

Bericht (für Fachkolleg:innen)

Moritz Laßnig, Moritz Schrom, Claudia Vötter

4. Dezember 2022

Kurzfassung

TODO:

Einleitung

Ausgangssituation

Die Tierkreiszeichen (im Allgemeinen vereinfacht Sternzeichen genannt) beschreiben zwölf gleich große Abschnitte im Tierkreis, einer 20 Grad breiten Zone um die Ekliptik (scheinbare Sonnenbahn). Sie werden noch heute häufig in der Astrologie für z.B. Horoskope verwendet, und spalten damit die Meinungen, ob durch sie valide Prognosen gestellt werden können. In dieser Kontroverse gründet die Motivation unserer Forschung, um dem oft sehr emotional geführten Diskurs ein empirisch untermauertes Argument hinzuzufügen.

Forschungsfrage: Gibt es einen Zusammenhang zwischen dem Sternzeichen einer Person und ihren Charaktereigenschaften?

Forschungshypothese: In vergleichbaren Gruppen von Menschen mit demselben Sternzeichen wird eine Gruppe eine stärkere oder schwächere Ausprägung einer Charaktereigenschaft aufweisen.

Nullhypothese: Es gibt keine signifikanten Unterschiede in den beobachteten Charaktereigenschaften zwischen vergleichbaren Gruppen von Menschen mit demselben Sternzeichen. Sternzeichen und Charaktereigenschaften sind unabhängig.

Alternativhypothese: Es gibt signifikante Unterschiede in beobachteten Charaktereigenschaft zwischen vergleichbaren Gruppen von Menschen mit demselben Sternzeichen. Zumindest eine Charaktereigenschaft ist bei zumindest einer Gruppe besonders stark oder schwach ausgeprägt. Sternzeichen und Charaktereigenschaft sind abhängig.

Es handelt sich hierbei um eine *vorhersagende Forschungsfrage*, da untersucht wird was passiert, wenn sich eine Variable (Sternzeichen) ändert. Es wird die Beeinflussung dieser Variable auf die zweite Variable (Charaktereigenschaft) untersucht. Dazu werden Daten erhoben, mit welcher die Hypothese getestet wird.

Des Weiteren handelt es sich um eine *korrelative Forschungshypothese*, da nur ein möglicher Zusammenhang zwischen den Variablen (welcher zum Beispiel in Form eines Modells für Prognosen verwendet werden könnte) wird. Unsere Forschungshypothese zielt nicht darauf ab einen kausalen Zusammenhang zwischen dem Sternzeichen und Charaktereigenschaft zu be- oder widerlegen.

Für die Charaktereigenschaften einer Person gibt es viele potenzielle Störfaktoren. Zum Beispiel die Sozialisation der Person, welche eine Anpassung an gesellschaftliche Denk- und Gefühlsmuster durch Verinnerlichung von sozialen Normen beschreibt. Wichtige Beiträge für die Erforschung der Entwicklung von Persönlichkeitsmerkmalen in der frühen Kindheit, Jugend bis hin zum mittleren und höheren Lebensalter (daher

über die gesamte Lebensspanne) kommen vor allem aus Humangenetik, Entwicklungsbiologie, Ethnologie und Kulturanthropologie, sowie aus sonstigen Arbeitsrichtungen der Psychologie. Für unsere konkrete Forschungshypothese hat dies jedoch keinen direkten Einfluss, da davon auszugehen ist, dass sich Störfaktoren auf die Gesamtheit auswirken, und nicht nur auf eine Gruppe mit einem Sternzeichen. Auf mögliche Störfaktoren, welche erwartete Verteilungen der Ausprägungen beeinflussen, wird in der deskriptiven Analyse genauer eingegangen. Ein weiterer Störfaktor ist die Person, welche die Daten erhebt. Um dies zu erläutern, wird im folgenden kurz auf die Methode der Datenerhebung und das Forschungsdesign im Allgemeinen eingegangen.

