

TAKE HOME EXAMINATION

MATHEMATISCHE METHODEN

WS 22/23 1. PRÜFUNGSZEITRAUM

Ronny Schomacker

VORGABEN

Nennen Sie im Titel/Kopf ihrer Take Home Examination ihre **Matrikelnummer** und ihre **Variablenauswahl**!

Die Take Home Examination sollte als **PDF-Datei** abgegeben werden und die Zeichenzahl von **17.000 Zeichen** inklusive Leerzeichen nicht überschreiten! Für die Zusatzaufgaben können 5000 Zeichen zusätzlich in Anspruch genommen werden. Die erstellten **Diagramme und Karten** sowie wichtige berechnete **Parameter** sollen im Text wiedergegeben werden!

Der Bericht sollte **selbständig** und ohne Hilfe anderer von ihnen eigenhändig erzeugt werden! Sollten Bericht- und Codeelemente bei zwei und mehr Abgaben Auffälligkeiten besitzen, die kaum durch Zufall erklärt werden können und damit auf eine nicht selbständige Arbeit schließen lassen, werden die betroffenen Berichte als nicht bestanden gewertet!

Layouten sie den Bericht und die Abbildungen und folgen sie der durch die Teilaufgaben vorgegebenen **Struktur (Gliederung, Aufgabennummerierung)**. Betten sie die Abbildungen (Diagramm) in das PDF-Dokument ein und vergeben sie Abbildungsnummern auf die sie sich im Text beziehen. Die Bewertung für ein ansprechendes **Layout und die Gestaltung** des Berichtes sowie eine übersichtliche R-Code Strukturierung und Kommentierung geht unabhängig von den unten gestellten Aufgaben mit einem Gewicht von 0.08 ein.

ABGABE

Geben Sie einen vollständig lauffähigen, strukturierten und kommentierten R-Code aller Teilaufgaben in der Form **Matrikelnummer.R** (alternativ auch **Matrikelnummer.Rmd**) ab! Fügen sie die erstellten R-Objekte (Daten) als **Matrikelnummer.rds** und **Matrikelnummer.csv** hinzu. Geben Sie die Lösung der Aufgaben in Form eines Berichtes ab: **Matrikelnummer.pdf** (bzw. in Rmd-Datei)!

Geben Sie eine unterschriebene **Eigenständigkeitserklärung** für den Bericht digital ab.

Abgabedatum: siehe Moodle!

BEWERTUNG:

- Summe der Gewichte: 1
- Endnote Take Home Examination = Summe(Teilnotengewichte * Teilnote) – Summe(Zusatzaufgabengewichte * Bewertung Zusatzaufgabe)
- Endnote Modul = Endnote Take Home Examination – 0.3 * Bewertung der Semesteraufgaben – 0.1 * Seminararbeit
- Generelle ZUSATZAUFGABE: Für besondere, brauchbare, individuelle Lösungen im Bericht, Diagramm und im R-Code werden weitere Zusatzpunkte vergeben. (bis max. 0.3 Gewichtspunkte)

AUFGABE

Erstellen Sie eine Abfrage bei INKAR (www.inkar.de) mit den separat in Moodle ausgewählten Variablen:

Variable 1: _____

Variable 2: _____

1. Laden Sie die Daten auf Kreisebene als **CSV-Datei (Matrikelnummer.csv)** herunter und bearbeiten Sie diese in R so, dass sie als dataframe-Objekt vorliegen. Speichern Sie dieses Objekt als **Matrikelnummer.rds** ab!

Gewicht R-Code	Gewicht Text
0.08	0

2. Erstellen Sie von ihren vorgegebenen Variablen die empirische Verteilungsfunktion (**ECDF**)! Entscheiden und Begründen sie aufgrund der univariaten Verteilungsform (Symmetrie) der Variablen, ob sie eine **Transformation** der Daten durchführen. Transformieren Sie die Daten gegebenenfalls unter Angabe der Formel und speichern sie diese als neue Variable ab!
 - a. ZUSATZ: bestimmen sie die Schiefe ihrer Daten.
 - b. Arbeiten sie mit den transformierten Variablen in den nachfolgenden Schritten weiter!

Gewicht R-Code	Gewicht Text
0.08/0.02 (Schiefe)	0.04

3. Entfernen sie die **NA-Werte mit R** (auch wenn keine NA-Werte vorhanden sind, sollte der Code durchlaufen!) und bilden sie ein entsprechendes neues **Subset**.
 - a. Für Daten die in **Aufgabe 2 nicht transformiert** werden: Erläutern sie kurz, wie NA-Werte prinzipiell gefüllt werden könnten. Welche Risiken und welche Chancen bestehen dabei?

