

# **Leistungsnachweis:**

## **Seminar: R you Ready? Reproduzierbare Datenaufbereitung und – analyse in R**

### **1. Rahmenbedingungen**

#### **Anwesenheitspflicht**

Maximal 2x unentschuldigtes Fehlen.

#### **Leistungsbestandteile**

<b>Leistung</b>	<b>Punkte</b>
Mitarbeit	14
Regelmässige Hausübungen	32
Abschlussprojekt	54
<b>Total</b>	<b>100</b>

#### **Notenskala**

<b>Punkte</b>	<b>Note</b>
100-92	6
91-84	5.5
83-76	5
75-68	4.5
67-60	4
< 60	ungenügend

#### **Wichtiger Hinweis:**

In jeder der drei Einzelleistungen (Mitarbeit, Hausübungen, Abschlussprojekt) müssen mindestens 50% der jeweiligen Punkte erreicht werden, um das Seminar erfolgreich zu absolvieren.

Eine Einsicht in den aktuellen Leistungsstand ist jederzeit auf Nachfrage möglich.

### **1.1 Mitarbeiter (14 Punkte)**

Mitarbeit umfasst unter anderem:

- Wortmeldungen im Plenum
- Aktive Teilnahme an Hands-on-Sessions
- Beteiligung im Ilias-Forum

7 Punkte werden durch die Lehrveranstaltungsleitung vergeben.

7 Punkte erfolgen am Ende des Semesters über eine Selbsteinschätzung.

### **1.2 Regelmässige Hausübungen (32 Punkte)**

Bestandteile:

- Erstellung Datenanalyseplan (4P) & Peer-Feedback (4P)

- Erstellung Codebook (4P) & Peer-Feedback (4P)
- 3 R-Übungen (je 4P) & Peer-Feedback zu Übung 2 (4P)

Insgesamt sind dies 8 Abgaben über das Semester hinweg.

Details zu den Hausübungen und deren Deadlines werden in den Einheiten erläutert; einen Überblick liefert der Semesterplan (Stand Februar 2026)

EH	Datum	Inhalte	Hausübung / Deadlines	R for Data Science / Einführung in R (Ellis & Mayer)
1	18.02.26	Organisatorisches; Vorstellung Ebook; Installation R & Pakete, Coding Basics	Bedarfsanalyse: <b>So 22.02.26</b>	Kapitel 2/6 Kapitel 1/2
2	25.02.26	Open Science Praktiken: Datenanalyseplan, Naming Conventions, Coding Basics	4P: Datenanalyseplan: <b>Mi 11.03.26</b>	Kapitel 2/6 Kapitel 1/2
3	04.03.26	Einführung Forschungsdatenmanagement: PsychDS, FAIR, Codebook, Projekte, Öffnen und Speichern von Daten, Coding Basics	Beginn Codebook Open Data	Kapitel 7/20 Kapitel 3.2
4	11.03.26	Datenaufbereitung (Merge); Error Detection & R Ressourcen	4P: Peer-Feedback Datenanalyseplan: <b>Mi 18.03.26</b> Muddiest Points: <b>So 15.03.26</b>	Kapitel 2 Kapitel 2
5	18.03.26	Muddiest Points; Reproduzierbarer Code, Quarto	4P: Codebook: <b>Mi 25.03.26</b>	Kapitel 8
6	25.03.26	Coding Basics: Matrizen, Listen, Dataframes, Faktoren	4P: Peer-Feedback Codebook: <b>Mi 01.04.26</b>	Kapitel 2/6/27 Kapitel 2
7	01.04.26	Abschluss Coding Basics; Datenaufbereitung (filter, subset)	4P: R Übung: <b>Mi 22.04.26</b> (Feedback durch Dozierende)	Kapitel 28/29/5 Kapitel 3
8	15.04.26	Skalenberechnung & - kodierung, Fehlende Werte (TIDY & TRANSFORM)	Fortsetzung	Kapitel: 3/18/12-13/16/26 Kapitel 4
9	22.04.26	Wide/long, Datenqualität: Schiefe, Kurtosis, Streudiagramme, Ausreisser, Skalenreliabilität (ANALYZE); Duplicate Values	Muddiest Points: <b>So 26.04.26</b> Aufarbeitung/Wiederholung/Vertiefung	Kapitel 5/13/16/26 Kapitel 3/4
10	29.04.26	Muddiest Points; Tabellen und Abbildungen	4P: R Übung: <b>Fr 08.05.26</b> 4P: Peerfeedback: <b>Mi 13.05.26</b>	
11	06.05.26	Fortsetzung Abbildungen und Tabellen, Beginn Inferenzstatistik	Fortsetzung	Kapitel 1/11 Kapitel 5
12	13.05.26	Inferenzstatistik: Korrelationen & Regressionen, t-Tests (ANALYZE); Details Abschlussprojekt.	4P: R Übung: <b>Mi 27.05.26</b>	Cheat Sheets I & III Mayer
13	20.05.26	Inferenzstatistik: ANOVAs (ANALYZE)	Fortsetzung, Muddiest Points: <b>So 24.05.26</b>	Cheat Sheet II Mayer
14	27.05.26	Recap; Offene Fragen/Muddiest Points; Selbsteinschätzung Mitarbeit; Evaluation	54P: Abschlussarbeit: <b>So 14.06.26</b>	