Forschungsdesign

Für die Beantwortung der Forschungsfrage wurden Daten von den Autor:innen gesammelt. Um Unschärfen in der Selbsteinschätzung zu verhindern, bewerteten die Autor:innen Personen aus dem Umfeld nach den Myers-Briggs Typenindikatoren. Es handelt sich dabei um vier binären Kategorien, welche unterschiedliche Charaktereigenschaften von Menschen kategorisieren. Zusätzlich wurde der Geburtstag (Tag und Monat) erfasst. Die Erfassung wurde in einer für alle Autor:innen zugänglichen geschützten Datenbank durchgeführt. Da es sich hierbei am Ende des Tages um eine subjektive Einschätzung handelt, ist die einschätzende Person ein potenzieller Störfaktor. Um diesen Störfaktor auszugleichen, wurde von allen Autor:innen gleichermaßen zum Datensatz beigetragen. Dies gleicht auch Unterschiede im Umfeld der erhebenden Personen aus. Eine weitere Fehlerquelle sind Doppelerfassungen im Datensatz. Da sich der Bekanntenkreis der Autor:innen sich kaum überlappt, und aktiv darauf geachtet wurde keine Duplikate zu erfassen, stellt dies nach Meinung der Autor:innen jedoch kein Problem dar.

Zwar wird von keinen signifikanten Unterschieden vom Einfluss (oder Nicht-Einfluss) von Sternzeichen auf Charaktereigenschaft in unterschiedlichen Altersgruppen und sozialen Kreisen ausgegangen, jedoch ist dies zum aktuellen Stand nicht empirisch belegt. Daher können wir nur von einer Grundpopulation ausgehen, für welche die Stichprobe repräsentativ ist. Es handelt sich bei den erfassten Personen um größtenteils österreichische Student:innen oder Akademiker:innen im Alter von 20-35 Jahren. Um schnelle und falsche Schlüsse auf eine zu große Grundgesamtheit bei unserem Stichprobenumfang zu vermeiden, bezieht sich die Grundpopulation auf genau ebendiesen beschriebenen Personenkreis.

Operationalisierung

Wie bereits allgemein beschrieben, werden Informationen zum Sternzeichen und zu den Charaktereigenschaften von Personen nach dem Myers-Briggs Typenindikator mit Hilfe einer geteilten Datenbank gesammelt. Im konkreten bedeutet dies, dass primär folgende Variablen mit den genannten Ausprägungsformen erfasst werden:

- Geburtsdatum (Tag und Monat, z.B. 04.11.)
- Charaktereigenschaften:
 - Extraversion / Introversion
 - Intuition / Sensorik
 - Denken / Fühlen
 - Wahrnehmen / Beurteilen

Es handelt sich bei den vier Merkmalen welche die Charaktereigenschaften von Personen beschreiben jeweils um kategoriale Variablen mit binärer Ausprägung: Jede der vier Kategorien wird in zwei mögliche Ausprägungen unterteilt. So kann die Variable "Extraversion / Introversion" entweder den Wert "Extraversion" oder "Introversion" annehmen, jedoch nie beide oder keinen der beiden möglichen Werte. Dies kann im Folgenden auch kodiert werden:

- Charaktereigenschaft1: Extraversion (0) ... Introversion (1)
- Charaktereigenschaft2: Intuition (0) ... Sensorik (1)

- Charaktereigenschaft3: Denken (0) ... Fühlen (1)
- Charaktereigenschaft4: Wahrnehmen (0) ... Beurteilen (1)

In der Datenerfassung wird ein verpflichtendes Auswahlfeld verwendet, welches den Wertebereich auf die beschriebenen binären Ausprägungen einschränkt.

Beim Geburtsdatum handelt es sich um ein verpflichtendes Datumsfeld, welches den Wertebereich auf gültige Tages und Monatswerte festlegt. Im folgenden wird aus dem Geburtsdatum das Tierkreiszeichen abgeleitet. Da das Geburtsjahr für die Bestimmung des Tierkreiszeichens nicht notwendig ist, wird dieses nicht erfasst.

