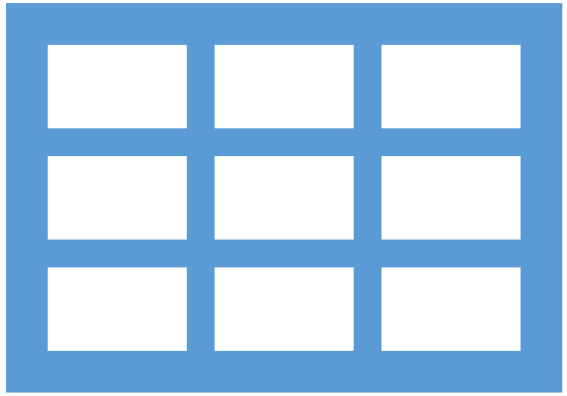
Statistik 1 – Tutorate
Einheit: Tabellenanalyse

Marco Giesselmann, Aurelia De Martinis, Alex Geistlich, Dominic Truxius, Nora Zumbühl

Kreuztabelle mit R



1 Kreuztabellen

- Welchen Zusammenhang vermutet ihr zwischen den Merkmalen ***Geschlecht*** und ***Rauchverhalten****?* Greift auch die Zusammenhangsform (Asymmetrie?, Tendenz?) mit in der Vermutung auf.
- Startet ein neues R-Skript und ergänzt es mit einem Header, der die wichtigsten Metadaten (Titel, Autor, Datum, Zweck) enthält. Kommentiert euren Code durchgehend.
- Ladet die Kursdaten in R, aktiviert die tidyverse-Packages (*library (tidyverse)*)
- Inspiziert die zu den Merkmalen korrespondierenden Variablen ***geschlecht*** und ***rauchen_aktuell*** (*attributes*, *table*, *class*)
- Führt ggf. Variablenbereinigungen durch!!
- Sinnvoll, da es sich um kategoriale Variablen handelt: Faktorisierungen per ***as_factor***.

geschlecht	**rauchen_aktuell** letzte Woche geraucht?
maennlich	0
weiblich	0
maennlich	1
weiblich	0
weiblich	1
weiblich	0
maennlich	1
weiblich	1
maennlich	0
maennlich	0
maennlich	0
maennlich	0

1 Kreuztabellen

```
[1] "letzte Woche geraucht?"

$format.stata
[1] "%12.0g"

$class
[1] "haven_labelled" "vctrs_vctr"     "double"

$labels
Keine Angabe          nein            ja
         -99             0             1

> table(kursdata_anon$rauchen_aktuell)

-99   0   1
 1  46  14
```

Inspektion z.B. per attributes: «rauchen_aktuell» enthält einen nicht korrekt als NA codierten fehlenden Wert!

daher…

```
kursdata_anon$rauchen_aktuell[kursdata_anon$rauchen_aktuell==-99]<-NA
table(kursdata_anon$rauchen_aktuell)
```

Faktorisierung:

```
kursdata_anon$geschlecht <- as_factor(kursdata_anon$geschlecht)
kursdata_anon$rauchen_aktuell <- as_factor(kursdata_anon$rauchen_aktuell)
```

Häufiges Problem nach *as_factor*: «Phantomkategorie»*

```
> table(kursdata_anon$rauchen_aktuell)

Keine Angabe          nein            ja
           0            46            14
```

Löschung der Phantomkategorie durch:

```
kursdata_anon$rauchen_aktuell<-fct_drop(kursdata_anon$rauchen_aktuell)
```

geschlecht	**rauchen_aktuell** letzte Woche geraucht?
maennlich	0
weiblich	0
maennlich	1
weiblich	0
weiblich	1
weiblich	0
maennlich	1
weiblich	1
maennlich	0
maennlich	0
maennlich	0
maennlich	0

*Durch die Faktorisierung über «as_factor» wird das label einer NA-Kategorie als (leere) Kategorie neu angelegt und stört nun die Darstellung von Tabellen und Abbildungen.

1.1 Kreuztabellen

Über den **tab_xtab()** Befehl aus dem „**sjPlot**“ Package lassen sich anschauliche Kreuztabellen erstellen.

```
tab_xtab(var.row = kursdata_anon$rauchen_aktuell,
         var.col = kursdata_anon$geschlecht,
         show.col.prc = TRUE,
         show.obs = TRUE)
```

letzte Woche geraucht?	*geschlecht* weiblich	maennlich	***Total***
nein	36 81.8 %	10 62.5 %	46 76.7 %
ja	8 18.2 %	6 37.5 %	14 23.3 %
Total	44 100 %	16 100 %	60 100 %

χ^2=1.487 · df=1 · &phi=0.202 · Fisher's p=0.168

- Beschreibt den Tabellenaufbau
- Beschreibt die einzelnen Elemente des Befehls
- Wie viele Befragungspersonen rauchen aktuell?
- Wie gross ist deren Anteil?
- Was sagt der Prozentwert im Feld i=21 („ja“ & „weiblich“) aus?
- Unterscheidet sich der Anteil aktuell Rauchender zwischen den Geschlechtern? Ermittle und Interpretiere die *Prozentsatzdifferenz*.
- Produziere eine Tabelle mit Zeilen- statt Spaltenprozenten

1.1 Kreuztabellen

Über den **tab_xtab()** Befehl aus dem „**sjPlot**“ Package lassen sich anschauliche Kreuztabellen erstellen.

```
tab_xtab(var.row = kursdata_anon$rauchen_aktuell,
         var.col = kursdata_anon$geschlecht,
         show.col.prc = TRUE,
         show.obs = TRUE)
```

letzte Woche geraucht?	*geschlecht* weiblich	maennlich	***Total***
nein	36 81.8 %	10 62.5 %	46 76.7 %
ja	8 18.2 %	6 37.5 %	14 23.3 %
Total	44 100 %	16 100 %	60 100 %

χ^2=1.487 · df=1 · &phi=0.202 · Fisher's p=0.168

Unter den weiblichen Personen rauchen aktuell etwa 18.2%, also ein knappes Fünftel

- Was sagt der Prozentwert im Feld unten links („ja“ & „weiblich“) aus?
- Unterscheidet sich der Anteil aktuell Rauchender zwischen den Geschlechtern? Ermittle und Interpretiere die *Prozentsatzdifferenz*.

%d=19.3: Der Anteil aktuell Nicht-Rauchender ist unter den Frauen 19.3 Prozentpunkte grösser als unter den Männern. Oder:
%d=-19.3: Der Anteil aktuell Rauchender ist unter Frauen 19.3 Prozentpunkte kleiner als unter den Männern.

