

8 Daten beurteilen mit Simulationen – Lösungshinweise

Aufgabe 8.1: Gegeben ist der Datensatz zu den Studierenden der Hochschulen in Freiburg und Münster. Untersuchen Sie mit dem simulierten Permutationstest, ob die Studierenden beider Hochschulen sich

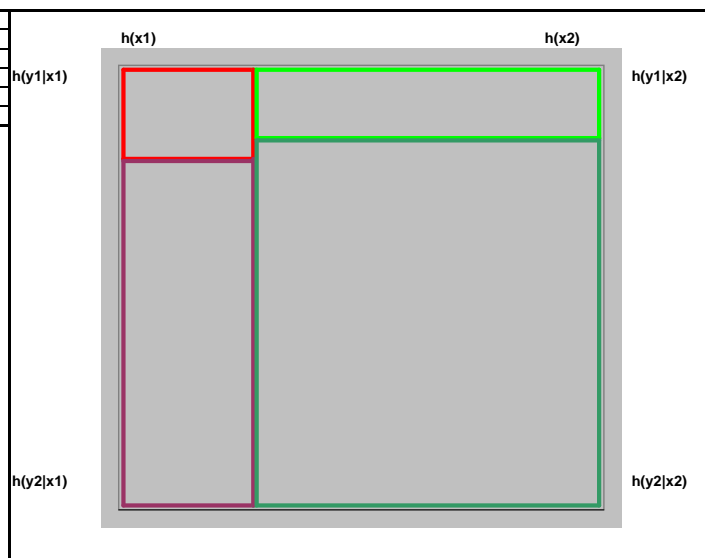
- in der Eigenschaft, noch bei den Eltern zu wohnen, unterscheiden.
- im Beziehungsverhalten unterscheiden.

Lösungsskizze Aufgabe 8.1 Gibt man die Daten in die im online-Material vorhandene Vierfeldertafel in der Datei assoziationsmaß.xls ein (s.u.), so wird das Assoziationsmaß hinsichtlich der Merkmale Hochschule und Wohnverhalten mit $A = 0,047$ angegeben ($x_1 = \text{Freiburg}$, $x_2 = \text{Münster}$; $y_1 = E$, $y_2 = \bar{E}$).

Vierfeldertafel (absolut)			
	x1	x2	Summe
y1	81	170	251
y2	315	912	1227
Summe	396	1082	1478

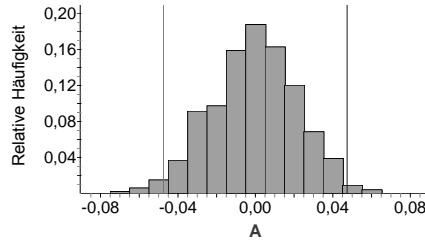
Vierfeldertafel (relativ)			
	x1	x2	Summe
y1	0,05	0,12	0,17
y2	0,21	0,62	0,83
Summe	0,27	0,73	1

$h(x_1) =$	0,27
$h(x_2) =$	0,73
$h(y_1 x_1) =$	0,20
$h(y_2 x_1) =$	0,80
$h(y_1 x_2) =$	0,16
$h(y_2 x_2) =$	0,84



Wir prüfen per Simulation, ob bei einer zufälligen Zuordnung von der Merkmalsausprägungen zu den Merkmalen Hochschule und Wohnverhalten ein Assoziationsmaß von mindestens 0,047

häufig vorkommt oder selten ist. 1000 Simulationen ergeben folgende Verteilung von Assoziationsmaßen.



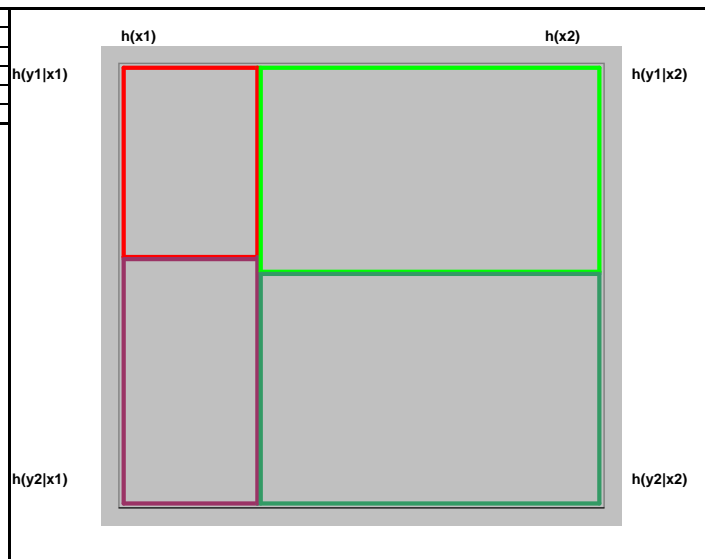
Es ist $h_{1000}(-0,047 < A < 0,047) \approx 0,015 \approx P(-0,047 < A < 0,047)$. Auf der Basis der Testsimulation ließe sich also die Hypothese ablehnen, dass das Assoziationsmaß in der Stichprobe von 0,047 zufällig zustande gekommen ist. In diesem Fall (siehe Vierfelder-Tafel) könnte man von einem leicht höheren Hang der PH-Studierenden ausgehen, zu Hause zu wohnen.