Es handelt sich beim Tierkreiszeichen um eine kategoriale Variable, mit zwölf möglichen Ausprägungen:

- Widder (21.03. - 20.04.)
- Stier (21.04. - 21.05.)
- Zwillinge (22.05. - 21.06.)
- Krebs (22.06. - 22.07.)
- Löwe (23.07. - 22.08.)
- Jungfrau (23.08. - 22.09.)
- Waage (23.09. - 22.10.)
- Skorpion (23.10. - 22.11.)
- Schütze (23.11. - 20.12.)
- Steinbock (21.12. - 19.01.)
- Wassermann (20.01. - 18.02.)
- Fische (19.02. - 20.03.)

Um den Datensatz auch hinsichtlich möglichen Störvariablen zu untersuchen, werden außerdem folgende Metadaten erfasst:

- Erfassende Person (Freitextfeld welcher mit dem Namen der erfassenden Person befüllt wird, in unserem Fall der Vor- und Nachname einer der drei Autor:innen)
- Erfassungsdatum (Tag, Monat und Jahr, z.B. 09.01.2023)

Statistische Methoden

TODO

Datenmanagement

Zur Untersuchung gelangt eine Stichprobe mit TODO Beobachtungen und fünf Merkmalen:

- Geburtsdatum, Tag und Monat im Format TT.MM. (*geburtstag*)
- Charaktereigenschaft 1, binäre kategoriale Variable mit den Ausprägungen: Extraversion / Introversion (*extraversion_introversion*)
- Charaktereigenschaft 2, binäre kategoriale Variable mit den Ausprägungen: Intuition / Sensorik (*intuition_sensorik*)
- Charaktereigenschaft 3, binäre kategoriale Variable mit den Ausprägungen: Denken / Fühlen (*denken_fuehlen*)
- Charaktereigenschaft 4, binäre kategoriale Variable mit den Ausprägungen: Wahrnehmen / Beurteilen (*wahrnehmen_beurteilen*)

Die Daten wurden in einer geteilten Datenbank gesammelt, liegen in einer exportierten CSV-Datei vor, und werden mit `read.table()` eingelesen:

```
daten = read.table("daten.csv", header = TRUE, sep=",")
head(daten)
```

```
##   id geburtstag extraversion_introversion intuition_sensorik denken_fuehlen
## 1  1    05.06.          Introversion          Sensorik          Denken
## 2  2    14.03.          Introversion          Intuition          Fühlen
## 3  3    29.11.          Extraversion          Sensorik          Fühlen
## 4  4    22.05.          Introversion          Intuition          Fühlen
## 5  5    10.07.          Extraversion          Intuition          Denken
## 6  6    23.12.          Extraversion          Sensorik          Denken
##   wahrnehmen_beurteilen   erstellt_von   erstellt_am
## 1                Wahrnehmen Moritz Schrom 04.12.2022
## 2                Beurteilen Moritz Schrom 04.12.2022
## 3                Beurteilen Moritz Schrom 04.12.2022
## 4                Wahrnehmen Moritz Schrom 04.12.2022
## 5                Wahrnehmen Moritz Schrom 04.12.2022
## 6                Beurteilen Moritz Schrom 04.12.2022
```

Vorerst wird überprüft, ob es fehlende Werte gibt:

```
any(is.na(daten))
```

```
## [1] FALSE
```

```
summary(daten)
```

```
##           id           geburtstag           extraversion_introversion intuition_sensorik
##  Min.      : 1.0      Length:71      Length:71      Length:71
## 1st Qu.:18.5      Class :character      Class :character      Class :character
##  Median :36.0      Mode  :character      Mode  :character      Mode  :character
##   Mean      :36.0
## 3rd Qu.:53.5
##   Max.      :71.0
## denken_fuehlen   wahrnehmen_beurteilen   erstellt_von   erstellt_am
## Length:71      Length:71      Length:71      Length:71
## Class :character      Class :character      Class :character      Class :character
## Mode  :character      Mode  :character      Mode  :character      Mode  :character
##
##
##
```

Es gibt keine fehlenden Werte. Jedoch fällt beim `summary()` der Daten auf, dass einige der vorliegenden Variablen noch nicht den gewünschten Datentyp entsprechen. So sollte *geburtstag* ein `Date` (statt `character`) sein, sowie die 4 Charaktereigenschaften als Faktor vorliegen. Dazu müssen wir die eingelesenen Daten umwandeln.