1.1 Kreuztabellen

Über den **tab_xtab()** Befehl aus dem „**sjPlot**“ Package lassen sich anschauliche Kreuztabellen erstellen.

```
tab_xtab(var.row = kursdata_anon$rauchen_aktuell,
         var.col = kursdata_anon$geschlecht,
         show.row.prc = TRUE,
         show.obs = TRUE)
```

Kreuztabelle: Rauchstatus nach Geschlecht

letzte Woche geraucht?	*geschlecht* weiblich	maennlich	***Total***
nein	36 78.3 %	10 21.7 %	46 100 %
ja	8 57.1 %	6 42.9 %	14 100 %
Total	44 73.3 %	16 26.7 %	60 100 %

χ^2=1.487 · df=1 · &phi=0.202 · Fisher's p=0.168

- Beschreibt den Tabellenaufbau
- Beschreibt die einzelnen Elemente des Befehls
- Wie viele Befragungspersonen rauchen aktuell?
- Wie gross ist deren Anteil?
- Was sagt der Prozentwert im Feld unten links („ja“ & „männlich“) aus?
- Unterscheidet sich der Anteil aktuell Rauchender zwischen den Geschlechtern? Ermittle und Interpretiere die *Prozentsatzdifferenz*.
- **Produziert eine Tabelle mit Zeilen- statt Spaltenprozenten mit dem Befehl**

1.1 Kreuztabellen

Über den **tab_xtab()** Befehl aus dem „**sjPlot**“ Package lassen sich anschauliche Kreuztabellen erstellen.

```
tab_xtab(var.row = kursdata_anon$rauchen_aktuell,
         var.col = kursdata_anon$geschlecht,
         show.row.prc = TRUE,
         show.obs = TRUE)
```

- Was sagt der Prozentwert im Feld unten links („ja“ & „weiblich“) ***nun*** aus?
- Was sagt der Prozentwert im Feld oben rechts („männlich“ & „nein“) aus?
- Lässt die so formatierte Kreuztabelle einen Rückschluss auf den Zusammenhang zwischen den beiden Variablen zu?
- Warum ist dieser Differenzwert trotzdem nicht die *richtige* Prozentsatzdifferenz des Zusammenhangs?

Kreuztabelle: Rauchstatus nach Geschlecht

letzte Woche geraucht?	*geschlecht* weiblich	maennlich	***Total***
nein	36 78.3 %	10 21.7 %	46 100 %
ja	8 57.1 %	6 42.9 %	14 100 %
Total	44 73.3 %	16 26.7 %	60 100 %

χ^2=1.487 · df=1 · &phi=0.202 · Fisher's p=0.168

1.1 Kreuztabellen

Ist die Tabelle in dieser Form vollständig und publikationswürdig?

letzte Woche geraucht?	*geschlecht* weiblich	maennlich	***Total***
nein	36 81.8 %	10 62.5 %	46 76.7 %
ja	8 18.2 %	6 37.5 %	14 23.3 %
Total	44 100 %	16 100 %	60 100 %

χ^2=1.487 · df=1 · &phi=0.202 · Fisher's p=0.168

Weitere Bearbeitungsschritte zur Publikation:

- Titel, Untertitel, Datenquelle
- Generelle Formatierungsarbeiten, Schriftgrösse?
- Kann z.T. über Suboptionen innerhalb des Befehls spezifiziert werden, grundsätzlich aber extern (z.B. Word oder Powerpoint)

Externe Weiterverarbeitung / Export:

- Die Tabelle wird automatisch im „Viewer“-Tab der R-Studio Konsole (rechts unten) angezeigt.
- Einfach per select/copy/paste in andere Dokumente bzw. Formate einfügen

Grafische Darstellung kreuztabellarischer Zusammenhänge

1.2 Visualisierung von Kreuztabellen

Achtung: Anders als Tabellenkommandos integrieren ggplot-Befehle Fehlende Werte (NAs) in die Darstellung. Das ist meistens schlecht – siehe HP

Daher vorab:

```
kursdata_rauchplot <- filter(kursdata_anon, !is.na(geschlecht) & !is.na(rauchen_aktuell))
```

Analysespezifischer Datensatz

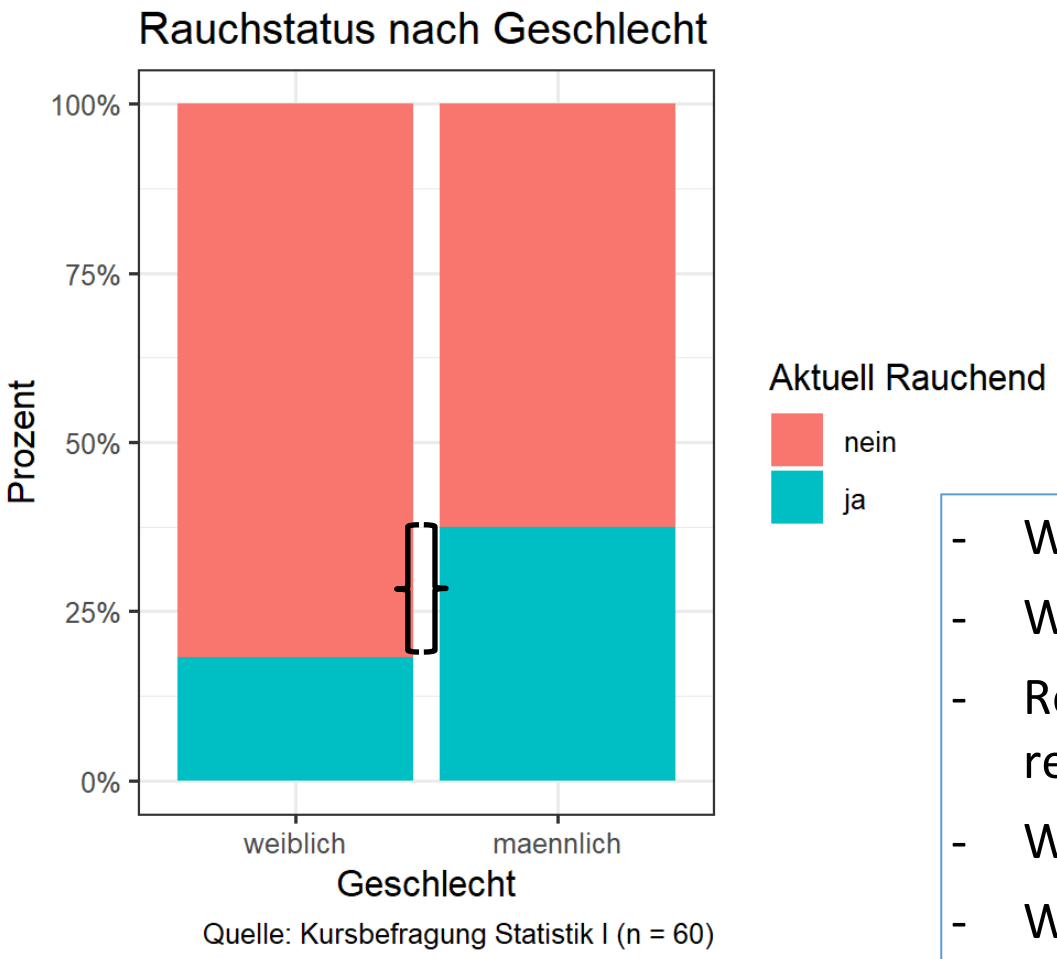
Achtung:

Funktioniert (natürlich) nur dann, wenn fehlende Werte korrekt als «NA» definiert wurden. Ggf. nochmal checken!