### Bewertungsrelevant:

Die Auseinandersetzung mit dem Stoff und den aufgetragenen Aufgaben steht im Vordergrund. Die formale Korrektheit ist zweitrangig.

Für individuelles Feedback oder bei Fragen sind die Dozierenden aktiv zu kontaktieren.

## 1.3 Abschlussprojekt

Bestandteile:

- Überarbeiteter Datenanalyseplan

(Der Datenanalyseplan war bereits Teil einer Hausübung und wird hier erneut bewertet → doppelte Punktevergabe)

b) Überarbeiteter Codebook für den kombinierten Datensatz (dat\_full – simulierte Daten)  
(Das Codebook war bereits Teil einer Hausübung und wird hier erneut bewertet → doppelte Punktevergabe)

- c) Datenfiles der simulierten Daten
  - 7 Einzel-Datensätze
  - Datensatz «merged»
  - Long-Datensatz
- d) Quarto-Analyseskripte
  - Processing (.qmd + gerenderte html-Datei)
  - Analysis (.qmd + gerenderte html-Datei)

### Ziel des Abschlussprojekts

Reanalyse der Studie von Grinschgl et al. (2021) mit simulierten Daten.

Die Datenaufbereitung und -analyse sind im Quarto-Skript nachvollziehbar zu kommentieren.  
Die statistischen Analysen sind im Skript zu interpretieren.

### • Schritte der Datenaufbereitung

Achtung: Hier kann es noch zu Updates kommen! Stand Februar 26

- Einlesen der 7 Einzeldatensätze
- Zusammenführen (mergen)
- Data-Wrangling, z.B.:
  - Rekodierung von Items
    - Gibt es Items, welche noch zu rekodieren sind?
  - Umgang mit fehlenden Werten
    - Gibt es Personen mit fehlenden Werten, die von der Analyse ausgeschlossen werden müssen?
  - Identifikation doppelter Personen im Datensatz
    - Gibt es Personen, welche im Datensatz doppelt vorkommen?
  - Umgang mit unmöglichen Werten
    - Gibt es unmögliche Werte in den Daten (z.B. -999)? Wie können diese Daten sinnvoll ersetzt werden?
  - Überprüfung und Anpassung von Datentypen
    - Gibt es Variablen, welche nicht den korrekten Datentyp haben (z.B. numerische Variablen, die als Charakter definiert sind)? Schaut euch die Variablen genau an (z.B. 3 wird fälschlicherweise als *drei* angezeigt).
  - Umbenennung von Variablen in snake\_case
  - Berechnung neuer Variablen (z.B. mmq\_mean)
  - Umwandlung von Wide- in Long-Format
  - Umwandlung relevanter Variablen in factor() (im Analysis-Skript)

