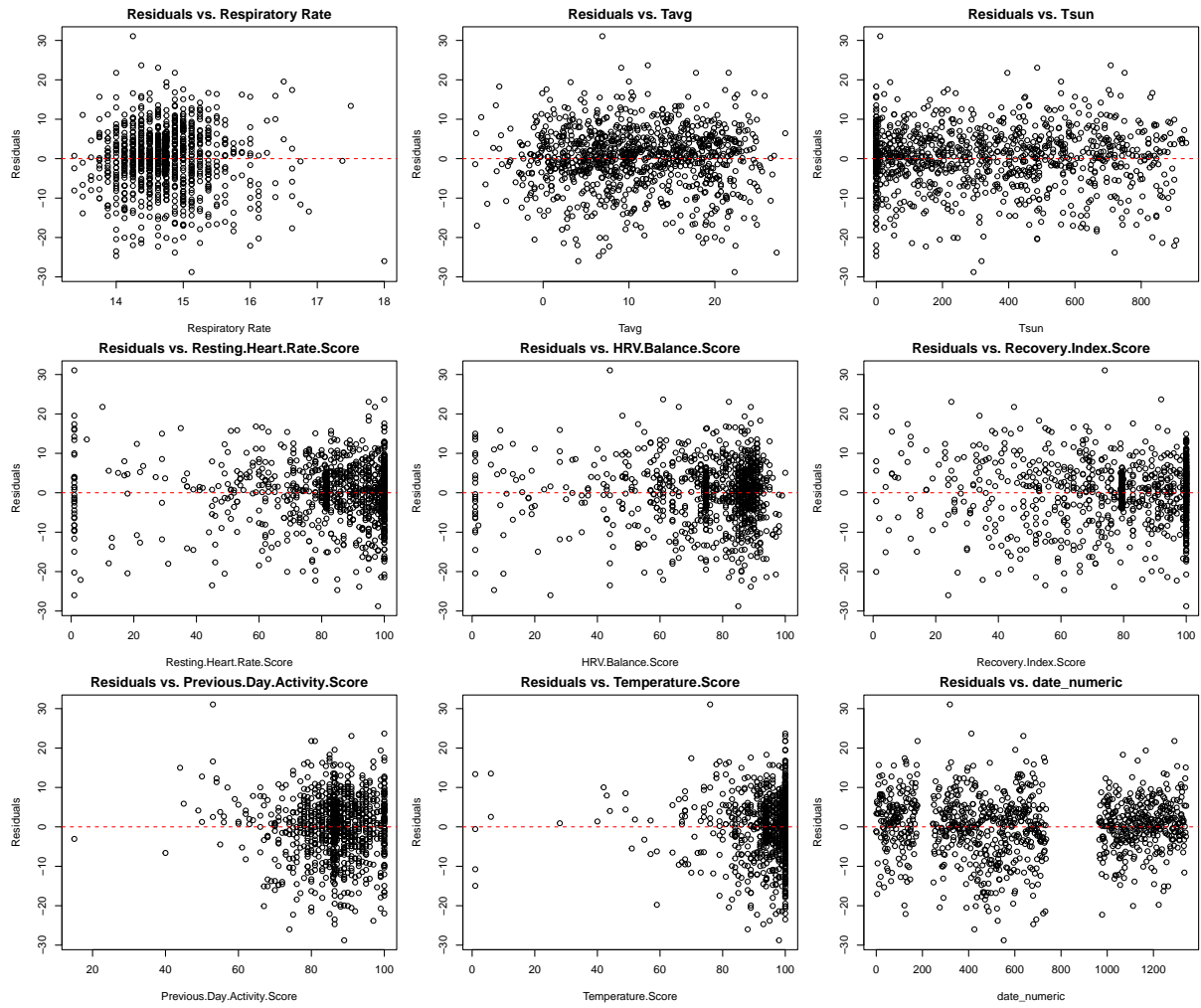


Abbildung 11: Residuen vs. angepasste Werte: Lineares Modell für die Zielvariable Schlafwert. Die roten Linien repräsentieren die Nulllinie, um Abweichungen der Residuen von der Idealverteilung darzustellen.

B Elektronischer Anhang

Für die Implementierung wurde ausschließlich die statistische Software R verwendet. Der vollständige Code und die zugrunde liegenden Daten sind auf GitHub verfügbar: <https://github.com/biancazettler/oura>.

