

Computergestützte Datenanalyse: DATA-Übung mit R

Tag 3 – 24.07.2025

UNSER PLAN

- **Tag 1**

- Einführung in R und RStudio
- „Basics“:
- Coding Konventionen
- Objekte, Datenimport & Co

- **Tag 2**

- Skalenniveau
- Troubleshooting
- Datenaufbereitung
- Datenvisualisierung
- Deskriptive Statistik

- **Tag 3**

- Inferenzstatistik I
- Bivariate Analyse

- **Tag 4**

- Indexbildung
- Inferenzstatistik II
- Abschluss

Genereller Ablauf

- Vier Tage geblockt
- Mischung aus Input- und Übungssessions
- Anwesenheitsabfrage alle 90 Minuten

Heute

- Zwei 15 Minuten Pause
- Eine Mittagspause

REFRESHER

- Welche Datentypen/-klassen gibt es?
- Was ist der Unterschied zwischen Subsetting und Rekodieren?
Welche Packages, Funktionen?
- Was untersucht der Chi^2 -Test?
- Was ist der Unterschied zwischen Cramer's V und Pearson's R?

REFRESHER

- Was wird hier gerechnet?
- Wie interpretieren wir diesen Output?

```
cor.test(allbus$pt15,  
allbus$pt03)
```

```
##  
## Pearson's product-moment correlation  
##  
## data: allbus$pt15 and allbus$pt03  
## t = 41.087, df = 3323, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not e  
qual to 0  
## 95 percent confidence interval:  
## 0.5574159 0.6025157  
## sample estimates:  
## cor  
## 0.5804107
```

Getting started

1. Rmarkdown file erstellen und einstellen
2. Packages installieren und/ oder aktivieren: rio und tidyverse
3. Import der ALLBUS 2018 Daten
4. In einem **neuem** Dataframe werden die Variablen pt01 bis pt20 gespeichert.
Nutzt dafür den Befehl `select()` und recherchiert, wie das funktioniert

Indexbildung und Reliabilitätsanalyse

Indexbildung

- Manchmal müssen wir, um ein Konstrukt angemessen darzustellen, mehrere Einzelindikatoren zusammenfassen
- Erfassung vieler theoretische Merkmale eines abstrakten Begriffs
 - Beispielsweise setzt sich das Phänomen der **Lebensqualität** aus den Einzelindikatoren Glück, Zufriedenheit und Wohlbefinden zusammen
- Reduktion sozialer Erwünschtheit
 - Beispielsweise Messung rechtsextremer Einstellung
- Reduktion des Messfehlers
 - Testtheorie: Reliabilität definiert als Genauigkeit, „mit der eine Skala ein Merkmal misst.“ (Rammstedt, 2010, S. 242)

Voraussetzung

- Alle Variablen müssen in die gleiche Richtung kodiert sein (ggf. rekodieren)
- Beispiel „Big Five“ zur Messung von Persönlichkeit (Rammstedt et al., 2014)
 - „Extraversion“ als eine Dimension der Persönlichkeit, wird über zwei Einzelindikatoren abgefragt
 - Für die Antworten der Befragungsperson steht eine fünfstufige Ratingskala von "trifft überhaupt nicht zu" (1) bis "trifft voll und ganz zu" (5) zur Verfügung.
 - „Ich bin eher zurückhaltend, reserviert.“ (negative Polung)
 - „Ich gehe aus mir heraus, bin gesellig.“ (positive Polung)

Voraussetzung

- Reliabilitätsanalyse: Interne Konsistenz im Antwortverhalten
 - Können wir die Einzelindikatoren zu einem Index zusammenfassen?
- Maßzahl Cronbachs Alpha
 - Ergibt sich aus der Anzahl der Einzelindikatoren und der (durchschnittlichen) Korrelation
 - Nimmt maximal den Wert 1 an
 - Unser Cut-Off-Kriterium > 0.7
 - Alpha steigt allerdings mit der Anzahl der Indikatoren
 - Es kann sein, dass sich Alpha stark verbessert, wenn wir ein Item aus dem Index weglassen (ggf. theoretisch begründen!)

Aufgabe

Korrelationsmatrix:

```
allbus_sub <- allbus %>%  
select(px01:px10)
```

Um welche Variablen handelt es sich?

```
cor(allbus_sub,  
     use = "pairwise.complete.obs",  
     method = "pearson")
```

Was seht ihr?

Cronbachs Alpha

- Wir nutzen die Funktion `alpha()` aus dem Psych-Package

Reliability analysis

Call: `alpha(x = df[trust], check.keys = TRUE)`

raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
0.9	0.9	0.91	0.4	8.8	0.0025	4.2	0.94	0.38

Reliability if an item is dropped:

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r	med.r
pt01	0.90	0.90	0.92	0.43	8.9	0.0025	0.016	0.39	
pt02	0.89	0.89	0.91	0.40	8.1	0.0027	0.017	0.37	
pt03	0.88	0.88	0.90	0.39	7.7	0.0029	0.015	0.36	
pt04	0.90	0.90	0.91	0.42	8.6	0.0026	0.017	0.39	
pt08	0.89	0.89	0.91	0.40	7.9	0.0028	0.018	0.37	
pt09	0.90	0.89	0.91	0.41	8.5	0.0026	0.016	0.39	
pt10	0.89	0.89	0.90	0.41	8.2	0.0027	0.017	0.37	
pt11	0.89	0.89	0.91	0.41	8.5	0.0026	0.018	0.38	
pt12	0.88	0.88	0.90	0.39	7.6	0.0029	0.015	0.36	
pt14	0.89	0.89	0.91	0.41	8.3	0.0026	0.018	0.38	
pt15	0.89	0.89	0.91	0.39	7.8	0.0028	0.016	0.37	
pt19	0.89	0.88	0.89	0.39	7.7	0.0028	0.013	0.37	
pt20	0.89	0.89	0.90	0.39	7.7	0.0028	0.013	0.37	

Aufgabe

Korrelationsmatrix:

```
allbus_sub <- allbus %>%  
select(px01:px10)
```

Um welche Variablen handelt es sich?

```
cor(allbus_sub,  
     use = "pairwise.complete.obs",  
     method = "pearson")
```

Was seht ihr?