1.2 Stacked Barplot: Visualisierung gemeinsamer Verteilung

```
ggplot(kursdata_rauchplot, aes(x = geschlecht, fill = rauchen_aktuell)) +
  geom_bar(position = "fill") +
  labs(title = "Rauchstatus nach Geschlecht",
       x = "Geschlecht", y = "Prozent", fill="Aktuell Rauchend",
       caption="Quelle: Kursbefragung Statistik I (n = 53)") +
  scale_y_continuous(labels = scales::percent) +
  theme_bw()
```

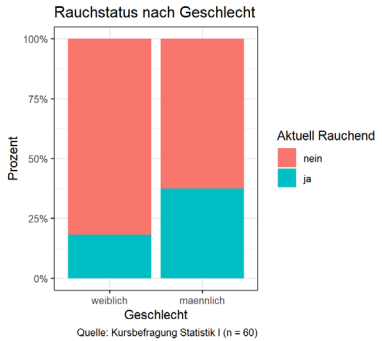
1.2 Stacked Barplot: Visualisierung gemeinsamer Verteilung




```r
ggplot(kursdata_rauchplot, aes(x = geschlecht, fill = rauchen_aktuell)) +
  geom_bar(position = "fill") +
  labs(title = "Rauchstatus nach Geschlecht",
       x = "Geschlecht", y = "Prozent", fill="Aktuell Rauchend",
       caption="Quelle: Kursbefragung Statistik I (n = 53)") +
  scale_y_continuous(labels = scales::percent) +
  theme_bw()
```

- Wofür stehen hier jeweils die beiden Säulen?
- Wo werden die Säulenkategorien im Code definiert?
- Repräsentieren die linken 100% gleich viele Personen wie die rechten 100%?
- Was kennzeichnet jeweils die rote Fläche?
- Wo werden die säuleninternen Farbkategorien im Code definiert?
- Wo wird in dieser Abbildung die Prozentsatzdifferenz visualisiert?

1.2 Stacked Barplot: Visualisierung gemeinsamer Verteilung



Bilde ein Säulendiagramm ab

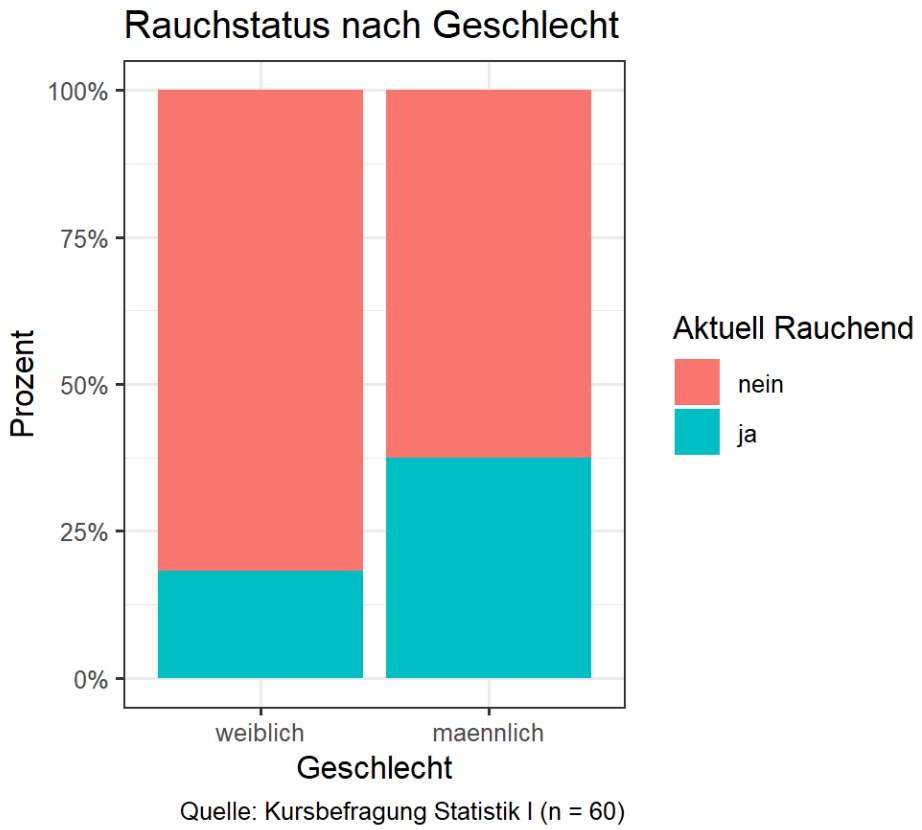
Fülle die «geoms» (hier: Säulen) *nicht* mit einer bestimmten Farbe (z.B. fill=«red»), sondern jeweils entsprechend der Verteilung der «rauchen»-Variable

```
ggplot(kursdata_rauchplot, aes(x = geschlecht, fill = rauchen_aktuell)) +
  geom_bar(position = "fill") +
  labs(title = "Rauchstatus nach Geschlecht",
       x = "Geschlecht", y = "Prozent", fill="Aktuell Rauchend",
       caption="Quelle: Kursbefragung Statistik I (n = 53)") +
  scale_y_continuous(labels = scales::percent) +
  theme_bw()
```

