

[GK] DaSci 7.2 Explorative Datenanalyse und Visualisierung – 4h

Done: View

Done: Make a submission

Done: Receive a grade

Opened: Monday, 18 September 2023, 12:00 AM

Due: Wednesday, 20 December 2023, 12:00 AM

Data Science ``Explorative Datenanalyse und Visualisierung" - Taskdescription

Einführung

Die Theorieeinheiten sollen als Einführung in die explorative Datenanalyse und Visualisierung dienen. Zur Einführung in R, RStudio oder Python ist der Link aus den Theorie Einheiten empfohlen.

Ziele

Das Ziel ist es das Grundwerkzeug zur Datenexploration und Visualisierung zu verstehen, einsetzen und interpretieren zu können.

Voraussetzungen

- Kenntnis von Markdown, LaTeX und Grundkenntnisse in R und Python
- funktionstüchtige Installation von R, RStudio, Python auf eurem Rechner, virtueller Maschine etc.
- Kenntnis von Variablentypen und Schätzern bzw. Methoden zur Informationszusammenfassung für unterschiedliche Variablentypen (Häufigkeiten, Lage, Streuung, etc.)
- Kenntnis von Möglichkeiten zur Visualisierung unterschiedlicher Datentypen

Aufgabe 1 (GK)

Exploration numerischer Daten (Wirtschaftsdaten der US Bundesstaaten)

1. Lade den Datensatz 'state.x77' in R. Beschreibe die Daten anhand der internen Hilfe.
2. Ermittle mithilfe geeigneter Schätzer für die Lage (arithmetischer Mittelwert und Median sollen verglichen werden) und Streuung (Standardabweichung und Interquartilsdistanz sollen verglichen werden) der ersten 5 Variablen: Population, Income, Illiteracy, Life Exp(ectancy) und Murder.
3. Stelle die Daten der ersten 5 Variablen, Population, Income, Illiteracy, Life Exp(ectancy) und Murder in geeigneter Weise graphisch dar, indem du Boxplot, Histogramm mit Dichteschätzung, ECDF und QQ-Plot verwendest.
4. Begründe anhand der graphischen Darstellung, ob es sich symmetrische oder schiefe Datenverteilungen handelt. Begründe anhand der graphischen Darstellungen, ob schwere oder leichte Ränder vorliegen (und auf welcher Seite).
5. Bestimme anhand graphischen Darstellungen aus Punkt 3. und der Erkenntnisse aus 4., ob Ausreißer vorliegen und welche Punkte dies sind.

Anfrage 2 (EK)

Exploration kategorialer Daten (Überlebende der Titanic)

1. Lade den Datensatz 'Titanic', welche im R base Paket inkludiert ist. Beschreibe die Daten anhand der internen Hilfe. (empfohlen sind dafür R Befehle: '?', help() )
2. Untersuche den "Titanic" Datensatz mithilfe von Tabellen von absoluten und relativen Häufigkeiten der Ränder heruntergebrochen auf je 2 oder 3 Dimensionen von Variablen: Stelle dabei die Überlebenden aufgeteilt nach dem Geschlecht, nach dem Alter und nach der Klasse als zweidimensionale Tabelle dar. Gib die Aufteilung der Überlebenden nach Geschlecht und Klasse zugleich, nach Alter und Klasse zugleich wieder. (empfohlen sind dafür die R Befehle: table(), ftable(), colSums, rowSums(), apply(), xtable())
3. Stelle die Daten in geeigneter Weise mithilfe von barplots und mosaicplots graphisch dar. (empfohlen sind dafür R Befehle: barplot(), mosaicplot(), assocplot())
4. Interpretiere die Tabellen und ihre dazugehörigen Grafiken. Begründe anhand der quantitativen Ergebnisse, ob Frauen und Kinder zuerst bei der Seentrettung berücksichtigt wurde.
5. Erkläre, wo sich Anzeichen für Simpson's Paradoxon finden lassen.

Abgabe

Das Protokoll ist als PDF-Dokument abzugeben vorzulegen, welches die graphischen Darstellungen und Interpretationen in ganzen deutschen Sätzen enthält.

Bewertung

Gruppengröße: 1-2 Person(en)

Anforderungen überwiegend erfüllt

- aktuelle Markdown- oder LaTeX-Protokollvorlage aus Github bzw. Moodle verwendet
- grundlegende Beschreibung und Verwendung der im Unterricht angeführten Begriffe: ordinale, nominal, metrisch rational, metrisch intervallskaliert; absolute und relative Häufigkeiten; Mittelwert, Median, Varianz; Standardabweichung, Interquartilsdistanz; Symmetrie, Schiefe, schwere Ränder, Ausreißer; Histogramm, Boxplot, QQ-Plot
- Codebeispiele referenziert
- Dokumentation aller Inhalte des DataCamp Kurses

Anforderungen zur Gänze erfüllt

- Verbale Beschreibung und Erklärung aller angeführter Begriffe und deren Anwendung in konkreten Beispielen in vollständigen deutschen Sätzen
- ausführliche Codebeispiele und Visualisierungen dokumentiert

Quellen

Edit submission

Remove submission

Submission status

Submission status	Submitted for grading	
Grading status	Graded	
Time remaining	Assignment was submitted 34 days 12 hours late	
Last modified	Tuesday, 23 January 2024, 12:55 PM	
File submissions	<div><div></div><div>Explorative.pdf</div><div>23 January 2024, 12:55 PM</div></div>	
Submission comments	<div><div></div><div>Comments (0)</div></div>	

Feedback

Grade	100.00 / 100.00
Graded on	Tuesday, 23 January 2024, 2:52 PM
Graded by	<div></div> Posekany Alexandra