Zuerst werden die Charaktereigenschaften mit `as.factor()` in einen Faktor umgewandelt:

```

daten$extraversion_introversion<-as.factor(daten$extraversion_introversion)
daten$intuition_sensorik<-as.factor(daten$intuition_sensorik)
daten$denken_fuehlen<-as.factor(daten$denken_fuehlen)
daten$wahrnehmen_beurteilen<-as.factor(daten$wahrnehmen_beurteilen)
summary(daten)

```

```

##          id          geburtstag      extraversion_introversion intuition_sensorik
## Min.      : 1.0      Length:71      Extraversion:46      Intuition:35
## 1st Qu.:18.5      Class :character      Introversion:25      Sensorik :36
## Median :36.0      Mode  :character
## Mean      :36.0
## 3rd Qu.:53.5
## Max.      :71.0
## denken_fuehlen wahrnehmen_beurteilen erstellt_von      erstellt_am
## Denken:39      Beurteilen:38      Length:71      Length:71
## Fühlen:32      Wahrnehmen:33      Class :character      Class :character
##                                     Mode  :character      Mode  :character
##
##
##

```

Aus technischen Gründen liegen Metadaten vor, welche nicht benötigt werden (z.B. `erstellt_von`). Diese werden im nächsten Schritt entfernt:

```

daten<-daten[,-c(7,8)]
summary(daten)

```

```

##          id          geburtstag      extraversion_introversion intuition_sensorik
## Min.      : 1.0      Length:71      Extraversion:46      Intuition:35
## 1st Qu.:18.5      Class :character      Introversion:25      Sensorik :36
## Median :36.0      Mode  :character
## Mean      :36.0
## 3rd Qu.:53.5
## Max.      :71.0
## denken_fuehlen wahrnehmen_beurteilen
## Denken:39      Beurteilen:38
## Fühlen:32      Wahrnehmen:33
##
##
##
##

```

Die Variable `geburtstag` liegt im Format TT.MM. vor und muss in ein `Date` umgewandelt werden.

```

daten$geburtstag<-as.Date(daten$geburtstag, format='%d.%m.')
summary(daten)

```

```

##          id          geburtstag      extraversion_introversion
## Min.      : 1.0      Min.      :2023-01-04      Extraversion:46
## 1st Qu.:18.5      1st Qu.:2023-04-03      Introversion:25
## Median :36.0      Median :2023-07-15

```

```
## Mean :36.0 Mean :2023-07-09
## 3rd Qu.:53.5 3rd Qu.:2023-10-13
## Max. :71.0 Max. :2023-12-26
## intuition_sensorik denken_fuehlen wahrnehmen_beurteilen
## Intuition:35 Denken:39 Beurteilen:38
## Sensorik :36 Fühlen:32 Wahrnehmen:33
##
##
##
##
```

Da das Jahr für die Ermittlung des Sternzeichens keine Relevanz hat, ist die Annahme des aktuellen Jahres (2023) von R für uns in Ordnung. Mit Hilfe der `Zodiac()` Funktion aus dem `DescTools` Paket wird der Datumswert in das Sternzeichen (in Form einer Zeichenkette) umgewandelt.