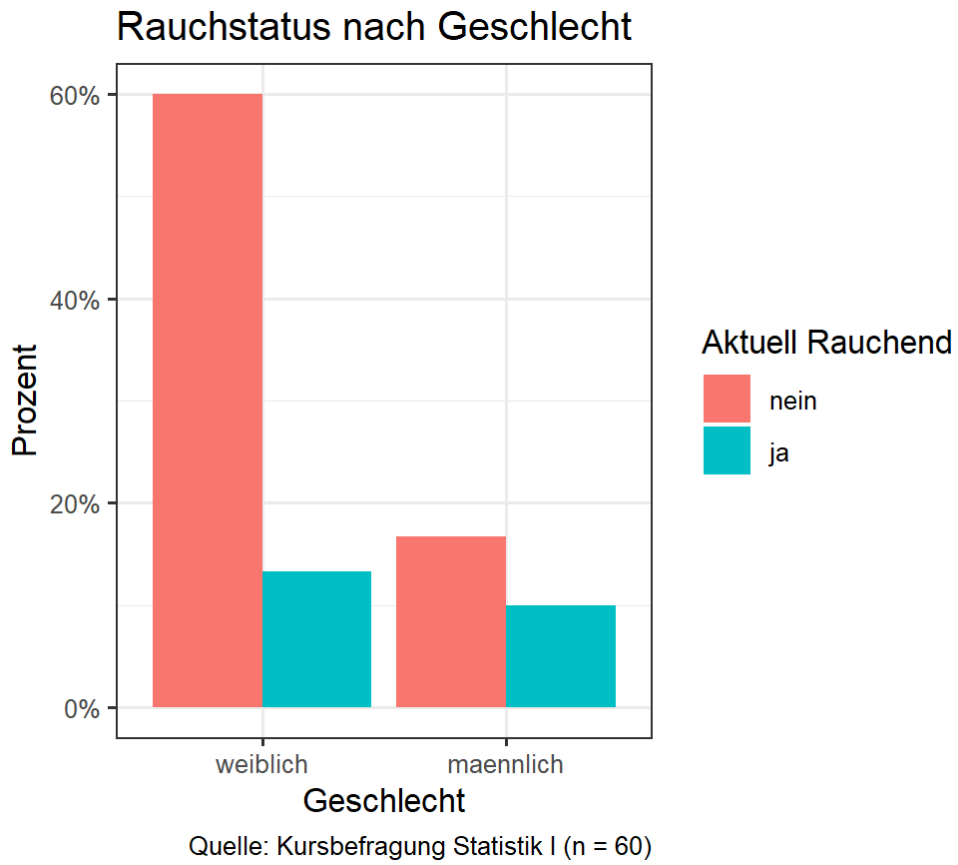
- Alle Säulen sollen die gleiche Höhe haben
- die Kategotieren der Füllvariable sollen dabei als Anteilswerte dargestellt werden

Multipliziere die y-Werte mit 100 und stelle sie mit Prozentzeichen dar

1.2 Alternative „Dodge“-Plot – Unterschiede in der Darstellung?




```
ggplot(kursdata_rauchplot, aes(x = geschlecht, fill = rauchen_aktuell)) +
  geom_bar(position = "fill") +
  labs(title = "Rauchstatus nach Geschlecht",
       x = "Geschlecht", y = "Prozent", fill="Aktuell Rauchend",
       caption="Quelle: Kursbefragung Statistik I (n = 53)") +
  scale_y_continuous(labels = scales::percent) +
  theme_bw()
```




```
ggplot(kursdata_rauchplot, aes(x = geschlecht, fill = rauchen_aktuell)) +
  geom_bar(position="dodge") +
  aes(y = after_stat(count / sum(count)))+
  labs(title = "Rauchstatus nach Geschlecht",
       x = "Geschlecht", y="Prozent", fill="Aktuell Rauchend",
       caption="Quelle: Kursbefragung Statistik I (n = 53)") +
  scale_y_continuous(labels = scales::percent) +
  theme_bw()
```

2. Kreuztabelle: Weiteres Beispiel aus der Kursbefragung

Wir wollen prüfen, in welchem Zusammenhang die *Elterliche Bildung* und das *allgemeine Vertrauen* innerhalb des Kurses stehen. Dazu verwenden wir die Variablen **akback** und zusätzlich **trustkat**.

I. Inspiziert die generierte (rekodierte) Variable **trustkat**. In welchem Verhältnis steht diese zur (Originalvariable) **trust**?
II. Formuliert und begründet eine **Hypothese** zu den beiden Variablen
III. Erstellt eine **Kreuztabelle** welche die gemeinsame Verteilung der beiden Variablen sinnvoll (im Sinne der formulierten Hypothese) abbildet.
IV. Wertet die Tabelle in einem inhaltlich gehaltvollen Antwortsatz aus (**Prozentsatzdifferenz**!).
V. Visualisiert den Zusammenhang
VI. Berechnet den Wert von Cramers V per Hand aus dem Chi2-Wert im Output.
VII. Stützt Eure Auswertung der Prozentsatzdifferenz durch Cramers V
VIII. Stützt Eure Auswertung durch Einbindung der Test-Statistik des Chi-Quadrat Tests

2. Kreuztabelle: Weiteres Beispiel aus der Kursbefragung

Wir wollen prüfen, in welchem Zusammenhang die *Elterliche Bildung* und das *allgemeine Vertrauen* innerhalb des Kurses stehen. Dazu verwenden wir die Variablen **akback** und zusätzlich **trustkat**.

I. Inspiziert die generierte (rekodierte) Variable **trustkat**. In welchem Verhältnis steht diese zur (Originalvariable) **trust**?
II. Formuliert und begründet eine **Hypothese** zu den beiden Variablen
III. Erstellt eine **Kreuztabelle** welche die gemeinsame Verteilung der beiden Variablen sinnvoll (im Sinne der formulierten Hypothese) abbildet.
IV. Wertet die Tabelle in einem inhaltlich gehaltvollen Antwortsatz aus (**Prozentsatzdifferenz**!).
V. Ermittelt Lambda
VI. Visualisiert den Zusammenhang
VII. Ermittelt die Indifferenztabelle per Hand (vollständig, definiert für alle 6 Felder)
VIII. Ermittelt die Indifferenztabelle mit tab_xtab
IX. Ermittelt den Chi-Quadrat-Wert der Tabelle per Hand
X. Berechnet den Wert von Cramers V per Hand aus dem Chi2-Wert im Output.
XI. Stützt Eure Auswertung der Prozentsatzdifferenz durch Cramers V
XII. Stützt Eure Auswertung durch Einbindung der Test-Statistik des Chi-Quadrat Tests
XIII. Vertauscht Spalten und Zeilen – welche Werte ändern sich? Welche bleiben gleich?

2. Kreuztabelle: Weiteres Beispiel aus der Kursbefragung

Wir wollen prüfen, in welchem Zusammenhang die *Elterliche Bildung* und das *allgemeine Vertrauen* innerhalb des Kurses stehen. Dazu verwenden wir die Variablen **akback** und zusätzlich **trustkat**.

I. Inspiziert die generierte (rekodierte) Variable **trustkat**. In welchem Verhältnis steht diese zur (Originalvariable) **trust**?

II. Formuliert und begründet eine **Hypothese** zu den beiden Variablen

III. Erstellt eine **Kreuztabelle** welche die gemeinsame Verteilung der beiden Variablen sinnvoll (im Sinne der formulierten Hypothese) abbildet.