Aufgabe

Korrelationsmatrix:

```
install.packages("psych")
```

```
library(psych)
```

```
alpha(allbus_sub)
```

Was seht ihr?

raw_alpha <dbl>	std.alpha <dbl>	G6(smc) <dbl>	average_r <dbl>	S/N <dbl>	ase <dbl>	mean <dbl>	sd <dbl>	median_r <dbl>
0.8176307	0.8176685	0.8364781	0.3096076	4.484517	0.004531764	2.253593	0.6997273	0.3046075

	raw_alpha <dbl>	std.alpha <dbl>	G6(smc) <dbl>	average_r <dbl>	S/N <dbl>	alpha se <dbl>	var.r <dbl>	med.r <dbl>
px01	0.8179796	0.8199486	0.8247519	0.3359879	4.553970	0.004544392	0.01240284	0.3182501
px02	0.8064941	0.8076461	0.8136710	0.3181173	4.198751	0.004793231	0.01703049	0.3141702
px03	0.8118608	0.8119047	0.8300114	0.3241444	4.316454	0.004704338	0.01869410	0.3141702
px04	0.7947232	0.7934270	0.8113661	0.2991147	3.840903	0.005106118	0.01963501	0.2861827
px05	0.8070169	0.8065751	0.8252992	0.3166268	4.169964	0.004811441	0.01953043	0.3075234
px06	0.7889937	0.7902014	0.8119407	0.2950286	3.766476	0.005362700	0.02031510	0.2745536
px07	0.7943422	0.7946189	0.8163511	0.3006449	3.868998	0.005153922	0.01971158	0.3041859
px08	0.7892032	0.7891120	0.8025894	0.2936664	3.741854	0.005272843	0.01640828	0.2997131
px09	0.7910886	0.7897658	0.8021216	0.2944829	3.756600	0.005214541	0.01621413	0.2997131
px10	0.8077998	0.8077501	0.8287183	0.3182625	4.201562	0.004824466	0.02018835	0.3015367

Indexbildung

- Sind die Variablen richtig kodiert und Cronbachs Alpha entsprechend hoch, können wir unseren Index bilden (zumindest aus einer ersten methodischen Perspektive)
 - Mittelwertindex
 - Additiver Index

Mittelwertindex

- Der Mittelwertindex wird errechnet, in dem der Mittelwert aus den Einzelitems der Skala gebildet wird.
 - Mit `rowMeans()`
 - Oder mit `mutate`

```
allbus$? <- rowMeans(allbus_sub, na.rm = FALSE)
```

```
allbus$? <- allbus %>%  
  mutate(? = rowMeans(across(px01:px10), na.rm = FALSE)))
```

Additiver index

- Der Mittelwertindex wird errechnet, in dem der Mittelwert aus den Einzelitems der Skala gebildet wird.
 - Mit `rowMeans()`
 - Oder mit `mutate`

```
allbus$? <- rowSums(allbus_sub, na.rm = FALSE)
```

```
allbus <- allbus %>%  
  mutate(? = rowSums(across(px01:px10), na.rm = FALSE)))
```


Additiver index

- Der Mittelwertindex wird errechnet, in dem der Mittelwert aus den Einzelitems der Skala gebildet wird.
 - Mit `rowMeans()`
 - Oder mit `mutate`

```
allbus$? <- rowSums(allbus_sub, na.rm = FALSE)
```

```
allbus <- allbus %>%  
  mutate(? = rowSums(across(px01:px10), na.rm = FALSE)))
```

Standardisierung, Additiver Index

- Der Mittelwertindex wird errechnet, in dem der Mittelwert aus den Einzelitems der Skala gebildet wird.
 - Mit `rowMeans()`
 - Oder mit `mutate`

```
allbus$? <- rowSums(allbus_sub, na.rm = FALSE)
```

```
allbus <- allbus %>%  
  mutate(? = rowSums(across(px01:px10), na.rm = FALSE)))
```

Weiterführende Methoden

- Explorative Faktorenanalyse (EFA)
- konfirmatorische Faktorenanalyse (CFA)
- Mit lavaan-Package in R

Aufgabe

- Ein Index für Vertrauen in Institutionen?
- Können Sie einen Index zum allgemeinen Vertrauen erstellen?
- Verbessert sich die Reliabilität, wenn Sie ein Item weglassen?
- Mittelwert und Additiven Index erstellen!
- Plus: Handelt es sich bei "Vertrauen" um ein eindimensionales Konzept? Und könnte es mit der Anzahl an Variablen zu Problemen mit Cronbach's Alpha kommen?

REFRESHER

- Warum wollen wir einen Index bilden?
- Was sind Voraussetzungen?

FRAGEN, UNKLARHEITEN, FEEDBACK?

VIELEN DANK 😊

ahrabhi.kathirgamalingam@cais-research.de