

TUTORIUM

Datenauswertung

DATENMANAGEMENT

AGENDA

- Installation: car
- Datenmanagement mit Vektoren
 - Umcodieren: einzelner Wert, mehrere Werte, Wertebereiche
 - Mathematische und logische Operationen
- Datenmanagement mit Faktoren
 - Umcodieren: Reihenfolge und Ausprägungen ändern
 - Level ohne Nennungen entfernen
 - Unterdatensatz erstellen

INSTALLATION: CAR

```
Console Terminal x
~/
> install.packages("car")
Installing package into 'C:/Users/vitus/Documents/R/win-library/3.5'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/car_3.0-2.zip'
Content type 'application/zip' length 1482141 bytes (1.4 MB)
downloaded 1.4 MB

package 'car' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\vitus\AppData\Local\Temp\Rtmp23X7pM\downloaded_packages
> library(car)
Lade nötiges Paket: carData
Warning message:
Paket 'car' wurde unter R Version 3.5.3 erstellt
> |
```

R UND VARIABLEN – BASICS

	Vektor erstellen	Faktor erstellen
Muster	<code>name <- c(Zahl₁, Zahl₂, ... Zahl_n)</code>	<code>name <- factor(c(„Wort₁“, „Wort₂“, ... „Wort_n“))</code>
Beispiel	<code>alter <- c(21, 20, 30)</code>	<code>geschlecht <- factor(c(„m“, „f“, „f“))</code>

- Wörter schreibt man in Anführungszeichen, Zahlen ohne
- Variablen innerhalb eines Datensatzes wählt man mit dem Dollarzeichen (\$) aus aus

VEKTOREN – EINZELNEN WERT ÄNDERN

Einen einzelnen Wert ändern

- Ist sinnvoll bei Tippfehlern
- Läuft über den Befehl mit dem Muster

`Datensatz$variable[Zeilennummer] <- neuer Wert`

- Beispiel: `beispiel$Haushaltsgröße[10]<- 2`

VEKTOREN – ALLE VORKOMMNISSE EINES WERTS ÄNDERN

Der Recode-Befehl

- Ermöglicht es, *alle Vorkommnisse eines Wertes* in einer Variable durch einen anderen Wert zu ersetzen
- Hat dieses Muster: `recode(datensatz$variable, 'alter Wert1 = neuer Wert1; alter Wert2 = neuer Wert2, ...')`
- Beispiel: `beispiel$Minuten.TV.Konsum <- recode (beispiel$Minuten.TV.Konsum, '120=NA')`

VEKTOREN – ALLE VORKOMMNISSSE EINES WERTEBEREICHS ÄNDERN

Wertebereiche

– Definiert man mit dem recode-Befehl und folgenden Operatoren:

Operator	Bedeutung	Beispiel
Doppelpunkt	von ... bis ...	1:60 = 1
lo (für „lowest“)	vom niedrigsten Wert bis ...	lo:60 = 1
hi (für „highest“)	von ... bis zum höchsten Wert	301:hi = 6
else	Für alles noch nicht Umcodierte	1:60=1, 61:120=2, else=3

– Beispiel: `beispiel$Minuten.TV.Konsum <-
recode(beispiel$Minuten.TV.Konsum, 'lo:60="wenig"')`

WERTE IN VEKTOREN ÄNDERN – AUFGABE

Aufgaben:

- Ladet die Datei „Beispieldatensatz.RData“ aus dem Learnweb herunter, speichert sie in eurem Arbeitsverzeichnis und öffnet sie in R
- Setzt den NA-Wert für die Variable Altersgruppe auf „25-34 Jahre“
- Setzt alle NAs der Variable Radiokonsum mit einem Befehl auf 0
- Erstellt eine Variable, die den Internetkonsum in Stunden statt in Minuten angibt

RECHNEN MIT VEKTOREN

Zeichen

+

*

^

exp(x)

sum(x)

Bedeutung

Addition

Multiplikation

Potenz

Exponentialfunktion

Summe

Zeichen

-

/

sqrt(x)

log(x)

abs(x)

Bedeutung

Subtraktion

Division

Wurzel

Natürlicher
Logarithmus

Absoluter Wert

LOGISCHE OPERATIONEN MIT VEKTOREN

Zeichen	Bedeutung	Zeichen	Bedeutung
==	Gleich	!=	Ungleich
>	Größer	<	Kleiner
>=	Größer gleich	<=	Kleiner gleich
&	Und		oder

Mathematische und Logische Operationen funktionieren auch für Vektoren!
Für Operationen, die zwei oder mehr Vektoren benötigen, müssen aber alle Vektoren gleich lang sein (d. h. gleich viele Fälle enthalten).