IV. Wertet die Tabelle in einem inhaltlich gehaltvollen Antwortsatz aus (**Prozentsatzdifferenz**!).

V. Visualisiert den Zusammenhang

VI. Ermittelt die Indifferenztabelle per Hand (vollständig, definiert für alle 6 Felder)

VII. Ermittelt die Indifferenztabelle mit tab_xtab

VIII. Ermittelt den Chi-Quadrat-Wert der Tabelle per Hand

IX. Berechnet den Wert von Cramers V per Hand aus dem Chi2-Wert im Output.

X. Stützt Eure Auswertung der Prozentsatzdifferenz durch Cramers V

XI. Stützt Eure Auswertung durch Einbindung der Test-Statistik des Chi-Quadrat Tests

XII. Vertauscht Spalten und Zeilen – welche Werte ändern sich? Welche bleiben gleich?

2 *trust* und *trustkat* – *wie rekodiert?*

baseR::table – nicht geeignet für inhaltlich motivierte Kreuztabellen, aber schnell und praktisch zum Check von Variablenrekodierungen

id	trust Kann man Menschen im Allg. vertrauen? (5-volle Zustimmung, 1-v...	trustkat Allg. Vertrauen (kat.)
77	4	Viel
76	2	Gering
78	5	Viel
49	3	Mittel
79	3	Mittel
81	2	Gering
82	3	Mittel
80	1	Gering
86	3	Mittel

```
> table(as_factor(kursdata_anon$trustkat),
+       as_factor(kursdata_anon$trust))

          1  2  3  4  5
  Gering  1 10  0  0  0
  Mittel  0  0 24  0  0
  Viel    0  0  0 23 11
```

Die ursprünglichen (relativ dünn belegten) Randausprägungen 1 und 5 in **trust** wurden in **trustkat** mit den anliegenden Ausprägungen 2 und 4 gruppiert, wodurch dann eine trichotome Variable mit den Ausprägungen 1 ("Gering") 2 ("Mittel") und 3 ("Viel") entsteht.

2 *eltern* und *akback* – *wie rekodiert?*

```
> table(as_factor(kursdata_anon$akback),
+       as_factor(kursdata_anon$eltern))

       beide einer keiner
  nein     0     0     21
  ja      25    19      0
```

Die dichotomisierte Variable *akback* unterscheidet zwischen Studierenden, bei denen mindestens ein Elternteil einen akademischen Abschluss hat und denen, bei denen kein Elternteil einen akademischen Abschluss hat.

2 Kreuztabelle und Abbildung – Code und Interpretation

```
kursdata_anon$trustkat<-as_factor(kursdata_anon$trustkat)

tab_xtab(var.row = kursdata_anon$trustkat, var.col = kursdata_anon$akback,
         title = "Kreuztabelle: Vertrauen nach Bildungshintergrund",
         var.labels = c("Vertrauen in Mitmenschen", "Akademikerkind?"),
         show.col.prc = TRUE,
         show.obs = TRUE)
```

Kreuztabelle: Vertrauen nach Bildungshintergrund

Vertrauen in Mitmenschen	*Akademikerkind?* nein	ja	***Total***
Gering	4 19 %	7 15.9 %	11 16.9 %
Mittel	8 38.1 %	14 31.8 %	22 33.8 %
Viel	9 42.9 %	23 52.3 %	32 49.2 %
Total	21 100 %	44 100 %	65 100 %

χ^2=0.504 · df=2 · Cramer's V=0.088 · Fisher's p=0.760

Auswertung nach Prozentsatzdifferenz?

2 Kreuztabelle und Abbildung – Code und Interpretation

```
kursdata_anon$trustkat<-as_factor(kursdata_anon$trustkat)

tab_xtab(var.row = kursdata_anon$trustkat, var.col = kursdata_anon$akback,
         title = "Kreuztabelle: Vertrauen nach Bildungshintergrund",
         var.labels = c("Vertrauen in Mitmenschen", "Akademikerkind?"),
         show.col.prc = TRUE,
         show.obs = TRUE)
```

Kreuztabelle: Vertrauen nach Bildungshintergrund

Vertrauen in Mitmenschen	*Akademikerkind?* nein	ja	***Total***
Gering	4 19 %	7 15.9 %	11 16.9 %
Mittel	8 38.1 %	14 31.8 %	22 33.8 %
Viel	9 42.9 %	23 52.3 %	32 49.2 %
Total	21 100 %	44 100 %	65 100 %

χ^2=0.504 · df=2 · Cramer's V=0.088 · Fisher's p=0.760

Musterauswertung, Beispiel (mehrere Varianten und Lösungen möglich):

….die bedingten Verteilungen der Vertrauensvariable unterscheiden sich grundsätzlich nicht stark zwischen Akademiker- und Arbeiterkindern – eine Wirkung der unabhängigen auf die abhängige Variable lässt sich nur ansatzweise erkennen. Gleichwohl beträgt die Prozentsatzdifferenz in der höchsten Kategorie immerhin fast 10 Prozentpunkte: Unter den Akademikerkindern ist der Anteil derjenigen, die hohes Vertrauen in ihre Mitmenschen haben, 9.4 Prozentpunkte grösser als unter Arbeiterkindern. Dies markiert einen inhaltlich bedeutsamen Unterschied.

2 Kreuztabelle und Abbildung – Code und Interpretation

```
kursdata_anon$trustkat<-as_factor(kursdata_anon$trustkat)

tab_xtab(var.row = kursdata_anon$trustkat, var.col = kursdata_anon$akback,
         title = "Kreuztabelle: Vertrauen nach Bildungshintergrund",
         var.labels = c("Vertrauen in Mitmenschen", "Akademikerkind?"),
         show.col.prc = TRUE,
         show.obs = TRUE)
```

Kreuztabelle: Vertrauen nach Bildungshintergrund

Vertrauen in Mitmenschen	*Akademikerkind?* nein	ja	***Total***
Gering	4 19 %	7 15.9 %	11 16.9 %
Mittel	8 38.1 %	14 31.8 %	22 33.8 %
Viel	9 42.9 %	23 52.3 %	32 49.2 %
Total	21 100 %	44 100 %	65 100 %

χ^2=0.504 · df=2 · Cramer's V=0.088 · Fisher's p=0.760

Lambda?

2 Kreuztabelle und Abbildung – Code und Interpretation

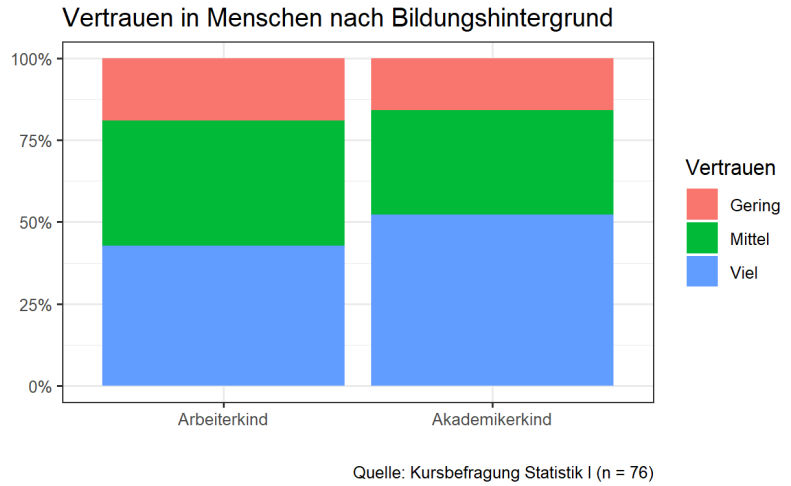
```
kursdata_anon$trustkat<-as_factor(kursdata_anon$trustkat)

tab_xtab(var.row = kursdata_anon$trustkat, var.col = kursdata_anon$akback,
         title = "Kreuztabelle: Vertrauen nach Bildungshintergrund",
         var.labels = c("Vertrauen in Mitmenschen", "Akademikerkind?"),
         show.col.prc = TRUE,
         show.obs = TRUE)
```

