Bachelorarbeit

Analyse der Studierendenbefragung der Universitätsbibliothek München mithilfe von Assoziationsregeln

Autor:
Robert Pietsch

Betreuerin:
Prof. Dr. Bettina Grün

Institut für Statistik
Ludwig-Maximilians-Universität München

3. August 2011

Zusammenfassung

In dieser Arbeit werden die Grundlagen der Datenanalyse mit Assoziationsregeln motiviert und eingeführt. Es werden die wichtigsten Bedeutungsmaße erklärt und anhand ihrer Eigenschaften verglichen. Zudem werden die Prinzipien und eine Veranschaulichung des Apriori-Algorithmus zur Gewinnung von Assoziationsregeln vorgestellt. Diese Methoden werden dann auf einen Datensatz der Universitätsbibliothek München angewendet. Hierzu wird das R-Paket **arules** verwendet. Ein wichtiges Resultat der Analyse ist, dass sich aus dem Studienfach viele Aussagen zum Lernverhalten der Studierenden treffen lassen.

Inhaltsverzeichnis

1. Einleitung

„We are drowning in information and starved for knowledge.“

- Naisbitt (1982)

1.1. Data Mining

Dieses Zitat von John Naisbitt beschreibt treffend die Motivation für „Data Mining“: Aus großen Datenmengen soll Wissen gewonnen werden, d. h., Muster, Strukturen und Gesetzmäßigkeiten. Dieses Wissen kann dazu genutzt werden, Voraussagen zu wagen, die einen praktischen Nutzen haben. Bekannte Verfahren des Data Mining sind z. B. Regressionsanalyse, Clusteranalyse, Zeitreihenanalyse und Assoziationsanalyse.

(vgl. Borgelt 1997)

1.2. Inhalt der Arbeit

In dieser Arbeit wird eine bestimmte Art der Assoziationsanalyse, nämlich Assoziationsregeln, vorgestellt und an einem Datensatz angewendet. Im Methodenteil (Kapitel 2) werden Bedeutungsmaße von Assoziationsregeln und ein Algorithmus zur Gewinnung von Assoziationsregeln erklärt.
Der im Anwendungsteil (Kapitel 3–4) betrachtete Datensatz wurde im Rahmen des „Statistischen Praktikums“ im Wintersemester 2010/11 an der Ludwig-Maximilians-Universität München (LMU) erhoben und enthält die Antworten der Studierendenbefragung der Universitätsbibliothek (UB). In Zusammenarbeit mit der UB der LMU wurden Daten zum Lernverhalten der Studierenden, deren Zufriedenheit mit der UB und deren personenbezogene Angaben gesammelt. Diese Daten wurden in Eifler u. a. (2011) schon univariat aufbereitet und werden hier nun multivariat mit Assoziationsregeln untersucht.

Diese Untersuchungen besitzen eventuell praktischen Nutzen für die Universitätsbibliothek. Eine Fragestellung wäre z. B., ob sich prognostizieren lässt, welche Angebote der UB bestimmte Fächergruppen, wie z. B. Naturwissenschaften, besonders nutzen. Eine weitere

Fragestellung wäre, ob sich aus dem Studienfach vorhersagen lässt, wo die Studierenden bevorzugt lernen. Bei beiden exemplarischen Fragen muss jedoch die Kausalität beachtet werden: D. h., wieso z. B. Studierende eines bestimmten Studienfachs an diesem Lernort bevorzugt lernen, ist nicht aus der Assoziationsstruktur ersichtlich und erfordert daher Hintergrundwissen.

1.3. Allgemeines zu Assoziationsregeln

Assoziationsregeln stammen aus dem Bereich der Statistik, der im Allgemeinen als unüberwachtes Lernen („Unsupervised Learning“) bezeichnet wird. Im unüberwachten Lernen wird versucht, auf Eigenschaften der gemeinsamen Dichte $\mathbb{P}(X)$ eines Zufallsvektor X (mit p Variablen und N Beobachtungen) rückzuschließen, ohne auf Fehlerwahrscheinlichkeiten oder ein direktes Maß für Erfolg zurückgreifen zu können. Da im unüberwachten Lernen oft die gemeinsame Dichte $\mathbb{P}(X)$ so hoch-dimensional ist, dass sie praktisch nicht berechnet werden kann, wird nur nach Bereichen gesucht, in denen $\mathbb{P}(X)$ (sehr) groß ist. Für hoch-dimensionale, binäre Daten versuchen Assoziationsregeln auf einfache Art und Weise Bereiche mit einer großen Dichte zu beschreiben (vgl. Hastie u. a. 2001, Seite 486–487).