In gleicher Weise behandeln wir das Beziehungsverhalten. Hier ergibt sich das Assoziationsmaß von $A = -0,015$ ($x_1 = \text{Freiburg}$, $x_2 = \text{Münster}$; $y_1 = S$, $y_2 = \bar{S}$)

Vierfeldertafel (absolut)			
	x1	x2	Summe
y1	172	486	658
y2	224	552	776
Summe	396	1038	1434

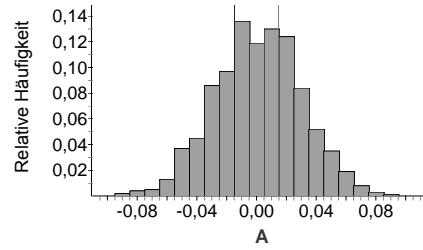
Vierfeldertafel (relativ)			
	x1	x2	Summe
y1	0,12	0,34	0,46
y2	0,16	0,38	0,54
Summe	0,28	0,72	1

$h(x_1) =$	0,28
$h(x_2) =$	0,72
$h(y_1 x_1) =$	0,43
$h(y_2 x_1) =$	0,57
$h(y_1 x_2) =$	0,47
$h(y_2 x_2) =$	0,53



1000 Simulationen ergeben die auf der Folgeseite dargestellte Verteilung von Assoziationsmaßen. Hier ist bereits optisch erkennbar, dass das Assoziationsmaß von -0,015 bei der zufälligen Zuordnung der Merkmale Hochschule und Beziehungsverhalten nicht auffällig ist. Tatsächlich ist

$h_{1000}(-0,015 < A < 0,015) \approx 0,657$. Man kann also von der Unabhängigkeit beider Merkmale ausgehen.



Aufgabe 8.2: Gegeben ist der Datensatz zu den Studierenden der Hochschulen in Freiburg und Münster. Untersuchen Sie mit dem simulierten Permutationstest, ob die Studierenden beider Geschlechter sich

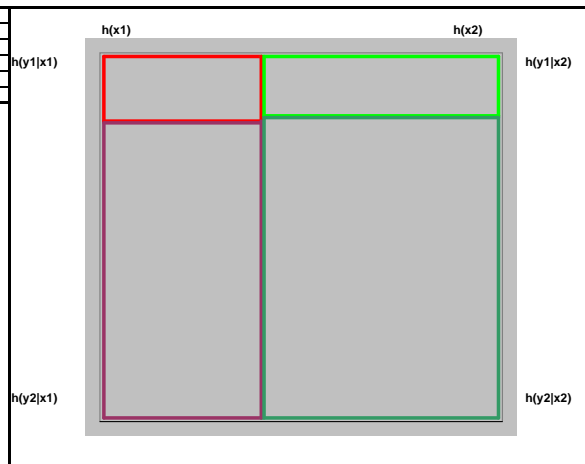
- in der Eigenschaft, noch bei den Eltern zu wohnen, unterscheiden.
- in der Präferenz für das Fortbewegungsmittel Fahrrad unterscheiden.

Lösungsskizze Aufgabe 8.2 Wir gehen wie in Aufgabe 8.1 vor. Das Assoziationsmaß hinsichtlich der Merkmale Geschlecht ($x_1 = m$, $x_2 = w$; $y_1 = E$, $y_2 = \bar{E}$) und Wohnverhalten ist mit $A = 0,014$ wiederum sehr klein.

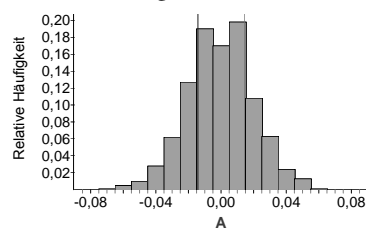
Vierfeldertafel (absolut)			
	x_1	x_2	Summe
y_1	104	147	251
y_2	478	748	1226
Summe	582	895	1477

Vierfeldertafel (relativ)			
	x_1	x_2	Summe
y_1	0,07	0,10	0,17
y_2	0,32	0,51	0,83
Summe	0,39	0,61	1

$h(x_1) =$	0,39
$h(x_2) =$	0,61
$h(y_1 x_1) =$	0,18
$h(y_2 x_1) =$	0,82
$h(y_1 x_2) =$	0,16
$h(y_2 x_2) =$	0,84



Die Simulation der zufälligen Zuordnung der Merkmale Geschlecht und Wohnverhalten ergibt bei 1000 Wiederholungen folgende Verteilung von Assoziationsmaßen.