Kreuztabelle: Vertrauen nach Bildungshintergrund

Vertrauen in Mitmenschen	*Akademikerkind?* nein	ja	***Total***
Gering	4 19 %	7 15.9 %	11 16.9 %
Mittel	8 38.1 %	14 31.8 %	22 33.8 %
Viel	9 42.9 %	23 52.3 %	32 49.2 %
Total	21 100 %	44 100 %	65 100 %

χ^2=0.504 · df=2 · Cramer's V=0.088 · Fisher's p=0.760

```
kursdata_plot2 <- filter(kursdata_anon, !is.na(akback) & !is.na(trustkat))

plot_fill_trust <- ggplot(kursdata_plot2, aes(x = akback, fill = trustkat)) +
  geom_bar(position = "fill") +
  labs(title = "Vertrauen in Menschen nach Bildungshintergrund",
       x = "", y = "",  fill="Vertrauen",
       caption="Quelle: Kursbefragung Statistik I (n = 76)") +

  scale_y_continuous(labels = scales::percent_format()) +
  scale_x_discrete(labels=c("Arbeiterkind", "Akademikerkind")) +
  theme_bw()
plot_fill_trust
```



2 Kreuztabelle und Abbildung – Code und Interpretation

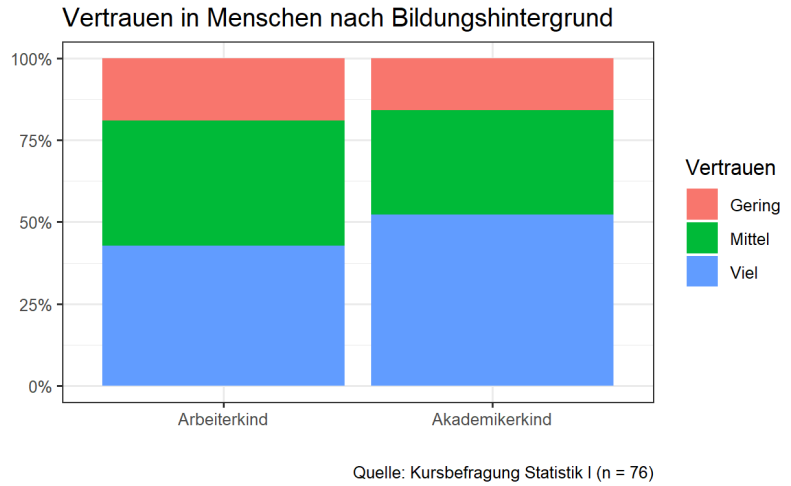
```
kursdata_anon$trustkat<-as_factor(kursdata_anon$trustkat)

tab_xtab(var.row = kursdata_anon$trustkat, var.col = kursdata_anon$akback,
         title = "Kreuztabelle: Vertrauen nach Bildungshintergrund",
         var.labels = c("Vertrauen in Mitmenschen", "Akademikerkind?"),
         show.col.prc = TRUE,
         show.obs = TRUE)
```

Kreuztabelle: Vertrauen nach Bildungshintergrund

Vertrauen in Mitmenschen	*Akademikerkind?* nein	ja	***Total***
Gering	4 19 %	7 15.9 %	11 16.9 %
Mittel	8 38.1 %	14 31.8 %	22 33.8 %
Viel	9 42.9 %	23 52.3 %	32 49.2 %
Total	21 100 %	44 100 %	65 100 %

χ^2=0.504 · df=2 · Cramer's V=0.088 · Fisher's p=0.760

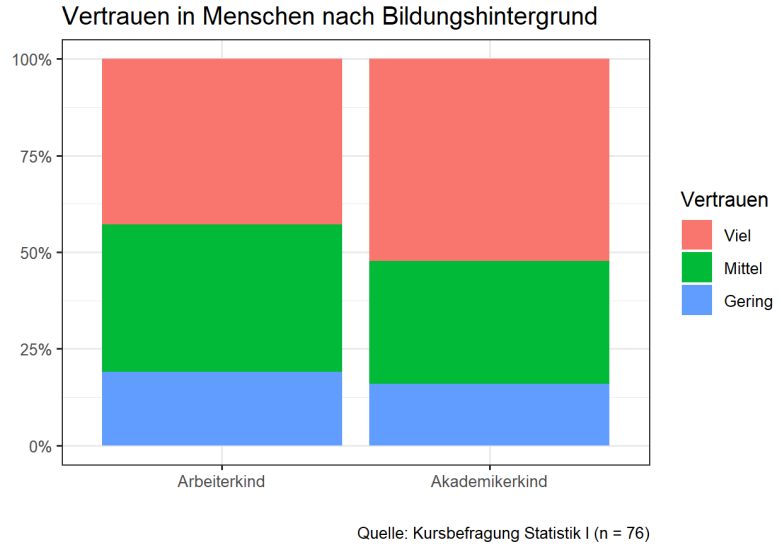
```
kursdata_plot2 <- filter(kursdata_anon, !is.na(akback) & !is.na(trustkat))

plot_fill_trust <- ggplot(kursdata_plot2, aes(x = akback, fill = trustkat)) +
  geom_bar(position = "fill") +
  labs(title = "Vertrauen in Menschen nach Bildungshintergrund",
       x = "", y = "",  fill="Vertrauen",
       caption="Quelle: Kursbefragung Statistik I (n = 76)") +

  scale_y_continuous(labels = scales::percent_format()) +
  scale_x_discrete(labels=c("Arbeiterkind", "Akademikerkind")) +
  theme_bw()
plot_fill_trust
```



Zusatzaufgaben (ChatGPT?):

- Ordnet die Kategorien in den Säulen neu und intuitiver: Niedriges Vertrauen unten, hohes Vertrauen oben
- Was tun, wenn Grautöne gefordert sind?

