

Übung 04

Hypothesen testen

Korrelation, Wilcoxon-test und χ^2 in R

INFI-IS

5xHWII

Albert Greinöcker, Thomas Kefer

December 9, 2021



CCA - COMPETENCE CENTRE

HTL Anichstraße

1 Hypothesen testen

So wie im Student Performance - Datensatz besprochen soll nun ein etwas größerer Fragebogen kennengelernt und teilweise ausgeweitet werden. Ein Download des Datensatzes plus einem PDF in dem genaue Beschreibungen zu den Variablen zu finden sind, befindet sich im Moodle unter: european social survey.

1.1 Importieren des Datensatzes

Die entsprechende Datei entpacken und die Datei ESS8e02.1_F1.csv so wie im gemeinsamen Skript zum European Social Survey beschrieben, importieren.

```
1 d <- read.csv("../datasets/ESS8e02.1_F1.csv", sep=";", encoding="UTF-8")
```

Im Datensatz sind die einzelnen Variablen in Zahlen kodiert (so wie es sich gehört). Will man diesen Zahlen Beschriftungen zuordnen, macht man das auf diese Art:

```
1 d$gndr <- factor(d$gndr, levels=c(1,2), labels = c("Male","Female"))
2 d$vote <- factor(d$vote, levels=c(1,2,3,7,8,9), labels = c("Yes","No", "Not eligible to vote", "Refusal", "Don't know", "No answer"))
```

Listing 1: Hier wird z.B. der Variable gndr (Geschlecht) die Werte "Male", "Female", "No Answer" für die Zahlen 1, 2, 9 zugeordnet.

In einzelnen Auswertungen wird der Vergleich zwischen 2 Ländern oder die Situation in nur einem Land untersucht. Aus diesem Grund müssen Subsets mit nur diesen Ländern gebildet werden. Das geht z.B. so:

```
1 d_at <- d[d$entry == 'AT',]
2 d_at_it <- d[d$entry %in% c('AT','IT'),]
```

Listing 2: Hier wird 1x ein Datensatz aller Österreicher und 1x ein Datensatz mit Österreichern und Italienern erzeugt.

Leider kann es dann sein, wenn - wie im Fall der Länder - schon Beschriftungen automatisch zugeordnet wurden, dass unbenutzte Labels (hier: die Länderkürzel für nicht mehr vorhandene Länder) in der Auswertung auftauchen. Diese kann man so löschen:

```
1 d_at_it$cntry <- droplevels(d_at_it$cntry)
```

1.2 Hypothesen Testen

In den eigentlichen Aufgabenstellungen sollen bestimmte Hypothesen getestet werden. Es soll bei jeder Hypothese...

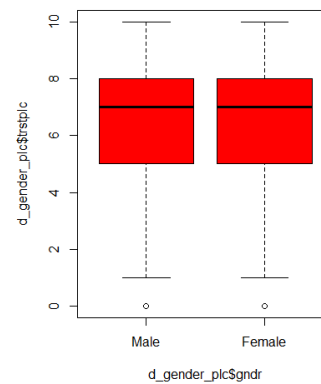
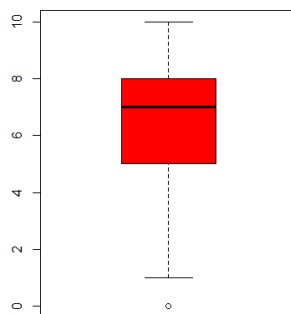
- a. eine Aufbereitung der Daten gemacht werden,
- b. eine grafische Veranschaulichung (sofern möglich) gemacht werden,
- c. ein passender Hypothesentest ausgewählt und ausgeführt werden und
- d. eine Interpretation der Ergebnisse erstellt werden.