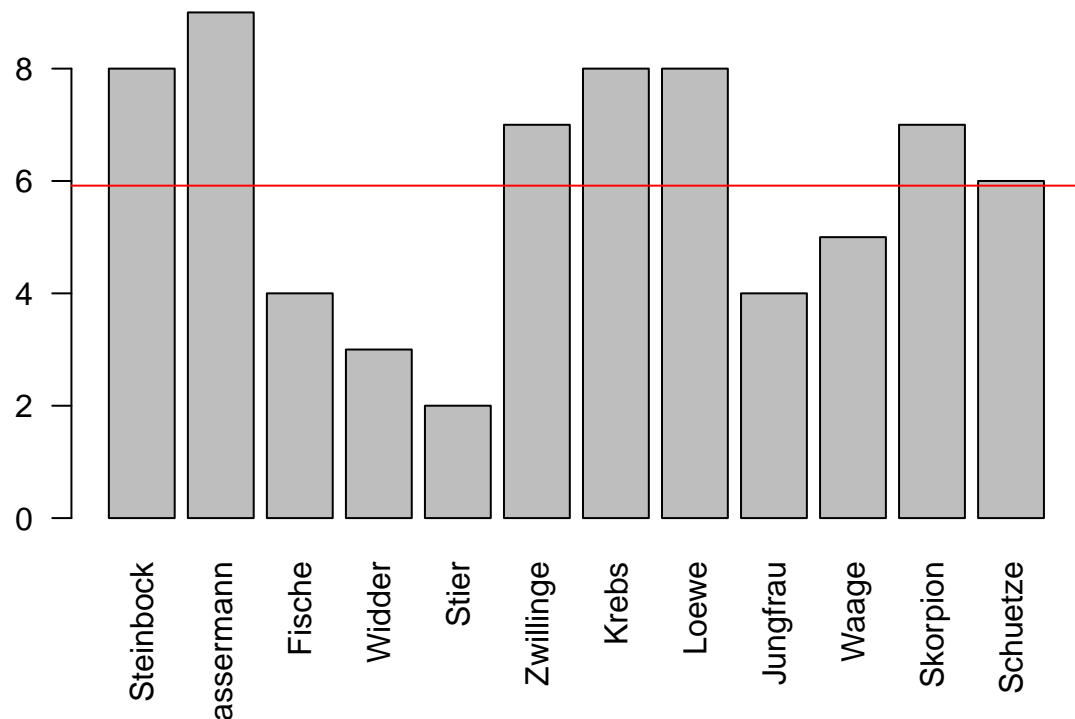
```
# install.packages("DescTools")
library(DescTools)
daten$sternzeichen<-Zodiac(daten$geburtstag, lang="deu")
summary(daten)
```

```
## id geburtstag extraversion_introversion
## Min. : 1.0 Min. :2023-01-04 Extraversion:46
## 1st Qu.:18.5 1st Qu.:2023-04-03 Introversion:25
## Median :36.0 Median :2023-07-15
## Mean :36.0 Mean :2023-07-09
## 3rd Qu.:53.5 3rd Qu.:2023-10-13
## Max. :71.0 Max. :2023-12-26
##
## intuition_sensorik denken_fuehlen wahrnehmen_beurteilen sternzeichen
## Intuition:35 Denken:39 Beurteilen:38 Wassermann: 9
## Sensorik :36 Fühlen:32 Wahrnehmen:33 Steinbock : 8
##
## Krebs : 8
## Loewe : 8
## Zwillinge : 7
## Skorpion : 7
## (Other) :24
```

Deskriptive Analyse der Stichprobe

Die Verteilung der Sternzeichen wird als Barplot visualisiert:

```
sternzeichen<-daten$sternzeichen
barplot(table(sternzeichen), las=2)
abline(h=mean(length(sternzeichen)/12), col="red")
```