## 2. Durchzuführende Analysen

- Deskriptive Statistik ( $M$  &  $SD$ ) entsprechend Tabelle 1
- 2x3 Mixed ANOVA (subjektive Leistungseinschätzungen) inkl. post-hoc t-Tests und Effektstärken ( $\eta^2$ , Cohen's d)
- One-Way ANOVA für jede Offloading-Variable (3x), inkl. Effektstärke  $\eta^2$
- One-Way ANOVA für Trial Duration (Pattern Copy Task), inkl. Effektstärke  $\eta^2$
- One-Way ANOVA für MMQ inkl. post-hoc t-Tests und Effektstärken ( $\eta^2$ , Cohen's d)
- One-Way ANOVA für Arbeitsgedächtnisleistung (Feature Switch Detection Task), inkl. Effektstärke  $\eta^2$
- Erstellung einer Tabelle oder Abbildung aus der Studie nach Wahl

### Optional:

Testen der Voraussetzungen bei den Analysen. Falls durchgeführt, muss dies im Datenanalyseplan beschrieben und bei Verletzungen angemessen berücksichtigt werden.

## 3. Bewertungskriterien Abschlussprojekt

Die Bewertung orientiert sich an den folgenden Kriterien. Im Fokus steht eine nachvollziehbare, reflektierte und reproduzierbare Umsetzung des Projekts.

### 4.1 Vollständigkeit

Im Rahmen der Bewertung wird geprüft, ob alle Bestandteile des Abschlussprojektes vollständig und konsistent eingereicht wurden. Dabei geht es nicht um die Perfektion einzelner Analysen, sondern darum, ob alle vorgesehenen Elemente vorhanden und inhaltlich bearbeitet wurden. Dazu gehören insbesondere:

- Ein **vollständiger und nachvollziehbarer Datenanalyseplan**, in dem versucht wurde, alle relevanten Fragestellungen systematisch zu bearbeiten (Beschreibung vorgesehener Analysen, Benennung und Definition zentraler Variablen, Darstellung des geplanten Vorgehens und relevanter Datenaufbereitungsschritte)
- Ein **vollständiges Codebook**, welches sämtliche im Datensatz enthaltenen Variablen dokumentiert (Angabe von Variablentyp, Kodierung, Wertebereich und neu berechneter Variablen)
- Alle geforderten **Datenfiles** (7 Einzel-Datensätze, kombinierter Datensatz, Long-Datensatz) in der vorgesehenen Struktur
- Vollständige **Processing- und Analysis-Skripte** (jeweils als .qmd und gerenderte Datei), in denen die geplanten Aufbereitungs- und Analyseschritte umgesetzt sind.

### 4.2 Reproduzierbarkeit

Es wird geprüft, ob das Projekt so dokumentiert ist, dass eine aussenstehende Person die Datenaufbereitung und Analyse nachvollziehen und reproduzieren kann. Dabei wird insbesondere auf Transparenz, Klarheit und Konsistenz zwischen Planung, Datensatzstruktur und Umsetzung geachtet. Relevante Aspekte sind unter anderem:

- Nachvollziehbare Beschreibung des Vorgehens im **Datenanalyseplan** (Datenanalyseplan als klare Anleitung für die spätere Analyse)
- Klare und konsistente Dokumentation der Variablen im **Codebook** (eindeutige und konsistente Variablennamen, klare Dokumentation von Kodierungen, nachvollziehbare Beschreibung von Transformationen)
- Verständliche Benennung und Strukturierung der **Datenfiles**

- Ausreichende Kommentierung und strukturierte Darstellung der Schritte in **Processing und Analysis Skripten** (inkl. nachvollziehbarer Interpretation der Ergebnisse)

### 4.3 Qualität der R-Skripte

Es wird die fachliche und technische Qualität der Quarto-Skripte beurteilt. Dazu zählen insbesondere:

- Sachgerechte Datenaufbereitung (z.B. Rekodierungen, Umgang mit fehlenden/unmöglichen Werten, korrekte Transformationen)
- Korrekte Durchführung und Interpretation der statistischen Analysen
- Technische Funktionsfähigkeit (Code ausführbar ohne manuelle Anpassungen, Verwendung relativer Pfade, erfolgreiches Rendern)
- Sinnvolle Strukturierung (klare Trennung von Processing und Analysis, logisch gegliederter Ablauf)
- Einheitlicher Stil gemäss Style-Guidelines sowie Vermeidung unnötiger Redundanzen

Treten Warn- oder Fehlermeldungen auf, sollen diese sichtbar bleiben und – falls relevant – im Skript nachvollziehbar eingeordnet werden.