2 Kreuztabelle und Abbildung – Code und Interpretation



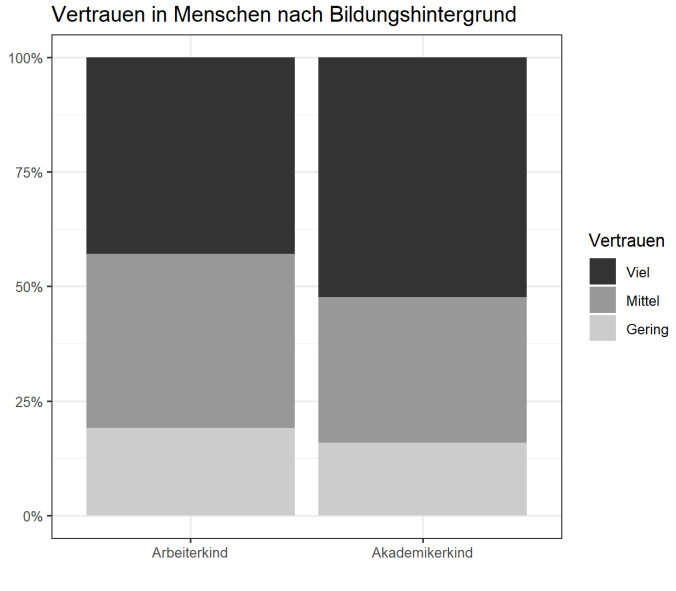

```
kursdata_plot2$trustkat<-factor(kursdata_plot2$trustkat, levels=c("Viel",
"Mittel", "Gering"))

plot_fill_trust <- ggplot(kursdata_plot2, aes(x = akback, fill = trustkat)) +
  geom_bar(position = "fill") +
  labs(title = "Vertrauen in Menschen nach Bildungshintergrund",
       x = "", y = "",  fill="Vertrauen",
       caption="Quelle: Kursbefragung Statistik I (n = 76)") +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_x_discrete(labels=c("Arbeiterkind", "Akademikerkind")) +
  theme_bw()
```

Neuordnung der Kategorien im Rahmen der Faktorisierung

Zusatzaufgaben (ChatGPT?):

- **Ordnet die Kategorien in den Säulen neu und intuitiver: Niedriges Vertrauen unten, hohes Vertrauen oben**
- Was tun, wenn Grautöne gefordert sind?

2 Kreuztabelle und Abbildung – Code und Interpretation



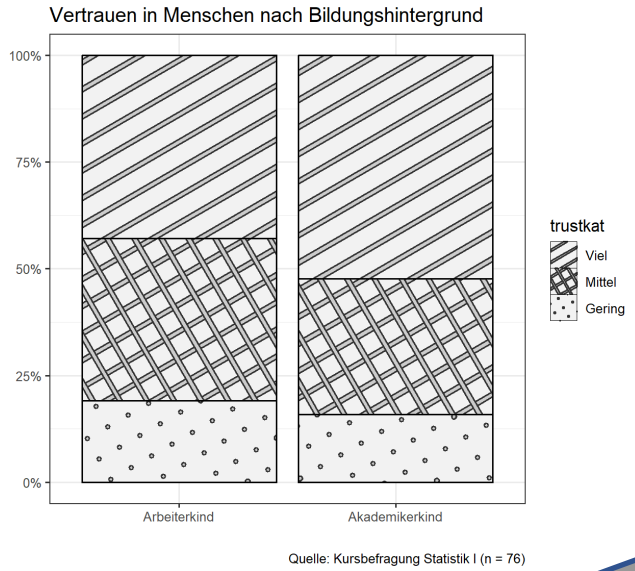

```
plot_fill_grey <- ggplot(kursdata_plot2, aes(x = akback, fill          ekat)) +
  geom_bar(position = "fill") +
  labs(title = "Vertrauen in Menschen nach Bildungshinte      nd",
       x = "", y = "",  fill="Vertrauen",
       caption="Quelle: Kursbefragung Statistik I (n = 76)") +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_x_discrete(labels=c("Arbeiterkind", "Akademikerkind")) + theme_bw() +
  scale_fill_grey(start = 0.2, end = 0.8)
```

Grautöne statt Standardfarbschmema in den Balken

Zusatzaufgaben (ChatGPT?):

- Ordnet die Kategorien in den Säulen neu und intuitiver: Niedriges Vertrauen unten, hohes Vertrauen oben
- **Was tun, wenn Grautöne gefordert sind?**

2 Kreuztabelle und Abbildung – Code und Interpretation



Fancy: ggpattern().
(Funktioniert leider nicht immer.)

```
library(ggpattern)
plot_fill_pattern <- ggplot(kursdata_plot2, aes(x = akk[illegible], fill = trustkat)) +
geom_bar_pattern(aes(pattern=trustkat),
                 color="black",
                 fill="grey95",
                 position = "fill")+
  labs(title = "Vertrauen in Menschen nach Bildungshintergrund",
       x = "", y = "",  fill="Vertrauen",
       caption="Quelle: Kursbefragung Statistik I (n = 76)") +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_x_discrete(labels=c("Arbeiterkind", "Akademikerkind")) + theme_bw()
plot_fill_pattern
```

Zusatzaufgaben (ChatGPT?):

- Ordnet die Kategorien in den Säulen neu und intuitiver: Niedriges Vertrauen unten, hohes Vertrauen oben
- **Was tun, wenn Grautöne gefordert sind?**

2 Indifferenztabelle und Chi-Quadrat

Kreuztabelle: Vertrauen nach Bildungshintergrund

Vertrauen in Mitmenschen	*Akademikerkind?* nein	ja	***Total***
Gering	4 4 19 %	7 7 15.9 %	11 11 16.9 %
Mittel	8 7 38.1 %	14 15 31.8 %	22 22 33.8 %
Viel	9 10 42.9 %	23 22 52.3 %	32 32 49.2 %
Total	21 21 100 %	44 44 100 %	65 65 100 %

χ^2=0.504 · df=2 · Cramer's V=0.088 · Fisher's p=0.784

Indifferenztabelle per Hand:

21*11/65 =3.55	44*11/65 =7.45
21*22/65 =7.11	44*22/65 =14.89
21*32/65 =10.34	44*32/65 =21.66

Chi_Quadrat per Hand:

-0.45^2 / 3.55 = 0.06	0.45^2 / 7.45 =0.03
0.89^2 / 7.11 =0.11	-0.89^2 / 14.89 =0.05
-1.34^2 / 10.34 =0.17	1.34^2 / 21.66 =0.08

2 Indifferenztabelle und Chi-Quadrat

Kreuztabelle: Vertrauen nach Bildungshintergrund

Vertrauen in Mitmenschen	*Akademikerkind?* nein	ja	***Total***
Gering	4 4 19 %	7 7 15.9 %	11 11 16.9 %
Mittel	8 7 38.1 %	14 15 31.8 %	22 22 33.8 %
Viel	9 10 42.9 %	23 22 52.3 %	32 32 49.2 %
Total	21 21 100 %	44 44 100 %	65 65 100 %

χ^2=0.504 · df=2 · Cramer's V=0.088 · Fisher's p=0.784

Indifferenztabelle per Hand:

21*11/65 =3.55	44*11/65 =7.45
21*22/65 =7.11	44*22/65 =14.89
21*32/65 =10.34	44*32/65 =21.66

Chi_Quadrat per Hand:

-0.45^2 / 3.55 = 0.06	0.45^2 / 7.45 =0.03
0.89^2 / 7.11 =0.11	-0.89^2 / 14.89 =0.05
-1.34^2 / 10.34 =0.17	1.34^2 / 21.66 =0.08

∑ = 0.5

2 Indifferenztabelle und Chi-Quadrat

Kreuztabelle: Vertrauen nach Bildungshintergrund

Vertrauen in Mitmenschen	*Akademikerkind?* nein	ja	***Total***
Gering	4 4 19 %	7 7 15.9 %	11 11 16.9 %
Mittel	8 7 38.1 %	14 15 31.8 %	22 22 33.8 %
Viel	9 10 42.9 %	23 22 52.3 %	32 32 49.2 %
Total	21 21 100 %	44 44 100 %	65 65 100 %

χ^2=0.504 · df=2 · *Cramer's V=0.088* · *Fisher's p=0.784*

Indifferenztabelle per Hand:

21*11/65 =3.55	44*11/65 =7.45
21*22/65 =7.11	44*22/65 =14.89
21*32/65 =10.34	44*32/65 =21.66

Chi_Quadrat per Hand:

-0.45^2 / 3.55 = 0.06	0.45^2 / 7.45 =0.03
0.89^2 / 7.11 =0.11	-0.89^2 / 14.89 =0.05
-1.34^2 / 10.34 =0.17	1.34^2 / 21.66 =0.08

∑ = 0.5

Cramer's V per Hand:

$$\sqrt{\frac{0.5}{65 * 1}} = 0.088$$

Bewertung Cramer's V:

Nach gängigen Klassifikationen (siehe Vorlesung) drückt Cramer's V hier einen schwachen Zusammenhang aus. Somit stützt es unsere Einschätzung aus der vergleichenden Betrachtung der Verteilungen der AV: Insgesamt sind diese sich zwischen den Akademiker- und Arbeiterkindern verhältnismässig ähnlich, insb. zur hohen Kategorie der AV ergibt sich jedoch eine veritable Prozentsatzdifferenz.

2 Indifferenztabelle und Chi-Quadrat

Kreuztabelle: Vertrauen nach Bildungshintergrund

Vertrauen in Mitmenschen	*Akademikerkind?* nein	ja	***Total***
Gering	4 4 19 %	7 7 15.9 %	11 11 16.9 %
Mittel	8 7 38.1 %	14 15 31.8 %	22 22 33.8 %
Viel	9 10 42.9 %	23 22 52.3 %	32 32 49.2 %
Total	21 21 100 %	44 44 100 %	65 65 100 %

χ^2=0.504 · *df*=2 · *Cramer's V*=0.088 *Fisher's p*=0.784

Indifferenztabelle per Hand:

21*11/65 =3.55	44*11/65 =7.45
21*22/65 =7.11	44*22/65 =14.89
21*32/65 =10.34	44*32/65 =21.66

Chi_Quadrat per Hand:

-0.45^2 / 3.55 = 0.06	0.45^2 / 7.45 =0.03
0.89^2 / 7.11 =0.11	-0.89^2 / 14.89 =0.05
-1.34^2 / 10.34 =0.17	1.34^2 / 21.66 =0.08

∑ = 0.5

Inferenzstatistische Hypothesenbewertung:

- Der Zufall produziert ***sehr oft*** ein Tabellenmuster wie das vorliegende oder ein extremeres. Die Nullhypothese, dass in der Population ***Unabhängigkeit zwischen dem Bildungshintergrund und dem Vertrauen*** besteht, kann auf also Basis des Stichprobenergebnisses ***nicht abgelehnt*** werden (chi^2=0.5, p>0,05).
- Eine unserer Analyse möglicherweise zugrunde liegende **einseitige Hypothese** ist mit der **nominalen Logik des Chi-Quadrat Unabhängigkeitstest nicht vereinbar** und somit auch nicht exakt testbar. Der p-Wert kann also dann nur sehr tentativ und defensiv in die Auswertung eingebunden werden.

2 Indifferenztabelle und Chi-Quadrat

Kreuztabelle: Vertrauen nach Bildungshintergrund

Vertrauen in Mitmenschen	*Akademikerkind?* nein	ja	***Total***
Gering	4 19 %	7 15.9 %	11 16.9 %
Mittel	8 38.1 %	14 31.8 %	22 33.8 %
Viel	9 42.9 %	23 52.3 %	32 49.2 %
Total	21 100 %	44 100 %	65 100 %

χ^2=0.504 · df=2 · Cramer's V=0.088 · Fisher's p=0.760

Kreuztabelle: Vertrauen und Bildungshintergrund

Akademikerkind?	*Vertrauen in Mitmenschen* Gering	Mittel	Viel	***Total***
nein	4 36.4 %	8 36.4 %	9 28.1 %	21 32.3 %
ja	7 63.6 %	14 63.6 %	23 71.9 %	44 67.7 %
Total	11 100 %	22 100 %	32 100 %	65 100 %

χ^2=0.504 · df=2 · Cramer's V=0.088 · Fisher's p=0.780

Spalten und Zeilen vertauschen – was ändert sich?

- Durch Chi^2 indizierte Unabhängigkeitsabweichung ist identisch.
- Dementsprechend auch Cramer‘s V und Teststatistik identisch
- Unter Beibehaltung von UV und AV verändert sich Berechnungslogik von Lambda (welches aber bei Beibehaltung von UV und AV stets identisch bleibt)
- Spaltenprozente der umgeformten Tabelle (wegen des unkonventionellen Aufbaus) nicht mehr im Sinne der antizipierten Kausalrichtung interpretierbar, aber:
- Spaltenprozente der Ausgangstabelle und Zeilenprozente der umgeformten Tabelle sind identisch (oben nicht explizit dargestellt).

2 Indifferenztabelle und Chi-Quadrat

Kreuztabelle: Vertrauen nach Bildungshintergrund

Vertrauen in Mitmenschen	*Akademikerkind?* nein	ja	***Total***
Gering	4 19 %	7 15.9 %	11 16.9 %
Mittel	8 38.1 %	14 31.8 %	22 33.8 %
Viel	9 42.9 %	23 52.3 %	32 49.2 %
Total	21 100 %	44 100 %	65 100 %

χ^2=0.504 · df=2 · Cramer's V=0.088 · Fisher's p=0.760

Kreuztabelle: Vertrauen und Bildungshintergrund

Akademikerkind?	*Vertrauen in Mitmenschen* Gering	Mittel	Viel	***Total***
nein	4 36.4 %	8 36.4 %	9 28.1 %	21 32.3 %
ja	7 63.6 %	14 63.6 %	23 71.9 %	44 67.7 %
Total	11 100 %	22 100 %	32 100 %	65 100 %

χ^2=0.504 · df=2 · Cramer's V=0.088 · Fisher's p=0.780

Fragen?

Weitere Übung

- http://www.suz.uzh.ch/dataforstat/