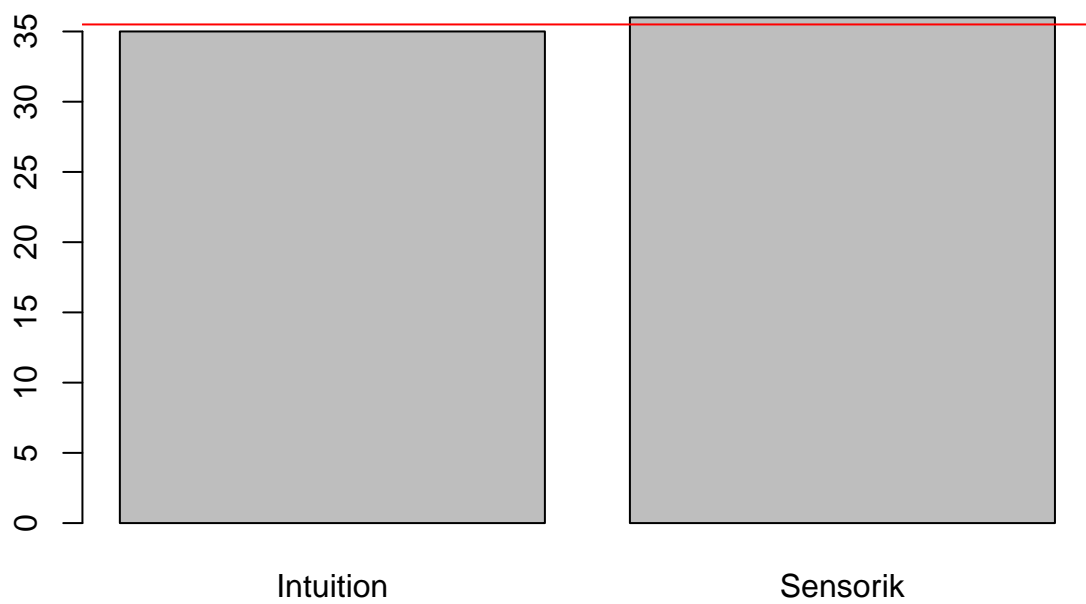
Es ist erkennbar, dass die Sternzeichen in der Stichprobe nicht gleich oft vorkommen. #TODO: So kommt das Sternzeichen ... mit einer Population von ... am häufigsten vor, das Sternzeichen ... mit einer Population von ... am seltensten (auf die Stichprobe bezogen).

Die Verteilung der vier erfassten Charaktereigenschaften werden auch als Barplots visualisiert:

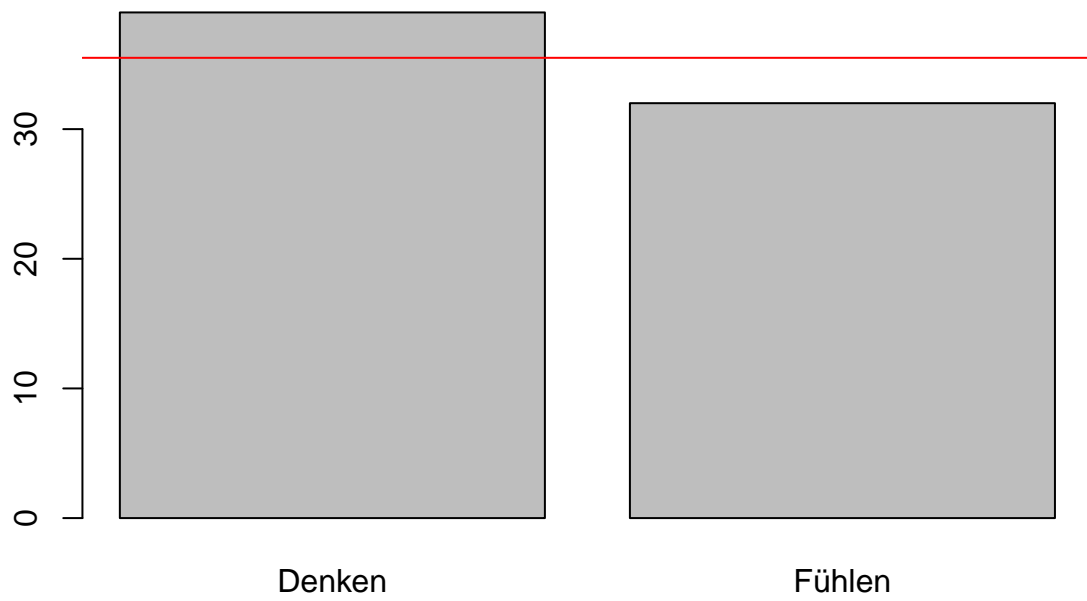
```
c1<-daten$extraversion_introversion
c2<-daten$intuition_sensorik
c3<-daten$denken_fuehlen
c4<-daten$wahrnehmen_beurteilen
barplot(table(c1))
abline(h=mean(length(c1)/2), col="red")
```