Hier die Hypothesen (in Klammern die betroffenen Variablen):

- a. Männer haben mehr Glauben in die Polizei (trstplc, gender).

```
1 table(d_gndr_2$gndr,d$trstplc)
2 d_gender_plc <- d[d$trstplc <= 10,]
3 table(d_gender_plc$gndr,d_gender_plc$trstplc)
4 boxplot(d_gender_plc$trstplc, col="red")
5 boxplot(d_gender_plc$trstplc~d_gender_plc$gndr, col="red")
6
7 wilcox.test(trstplc~gndr, data=d_gender_plc)
```

		0	1	2	3	4	5	6	7	8	9	10
1	Male	783	422	722	1015	1164	2642	2517	3700	4208	2258	1497
3	Female	622	347	696	1144	1320	3083	2866	4158	4711	2499	1684



```
Wilcoxon rank sum test with continuity correction
data: trstplc by gndr
W = 238346250, p-value = 0.005222
alternative hypothesis: true location shift is not equal to 0
```

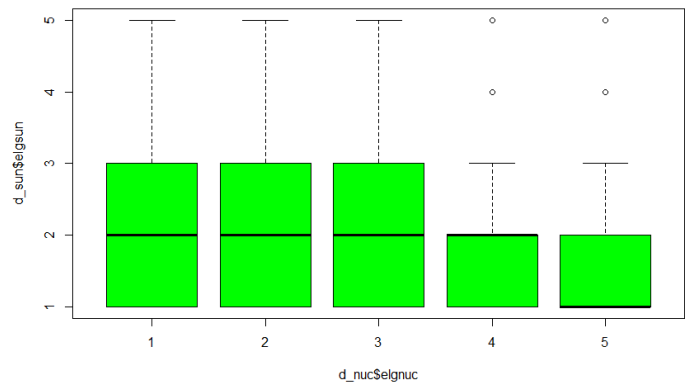
- b. Es besteht ein negativer Zusammenhang bei "mehr Strom aus nuklearer Energie" und "mehr Strom aus Solarenergie" (elgnuc, elgsun).

```

1 d$elgsun <- factor (d$elgsun , levels=c(1,2,3,4,5))
2 d$elgnuc <- factor (d$elgnuc , levels = c(1,2,3,4,5))
3
4 d_sun <- d[as.numeric(d$elgsun) <= 5,]
5 d_nuc <- d[as.numeric(d$elgnuc) <= 5,]
6
7 d_sun$elgsun <- droplevels(d_sun$elgsun)
8 d_nuc$elgnuc <- droplevels(d_nuc$elgnuc)
9
10 boxplot(d_sun$elgsun~d_nuc$elgnuc , col="green")
11 cor.test(as.numeric(d_sun$elgsun),as.numeric(d_nuc$elgnuc) , method="kendall")

```

	1	2	3	4	5
1	1137	1143	2059	3660	9140
2	479	1886	2395	3694	5886
3	400	937	1535	1211	1892
4	295	470	568	565	903
5	46	52	72	88	221



Kendall's rank correlation tau

```

data: as.numeric(d_sun$elgsun) and as.numeric(d_nuc$elgnuc)
z = -36.255, p-value < 2.2e-16
alternative hypothesis: true tau is not equal to 0
sample estimates:
tau
-0.1536407

```

- c. In Österreich ist der Eindruck, dass sich der Klimawandel schlecht auf die Menschen auswirkt, stärker als in Ungarn (ccgdbd).

```
1 d_at_hu <- d[d$cntry %in% c("AT", "HU"),]  
2 d_at_hu <- d_at_hu[as.numeric(d_at_hu$ccgdbd) <= 10,]  
3 d_at_hu$cntry <- droplevels(d_at_hu$cntry)  
4 table(d_at_hu$cntry, d_at_hu$ccgdbd)  
5 wilcox.test(ccgdbd~cntry, data=d_at_hu)
```

	0	1	2	3	4	5	6	7	8	9	10
AT	209	173	329	361	288	230	89	86	71	20	13
HU	267	183	269	246	177	206	78	55	17	3	3

Wilcoxon rank sum test with continuity correction

```
data: ccgdbd by cntry  
W = 1579664, p-value = 3.86e-10  
alternative hypothesis: true location shift is not equal to 0
```

d. Frauen stimmen einem bedingungslosen Grundeinkommen eher zu (basinc).

```
1 d_basicnc <- d[as.numeric(d$basinc) <=7,]
2 d_basicnc$basinc <- droplevels(d_basicnc$basinc)
3 table(d_basicnc$gndr, d_basicnc$basinc)
4 boxplot(d_basicnc$basinc ~ d_basicnc$gndr, col="blue")
5 cor.test(as.numeric(d_basicnc$basinc), as.numeric(d_basicnc$gndr), method="kendall")
```

	1	2	3	4	7
Male	2527	6527	8690	1815	105
Female	2361	7165	9569	1934	130

Bitte noch 2 Hypothesen selbst wählen und überprüfen!