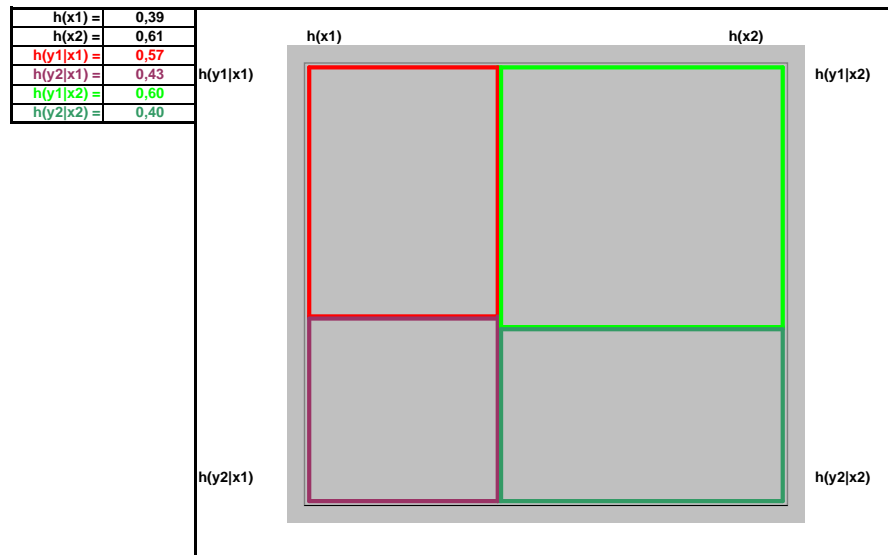


Hier ist wiederum bereits optisch erkennbar, dass das Assoziationsmaß von $-0,014$ bei der zufälligen Zuordnung der Merkmale Geschlecht und Wohnverhalten nicht auffällig ist. Tatsächlich ist $h_{1000}(-0,014 < A < 0,01447) \approx 0,478$. Man kann also von der Unabhängigkeit beider Merkmale ausgehen.

In gleicher Weise behandeln wir den Zusammenhang von Geschlecht und Beförderungsmittel. Hier ergibt sich das Assoziationsmaß von $A = -0,027$ ($x_1 = m$, $x_2 = w$; $y_1 = F$, $y_2 = \bar{F}$).

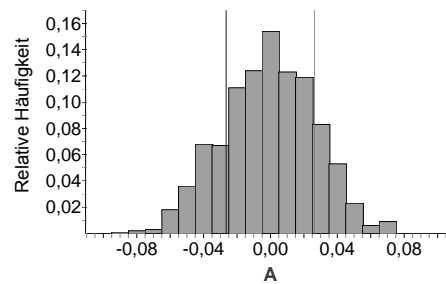
Vierfeldertafel (absolut)			
	x1	x2	Summe
y1	333	536	869
y2	247	359	606
Summe	580	895	1475

Vierfeldertafel (relativ)			
	x1	x2	Summe
y1	0,23	0,36	0,59
y2	0,17	0,24	0,41
Summe	0,39	0,61	1



1000 Simulationen ergeben die auf der Folgeseite dargestellte Verteilung von Assoziationsmaßen. Auch in diesem Fall ist das Assoziationsmaß von $-0,027$ bei der zufälligen Zuordnung der Merkmale Geschlecht und Beförderungsmittel nicht auffällig.

Tatsächlich ist $h_{1000}(-0,027 < A < 0,027) \approx 0,358$. Man kann also auch in diesem Fall von der Unabhängigkeit beider Merkmale ausgehen.



Aufgabe 8.3: Betrachten Sie eine der beiden Hochschulen, Freiburg oder Münster. Bestimmen Sie per Berechnung, Simulation und Bootstrap-Verfahren Konfidenzintervalle zu der in der Grundgesamtheit auftretenden Häufigkeit (p) für

- Single,
- Radfahrer,
- „schlechte“ Schüler mit einem Abiturschnitt über 3,0.

Lösungsskizze Aufgabe 8.3 Wir betrachten im Folgenden nur die Studierenden der PH Freiburg und Konfidenzintervalle zum Konfidenzniveau 0,05. Aus der Stichprobe erhalten wir:

$$H_{396}(S) = 172, h_{396}(S) \approx 0,434; \quad H_{396}(F) = 116, h_{396}(F) \approx 0,293; \quad H_{396}(\bar{G}) = 37, h_{396}(\bar{G}) \approx 0,093$$

wobei \bar{G} die Studierenden mit einem Abiturschnitt über 3,0 repräsentiert.