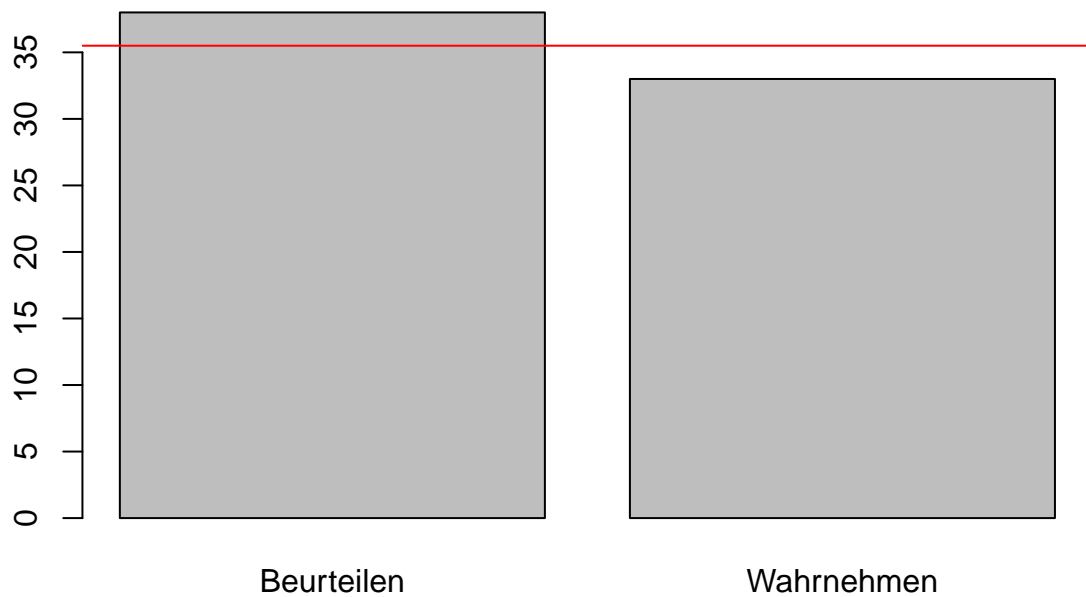
```
barplot(table(c2))  
abline(h=mean(length(c2)/2), col="red")
```

```
barplot(table(c3))  
abline(h=mean(length(c3)/2), col="red")
```



```
barplot(table(c4))  
abline(h=mean(length(c4)/2), col="red")
```



Untersuchung des Zusammenhangs Sternzeichen und Charaktereigenschaften

TODO: Durch 4

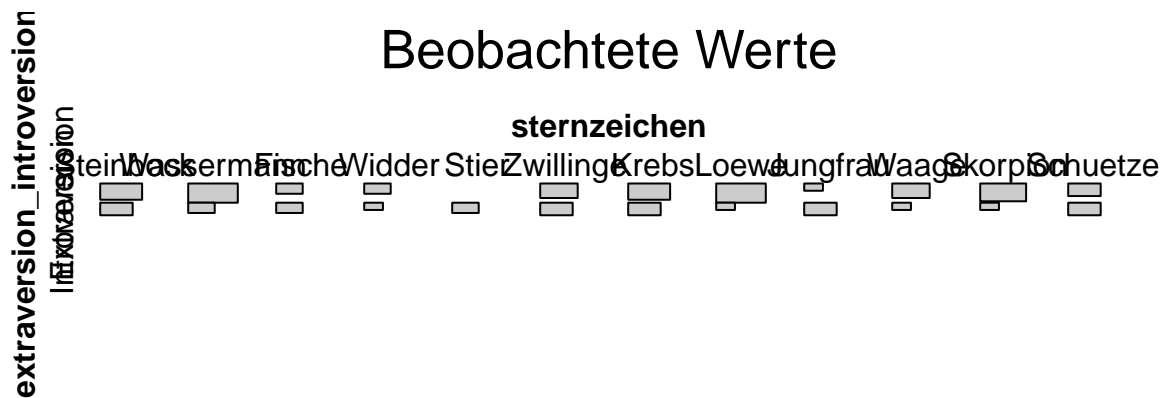
```
tab_extraversion_introversion_sternzeichen = with(daten, table(extraversion_introversion, sternzeichen))
tab_extraversion_introversion_sternzeichen
```

```
##                sternzeichen
## extraversion_introversion Steinbock Wassermann Fische Widder Stier Zwillinge
##           Extraversion      5      7      2      2      0      4
##           Introversion      3      2      2      1      2      3
##                sternzeichen
## extraversion_introversion Krebs Loewe Jungfrau Waage Skorpion Schuetze
##           Extraversion      5      7      1      4      6      3
##           Introversion      3      1      3      1      1      3
```