### 4.4 Grundprinzip der Bewertung

Im Vordergrund der Bewertung steht die nachvollziehbare und reflektierte Auseinandersetzung mit dem Analyseprozess. Dazu gehört insbesondere:

- Lösungswege werden transparent dokumentiert
- Denkprozesse werden kommentiert und reflektiert
- Auch Unsicherheiten, Schwierigkeiten oder unerwartete Ergebnisse werden reflektiert
- Fehler sind nicht per se negativ, sofern sie nachvollziehbar dokumentiert und kritisch eingeordnet werden

### 4.5 Formelle Vorgaben

Die folgenden formellen Anforderungen sind einzuhalten. Sie dienen der Sicherstellung einer professionellen, nachvollziehbaren und wissenschaftlich angemessenen Darstellung:

- Klare Struktur mit sinnvollen Überschriften zur Gliederung
- Übersichtliche Gestaltung ohne unnötige Formatierungen oder Redundanzen (Orientierung an APA 7)
- Abgabe auf Deutsch oder Englisch
- Die Zeitform ist durchgehend Präsens und die Ausführungen sind in der wissenschaftlichen Wir-Form zu formulieren
- Code-Kommentare begründen Vorgehen und Funktionswahl nachvollziehbar
- Warn- und Fehlermeldungen werden nicht unterdrückt, sondern transparent dargestellt und interpretiert
- Die Nutzung von LLMs ist zulässig, muss aber transparent dokumentiert und mit der Abgabe eingereicht werden (Link oder separates Dokument); die Verantwortung bleibt vollständig bei den Studierenden
- Tabellen/Abbildungen aus Grinschgl et al. (2021) werden sinngemäß und vollständig repliziert (kleine Formatabweichungen zulässig),
- Nur Inhalte mit direktem Bezug zur Endabgabe werden eingereicht; nicht relevante Übungs-/Hands-on-Inhalte sind zu entfernen.
- Die Gliederung folgt einer logisch nachvollziehbaren Reihenfolge

- Ungewöhnliche oder inkonsistente Formatierungen sind zu vermeiden
- Redundanzen (z.B. doppelte Codeblöcke oder unnötige Wiederholungen) sind zu entfernen

## 4. Organisatorische Angaben

**Deadline Abschlussprojekt:** **14.06.2026**

**Abgabe:** Upload via Ilias als ZIP-Ordner.

- Der ZIP-Ordner enthält:
  - Die vorgegebene Ordnerstruktur
  - Alle geforderten Dateien
  - Die Dokumentation der Nutzung von LLMs
- Benennung des ZIP-Ordners: vorname\_nachname\_abschlussarbeit.zip

Für individuelles Feedback zum Abschlussprojekt sind die Dozierenden aktiv zu kontaktieren.

## 5. Richtlinien zur Zusammenarbeit mit Peers

Die Abschlussarbeit ist eine individuelle Leistung.

### Erlaubte Zusammenarbeit

- Austausch über allgemeine methodische Fragestellungen (z.B. Wahl statistischer Verfahren)
- Diskussion über Strukturierung von Skripten
- Gemeinsames Besprechen und Verstehen von Fehlermeldungen
- Peer-Feedback zu Klarheit, Struktur und Vollständigkeit von Datenanalyseplan oder Codebook

### Nicht erlaubte Zusammenarbeit

- Austausch vollständiger R-Skripte, Quarto-Dokumente oder Codebooks
- Kopieren von Formulierungen oder Analyseplänen
- Gemeinsames vollständiges Erarbeiten der Datenaufbereitung oder Analyse

Die Eigenständigkeit der Arbeit muss jederzeit gewährleistet sein.