MIT VEKTOREN RECHNEN – AUFGABEN

Aufgaben:

- Erstellt eine Variable, die den gesamten Medienkonsum pro Person (TV, Radio, Internet, Zeitung) wiedergibt
- Findet mit einem einzelnen Befehl heraus, wie viele Personen mehr als 15 Minuten Zeitung pro Tag lesen
- Erstellt eine Variable, die in Prozent angibt, wie lange eine Person pro Tag das Internet nutzt

FAKTOREN: REIHENFOLGE VON AUSPRÄGUNGEN ÄNDERN

Der Factor-Befehl

- wird benutzt, um die Ausprägungen einer kategorialen Variable und deren Reihenfolge festzulegen
- Hat folgendes Muster: `factor(variable, levels=c(
„Ausprägung1“, „Ausprägung2“, ... , „Ausprägungn“))`
- Beispiel: `beispiel$Geschlecht <-
factor(beispiel$Geschlecht, levels=c("Frau", "Mann"))`

FAKTOREN: LEVEL OHNE NENNUNGEN LÖSCHEN

Level ohne Nennungen löschen

- Macht Sinn, wenn eine der theoretisch denkbaren Ausprägungen in keinem Fall realisiert wurde
- Läuft nach diesem Muster: `datensatz <- within(datensatz, {variable <- droplevels(variable)})`
- Beispiel: `beispiel <- within(beispiel, {Geschlecht <- droplevels(Geschlecht)})`

FAKTOREN – DATENSATZ NACH GRUPPEN AUFTEILEN

Die Subset-Funktion

- Wählt einen Teil des Datensatzes aus
- Ist sinnvoll für Gruppenbildung
- Funktioniert nach dem Muster: `datensatz_neu <- subset(datensatz_alt, variable == „Ausprägung“)`
- Beispiel: `beispiel_single <- subset(beispiel, Partnerschaft=="Single")`

FAKTOREN – AUFGABEN

Aufgaben:

- Ändert die Reihenfolge der Ausprägungen der Variable Partnerschaft in „Single“, „Liiert“
- Erstellt einen Datensatz, der nur diejenigen Personen enthält, die angegeben haben, zwischen 15 und 24 Jahre alt zu sein

WIEDERHOLUNG – VARIABLEN

Eine Variable

- Ist eine Eigenschaft, die man messen will
- Muss mindestens zwei Ausprägungen haben
- Ausprägungen sind die möglichen Werte, die eine Variable annehmen kann
- Beispiele:
 - Variable: Gereiztheit
 - Ausprägungen: sehr niedrig, niedrig, mittel, hoch, sehr hoch
 - Variable: Fernsehnutzung (täglich)
 - Ausprägungen: 0 Minuten, 1-30 Minuten, 31-60 Minuten, ...

WIEDERHOLUNG – SKALENNIVEAUS

	Skalenniveau	Beschreibung	Mathematische Operationen	Messbare Eigenschaften	Beispiele
Kategoriale Variablen	Nominalskala	Unterscheidung /Klassifizierung	$=/\neq$	Häufigkeit (Modus)	Geschlecht, Beruf
	Ordinalskala	Nominal + Reihen-/Rangfolge	$=/\neq, </>$	Häufigkeit (Modus), Reihenfolge (Median)	Schulnoten, Platzierung beim Wettbewerb
metrische Variablen	Intervallskala	Ordinal + gleiche Abstände	$=/\neq, </>, +/-$	Häufigkeit (Modus), Reihenfolge (Median), Abstand (Mittelwert)	Datum, Temperatur (in Celsius)
	Ratioskala	Intervall + natürlicher Nullpunkt	$=/\neq, </>, +/-, \bullet/\div$	Häufigkeit (Modus), Reihenfolge (Median), Abstand (Mittelwert), Verhältnisbildung	Alter, Fernsehnutzung pro Tag

WIEDERHOLUNG – HYPOTHESEN

Eine Hypothese

- Ist eine wissenschaftliche Behauptung über den Zusammenhang bzw. Unterschied mindestens zweier Variablen
- Sollte theoretisch fundiert sein
- Muss empirisch überprüfbar sein → kritischer Rationalismus

FORSCHUNGSBERICHT

- Ziel: Abgabe eines mit R Markdown erstellten Forschungsberichts bis zum 30.06.2019, 23:59 Uhr
- Vorlage für Bericht abhängig davon, ob man Ein-Fach- oder Zwei-Fach-Bachelor ist
- Übersicht zu Aufgaben: siehe erste PPP
- Erläuterungen zu den Aufgaben befinden sich in den dazugehörigen Markdowns

FRAGEN?

SELBSTSTUDIUM

- Ein-Fach-Bachelor: Bericht für die Stichprobe (3.3)
- Datenmanagement für daten_2019\$www_minuten

BIS NÄCHSTE WOCH!