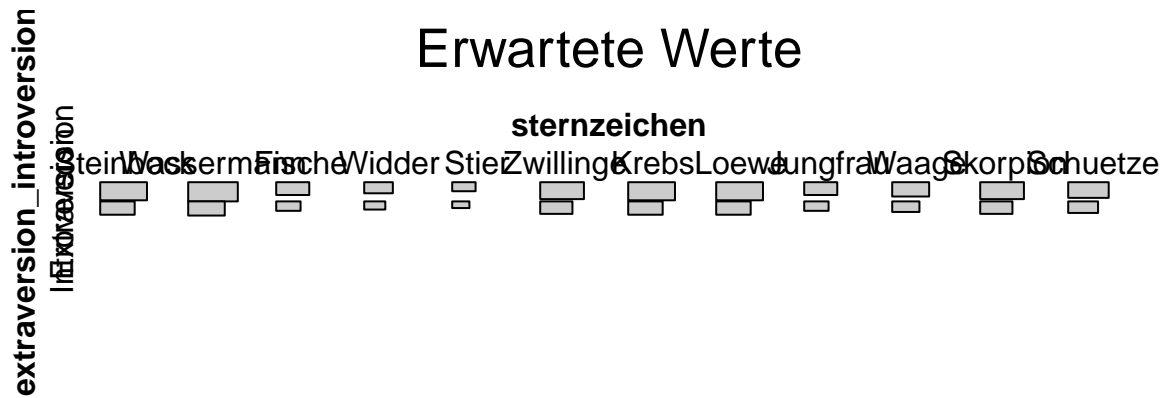
```
library(vcd)
```

```
## Loading required package: grid
```

```
tile(tab_extraversion_introversion_sternzeichen, main="Beobachtete Werte")
```



```
tile(tab_extraversion_introversion_sternzeichen, type="expected", main="Erwartete Werte")
```



```
chisq.test(tab_extraversion_introversion_sternzeichen)
```

```
## Warning in chisq.test(tab_extraversion_introversion_sternzeichen): Chi-squared
## approximation may be incorrect
```

```
##
## Pearson's Chi-squared test
##
## data:  tab_extraversion_introversion_sternzeichen
## X-squared = 11.96, df = 11, p-value = 0.3666
```

Überprüfung der Fragestellung bzgl. der Grundgesamtheit (Tests, Konfidenzintervalle, Modelle etc.)

- Bei Tests:
 - Angabe von Testbezeichnung
 - Begründung der Testauswahl (z.B. robuste Methoden bei Ausreißern etc.)
 - Null- und Alternativhypothese
 - Signifikanzniveau
 - P-Wert
 - Formale Schlussfolgerung (Vergleich P-Wert mit Signifikanzniveau)
 - Inhaltliche Schlussfolgerung

- Bei Modellen:
 - Begründung für die Wahl des Modells (z.B. abhängig vom Skalenniveau der abhängigen und unabhängigen Variablen)
 - Schätzung des Modells
 - Bei Regressionsmodellen: Ableitung der Modellformel, sowie Interpretation des Einflusses der unabhängigen Variablen auf die abhängige Variable (“Effektstärke”). Unterstützung via Effektplot bzw. Modellplot
 - Beurteilung der Signifikanz der Parameter, sowie des Gesamtmodells
 - Beurteilung der Modellgüte (z.B. R^2 bei linearer Regression)
 - Modelldiagnostik (Normalverteilung der Residuen, Restzusammenhang der Residuen mit den geschätzten Werten)
 - Diagramm mit erwarteten Werten (z.B. Regressionsgerade) samt Konfidenzbändern
 - Vorhersagen (erwartete Werte, bzw. Prognoseintervalle für Einzelwerte)

Fazit

- Zusammenfassung
- Diskussion allfälliger Probleme
- Generalisierbarkeit der Ergebnisse