Gewicht R-Code	Gewicht Text
0.02	0.04 (NA Alternativen)

4. Entfernen sie mithilfe eines geeigneten Verfahrens (Nennen!) die **Ausreißer** ihres (transformierten) Datensatzes, mit dem sie in den folgenden Schritten weiterarbeiten! Bessere ALTERNATIVE (schwierig): Sie können in den folgenden Schritten (6.) auch mit (Ausreißer) bivariat (!) robusten Verfahren weiterarbeiten!

Gewicht R-Code	Gewicht Text
0.04	0

5. Erzeugen sie ein **Punkstreudiagramm** (Scatterplot) der beiden unter 3. und 4. verarbeiteten Variablen in **R**! Entscheiden sie sich für die abhängige und unabhängige Variable. Begründen Sie ihre Entscheidung! Beschreiben sie die Beziehung der Variablen anhand der Muster (Anordnung) im Diagramm.
 - a. Für robuste Schätzung ZUSATZAUFGABE: markieren sie die Punkte in geeigneter Form, die zur robusten Modellierung herangezogen werden!

Gewicht R-Code	Gewicht Text
0.02/0.04 (robust)	0.06

6. Diskutieren sie anhand des Erscheinungsbildes des **Punkstreudiagramms**, inwiefern eine lineare Regression sinnvoll sein könnte (**Kriterien der Regression**)!

Gewicht R-Code	Gewicht Text
0.04	0.08

7. Führen sie die **Lineare Regression** unabhängig von ihrer (eventuell negativen) Bewertung in Aufgabe 6 in R durch und interpretieren Sie **Güte und Aussage ihres Regressionsmodells** quantitativ anhand der Modellergebnisse im erzeugten Regressionsobjekt.

Gewicht R-Code	Gewicht Text
0.04/+0.04 (robust)	0.04

8. Geben Sie die **lineare Regressionsgleichung** und gegebenenfalls die rücktransformierte Regressionsgleichung im Text an. Interpretieren sie diese — In welcher Relation stehen die beiden Variablen in den Landkreisen von Deutschland danach generell zueinander?

Gewicht R-Code	Gewicht Text
0	0.04

9. Verknüpfen Sie die ungefilterten und untransformierten Daten mit Geometriedaten der Kreise mithilfe von beispielsweise `merge()` in R. Bereiten sie die Daten gegebenenfalls so vor, dass dieser Schritt möglich ist.

Gewicht R-Code	Gewicht Text
0.02	0

10. Erstellen Sie eine **ansprechende Karte** für die beiden Variablen in R, beispielsweise in **tmap**! Achten Sie dabei insbesondere auf die Legende. Beachten Sie unbedingt auf eine sinnvolle Klassenbildung (Anzahl nach STURGES und Klassengrenzenbildungsverfahren (beispielsweise gibt die eigene Transformation ein Hinweis auf das Klassenbildungsverfahren)), passen sie dabei die Klassengrenzen auf leicht verständliche Zahlenwerte an (pretty).

Gewicht R-Code	Gewicht Text
0.06	0.04

11. Beschreiben Sie beobachtbare räumliche Muster (Nord/Süd, Ost/West, städtisch/ländlich, ...) der einzelnen Variablen (univariat)! Interpretieren sie ihre Beobachtung mit ihren Erfahrungen für die räumlich hervortretenden univariaten Muster!

Gewicht R-Code	Gewicht Text
0	0.08

12. Berechnen sie die Schätzwerte für die abhängige Variable nach dem Regressionsmodell (Achtung Rücktransformation!) im dataframe! Berechnen Sie damit Residuen!
- a. ZUSATZ: Wie bewerten sie die Eigenschaften der von ihnen berechneten Residuen bezüglich der Güte der (rücktransformierten) Regression! Diagramme sind erwünscht!

Gewicht R-Code	Gewicht Text
0.04	0

13. Erstellen Sie eine **ansprechende Karte** für die **Residuen ihrer linearen Regression** in R! Was sagt ihnen das in Erscheinung tretende räumliche Muster der Residuen auf der Karte? Welche Konsequenzen ergeben sich für die räumliche Güte des Linearen Regressionsmodells?

Gewicht R-Code	Gewicht Text
0.04	0.06

Viel Erfolg und gutes Gelingen!