Wir bestimmen für alle drei Merkmalsausprägungen zwei Grenzwahrscheinlichkeiten p_u und p_o symmetrisch um die empirisch ermittelten Häufigkeiten, so dass für die binomialverteilten Zufallsgrößen X_S , X_F und $X_{\bar{G}}$ gilt

- $P_{p_u}(X_S \leq 172) + P_{p_o}(X_S \geq 172) \approx 0,95$,
- $P_{p_u}(X_F \leq 116) + P_{p_o}(X_F \geq 116) \approx 0,95$ und
- $P_{p_u}(X_{\bar{G}} \leq 37) + P_{p_o}(X_{\bar{G}} \geq 37) \approx 0,95$.

Weiterhin bestimmen wir p_u und p_o durch Simulation (1000 Wiederholungen). Schließlich ermitteln wir die mit dem Bootstrap-Verfahren erzeugten Konfidenzintervalle.

Wir erhalten folgende Konfidenzintervalle:

	Single	Fahrrad	Abinote
Berechnung	$I = [0,386; 0,483]$	$I = [0,248; 0,338]$	$I = [0,064; 0,123]$
Simulation	$I = [0,385; 0,484]$	$I = [0,246; 0,340]$	$I = [0,063; 0,124]$
Bootstrap	$I = [0,386; 0,482]$	$I = [0,250; 0,336]$	$I = [0,065; 0,122]$

Alle drei Konstruktionen erzeugen damit nahezu identische Konfidenzintervalle. In 95 Prozent der Fälle enthält ein so konstruiertes Konfidenzintervall den wahren Parameter, d.h. den wahren Anteil der drei Merkmalsausprägungen (unter der Annahme, dass die Daten gut sind).

Aufgabe 8.4: In Kapitel 3.7 hatte wir folgende Unterschiede zwischen den Studierenden in Freiburg und Münster betrachtet:

„Gegeben sind unten die Datensätze zu Studierenden (Münster und Freiburg) zu der Parteipräferenz, dem bevorzugten Beförderungsmittel sowie zum Erhalt von BAföG. Sind Grünen-Wähler umweltbewusster? Sind BAföG-Bezieher SPD-Wähler?“

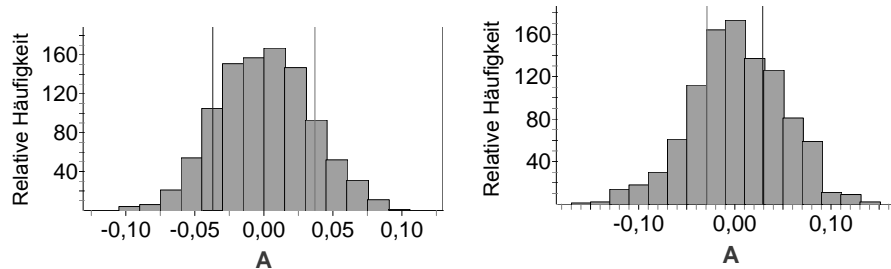
Beurteilen Sie nun die deskriptiv ermittelten Unterschiede.

	Grüne	CDU/FDP	Summe
Auto	13	9	22
Fahrrad	116	132	248
Summe	129	141	270

	BAföG	kein BAföG	Summe
SPD	49	107	156
CDU/FDP	65	163	228
Summe	114	270	384

Lösungsskizze Aufgabe 8.4 Wir können hier ein weiteres Mal den simulierten Permutationstest anwenden und damit die bisher rein deskriptiv erzeugten Aussagen zum Zusammenhang der jeweiligen Merkmale beurteilen.

Im ersten Beispiel hatte sich ein Assoziationsmaß $A = 0,037$, im zweiten von $A = 0,029$ ergeben. Bei 1000 Simulationen, in denen im ersten Fall das Beförderungsmittel zufällig der Parteienpräferenz und im zweiten Fall die Parteipräferenz zufällig dem BAföG-Status zugeordnet werden, hat sich folgende Verteilung von Assoziationsmaßen ergeben (links zu Datensatz 1, rechts zu Datensatz 2).



Da im ersten Fall $h_{1000}(|A| \geq 0,037) \approx 0,273$ gilt, kann von der Unabhängigkeit der Merkmale Parteipräferenz und Beförderungsmittel ausgegangen werden. Auch im zweiten Fall ergibt sich wegen $h_{1000}(|A| \geq 0,029) \approx 0,564$, dass weiterhin von der Unabhängigkeit der Merkmale Parteipräferenz und BAföG-Status ausgegangen werden kann.