C Abbildungsverzeichnis

1	Dichteplot des Schlafwerts zur Überprüfung der Verteilung.	8
2	Zusammenhang zwischen Schlafwert und Einflussgrößen.	11
3	Korrelationen der ausgewählten Variablen mit dem Schlafwert	13
4	Posterior Predictive Checks der Modelle	25
5	Residuen vs. Angepasste Werte für die drei Modelle	26
6	Korrelationen der ausgewählten Variablen mit der Atemfrequenz	29
7	Posterior Predictive Check des linearen Modells	34

D Tabellenverzeichnis

1	Überblick über die Variablen der Smart-Ring-Daten	4
2	Beschreibung der Wetterdaten für die Region München	4
3	Beschreibung der in den Modellen verwendeten Variablen	5
4	Deskriptive Statistiken der verwendeten Variablen	9
5	Variance-Inflation-Faktor (VIF) der ausgewählten Prädiktoren	15
6	Koeffizienten der Lasso-Regression mit optimalem λ	16
7	Ergebnisse des bayesianischen linearen Modells für den Schlafwert	18
8	Ergebnisse des bayesianischen Modells mit Spline für Atemfrequenz	22
9	Ergebnisse des bayesianischen Modells mit Polynom für die Atemfrequenz	24
10	Ergebnisse des Leave-One-Out Cross-Validation-Kriteriums	26
11	Variance-Inflation-Faktor (VIF) der ausgewählten Prädiktoren	30
12	Koeffizienten der Lasso-Regression mit optimalem λ	31
13	Ergebnisse des bayesianischen linearen Modells für Atemfrequenz	32

E Literaturverzeichnis

- Bloomfield, Leah S. P. u. a. (2022). “Predicting stress in first-year college students using sleep data from wearable devices”. In: *JMIR Mental Health* 9.6, e30323. DOI: [10.2196/30323](https://doi.org/10.2196/30323).
- Bürkner, Paul-Christian (2023). *brms: Bayesian Regression Models using 'Stan'*. R package version 2.19.0, Zugriff am 2. Dezember 2024, <https://CRAN.R-project.org/package=brms>.
- Buuren, Stef van und Karin Groothuis-Oudshoorn (2021). *mice: Multivariate Imputation by Chained Equations in R*. R package version 3.15.0, Zugriff am 2. Dezember 2024, <https://CRAN.R-project.org/package=mice>.
- Cowles, Mary K. und Bradley P. Carlin (1996). “Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review”. In: *Journal of the American Statistical Association* 91.434. Zugriff am 02. Dezember 2024, URL: <http://www.jstor.org/stable/2291683>, S. 883–904.
- Dragulescu, Adrian (2021). *xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files*. R package version 0.6.5, Zugriff am 2. Dezember 2024, <https://CRAN.R-project.org/package=xlsx>.
- Ellis, Andrew (2017). *lunar: Lunar Phase & Distance, Seasons and Other Environmental Factors*. R package version 0.1.4, Zugriff am 2. Dezember 2024, <https://CRAN.R-project.org/package=lunar>.
- Enders, Craig (2010). *Applied Missing Data Analysis*. New York: Guilford Press. ISBN: 978-1-60623-639-0.
- Fahrmeir, Ludwig, Rita Künstler u. a. (2004). *Statistik - Der Weg zur Datenanalyse*. Springer-Verlag. ISBN: 3-540-21232-9.
- Fahrmeir, Ludwig und Alfred Tutz Gerhard und Hamerle (1996). *Multivariate Statistische Verfahren*. 2., erweiterte Auflage. Berlin, New York: Walter de Gruyter. ISBN: 3110138069.
- Fahrmeir, Ludwig und Gerhard Tutz (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer-Verlag.
- Fox, John und Sanford Weisberg (2023). *car: Companion to Applied Regression*. R package version 3.1-2, Zugriff am 2. Dezember 2024, <https://CRAN.R-project.org/package=car>.

- Friedman, Jerome, Trevor Hastie und Rob Tibshirani (2023). *glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models*. R package version 4.1-8, Zugriff am 2. Dezember 2024, <https://CRAN.R-project.org/package=glmnet>.
- Gabry, Jonah und Thomas Mahr (2021). *bayesplot: Plotting for Bayesian Models*. R package version 1.8.1, Zugriff am 2. Dezember 2024, <https://mc-stan.org/bayesplot/>.
- Gelman, Andrew u. a. (2014). *Bayesian Data Analysis*. Chapman & Hall, CRC Press.
- Geyer, Charles J. (2011). “Introduction to MCMC”. In: *Handbook of Markov Chain Monte Carlo*. Hrsg. von Steve Brooks u. a. Zugriff am 02. Dezember 2024, URL: <http://www.mcmchandbook.net/HandbookChapter1.pdf>. Chapman & Hall, CRC Press.
- Handl, Anneliese (2010). *Multivariate Analysemethoden*. 2. Auflage. Springer. ISBN: 9783642149863.
- Hastie, Trevor J., Robert J. Tibshirani und Jerome H. Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Zugriff am 02. Dezember 2024, URL: http://web.stanford.edu/~hastie/local.ftp/Springer/OLD/ESLII_print4.pdf. New York: Springer.
- Held, Leonhard und David S. Bové (2014). *Applied Statistical Inference: Likelihood and Bayes*. Springer-Verlag.
- Henry, Lionel und Hadley Wickham (2023). *rlang: Functions for Base Types and Core R and 'Tidyverse' Features*. R package version 1.1.1, Zugriff am 2. Dezember 2024, <https://CRAN.R-project.org/package=rlang>.
- Imura, Ichiro u. a. (2022). “A method for estimating physician stress using wearable sensor devices”. In: *Sensors and Materials* 34.8, S. 2955–2971. DOI: [10.18494/SAM3908](https://doi.org/10.18494/SAM3908).
- Koskimäki, Heikki u. a. (2018). “How do we sleep: A case study of sleep duration and quality using data from Oura ring”. In: *Proceedings of the 2018 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2018 ACM International Symposium on Wearable Computers*. ACM, S. 714–717. DOI: [10.1145/3267305.3267697](https://doi.org/10.1145/3267305.3267697).
- Lai, Julia u. a. (2021). “Improving sleep quality through mindfulness interventions: A comprehensive study”. In: *Journal of Health and Sleep Research* 13.5, S. 102–116.
- Malakhatka, Elena u. a. (2021). “Monitoring and predicting occupant’s sleep quality by using wearable device OURA ring and smart building sensors data: A living laboratory case study”. In: *Buildings* 11.10, S. 459. DOI: [10.3390/buildings11100459](https://doi.org/10.3390/buildings11100459).

- Massar, Sarah A. A. u. a. (2021). “Trait-like nocturnal sleep behavior identified by combining wearable, phone-use, and self-report data”. In: *npj Digital Medicine* 4, S. 90. DOI: [10.1038/s41746-021-00466-9](https://doi.org/10.1038/s41746-021-00466-9).
- McCullagh, Peter und John Nelder (1989). *Generalized Linear Models*. 2nd ed. Chapman und Hall/CRC.
- Meteostat (2024). *München Wetterdaten*. [Meteostat](https://meteostat.net/en/). Zugriff am 2. Dezember 2024.
- Opoku Asare, Kofi u. a. (2022). “Mood ratings and digital biomarkers from smartphone and wearable data differentiates and predicts depression status: A longitudinal data analysis”. In: *Journal of Affective Disorders* 313, S. 111–122. DOI: [10.1016/j.pmcj.2022.101621](https://doi.org/10.1016/j.pmcj.2022.101621).
- Oura Ring (2024). *Oura Ring - Sleep and Activity Tracker*. Zugriff am 02. Dezember 2024, URL: <https://ouraring.com>.
- Pedersen, Thomas Lin (2022). *patchwork: The Composer of Plots*. R package version 1.1.2, Zugriff am 2. Dezember 2024, <https://CRAN.R-project.org/package=patchwork>.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Zugriff am 02. Dezember 2024, URL: <http://www.r-project.org>.
- Robert, Christian P. (2007). *The Bayesian Choice*. Springer Science+Business Media.
- Roberts, David M. u. a. (2020). “Detecting sleep using heart rate and motion data from multisensor consumer-grade wearables, relative to wrist actigraphy and polysomnography”. In: *Sleep* 43.7, zsaa045. DOI: [10.1093/sleep/zsaa045](https://doi.org/10.1093/sleep/zsaa045).
- Stroup, Walter W. (2012). *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. CRC Press.
- Truda, Giovanni (2020). “Machine learning for sleep data analysis: A case study in quantified self”. In: *Journal of Machine Learning in Sleep Science* 12.4, S. 285–301.
- Vries, H. J. de u. a. (2023). “Wearable-measured sleep and resting heart rate variability as an outcome of and predictor for subjective stress measures: A multiple N-of-1 observational study”. In: *Sensors* 23.1, S. 332. DOI: [10.3390/s23010332](https://doi.org/10.3390/s23010332).
- Wei, Taiyun und Viliam Simko (2023). *corrplot: Visualization of a Correlation Matrix*. R package version 0.92, Zugriff am 2. Dezember 2024, <https://CRAN.R-project.org/package=corrplot>.
- Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. Zugriff am 2. Dezember 2024, <https://ggplot2.tidyverse.org>. Springer-Verlag.

- Wickham, Hadley (2023). *tidyr: Tidy Messy Data*. R package version 1.3.0, Zugriff am 2. Dezember 2024, <https://CRAN.R-project.org/package=tidyr>.
- Wickham, Hadley u. a. (2023a). *tidyverse: Easily Install and Load the 'Tidyverse'*. R package version 2.0.0, Zugriff am 2. Dezember 2024, <https://CRAN.R-project.org/package=tidyverse>.
- Wickham, Hadley u. a. (2023b). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.3, Zugriff am 2. Dezember 2024, <https://CRAN.R-project.org/package=dplyr>.
- Xie, Yihui (2023). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.44, Zugriff am 2. Dezember 2024, <https://CRAN.R-project.org/package=knitr>.

F Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und ohne unzulässige fremde Hilfe angefertigt habe. Alle verwendeten Quellen und Hilfsmittel sind angegeben. Mir ist bekannt, dass ein Verstoß gegen diese Erklärung zur Aberkennung des Bachelor of Science führen kann.

München, 4. Dezember 2024

Bianca Laetitia Zettler