Die Verfahren des unüberwachten Lernens sind häufig eng mit der Informatik verbunden. Auch bei den Assoziationsregeln sind diese Wurzeln sichtbar, so würde z. B. ein gelernter Statistiker das Support-Maß (siehe Punkt 2.1.1) von vorne herein als relative Häufigkeit bezeichnen.

Häufig werden Assoziationsregeln im Bereich der Warenkorbanalyse angewendet, d. h., welche Waren werden zusammengekauft, und welche Waren implizieren den Kauf bestimmter weiterer Waren. Ein Anwendungsbeispiel aus dem täglichen Leben hierfür sind die Produktempfehlungen von Amazon, die im Wesentlichen auf Assoziationsregeln beruhen.

2. Methoden

In diesem Kapitel wird eine Auswahl an Bedeutungsmaßen, die im Rahmen der Assoziationsregeln eine Rolle spielen, vorgestellt. Diese werden nach wünschenswerten Eigenschaften verglichen und es wird auf die Problematik verschiedener Maße eingegangen. Im zweiten Teil wird der Apriori-Algorithmus vorgestellt, der mit Hilfe von Bedeutungsmaßen Assoziationsregeln aus Daten gewinnt.

Definition 1 (Notation) *Sei $I = \{i_1, i_2, \ldots, i_d\}$ eine Menge von d binären Attributen, die im folgenden Items genannt werden. Sei $T = \{t_1, t_2, \ldots, t_n\}$ eine Menge an Transaktionen, auch Datensatz genannt. Jede Transaktion aus T besitzt eine einzigartige Transaktions-ID (TID) und beinhaltet eine Untermenge von I. Eine Menge die k Items enthält, wird als k-Menge bezeichnet. Eine Assoziationsregel ist definiert als die Implikation der Form $X \Rightarrow Y$, wobei $X, Y \subseteq I$ und $X \cap Y = \emptyset$. Die Mengen X und Y werden Voraussetzung (oder Left-Hand-Side bzw. LHS) und Konsequenz (oder Right-Hand-Side bzw. RHS) einer Regel genannt. $\mathbb{P}(X)$ bezeichnet die relative Häufigkeit der Menge X.*

(vgl. Tan u. a. 2005, Seite 329 und Hahsler u. a. 2005, Seite 1)

2.1. Bedeutungsmaße

Für den regelgenerierenden Prozess werden nur die Maße Support (siehe Punkt 2.1.1) und Confidence (siehe Punkt 2.1.2) benötigt. Jedoch werden so meistens zu viele Assoziationsregeln generiert. Es bietet sich daher an, die Regeln mit Hilfe weiterer Maße zusätzlich zu filtern bzw. zu ordnen (siehe Hahsler u. Hornik 2007, Seite 2).

Alle hier vorgestellten Maße werden an einem fiktiven Datensatz eines Supermarktes vorgeführt (siehe Tabelle 2.1).

Transaktions ID	Items
1	Brot, Milch
2	Brot, Butter
3	Bier
4	Brot, Butter, Milch
5	Brot, Butter

Tabelle 2.1.: Beispiel: Datensatz aus Hahsler u. a. (2005, Seite 2)

Dieser Datensatz beinhaltet 5 Transaktionen und als Menge von Items $I = \{$Bier, Brot, Butter, Milch$\}$.

2.1.1. Support

Support ist das wichtigste Maß für die Bedeutung einer Assoziationsregel. Es ist definiert als der Anteil aller Transaktionen, die die Elemente einer bestimmten Assoziationsregel beinhalten, vonn der Grundmenge aller betrachteten Transaktionen. D. h., dieses Maß gibt an, wie häufig eine Assoziationsregel im Datensatz vorkommt.

$$\text{supp}(X \Rightarrow Y) = \text{supp}(X \cup Y) = \mathbb{P}(X \cup Y) \tag{2.1}$$

Hieraus folgt $\text{supp}(X \Rightarrow Y) = \text{supp}(Y \Rightarrow X)$. Dies bedeutet, dass es sich beim Support um ein symmetrisches Maß handelt. Eine weitere, wichtige Eigenschaft des Supports ist, dass Support nach unten abgeschlossen („downward closure property“) bzw. monoton fallend („anti-monotone property“) ist. Diese Eigenschaft spielt beim Apriori-Algorithmus eine entscheidende Rolle (siehe hierzu Definition 3).

Support ist auch deshalb ein wichtiges Maß, da Assoziationsregeln mit niedrigem Support möglicherweise nur zufällig auftreten und somit irrelevant sind. Zudem sind Assoziationsregeln mit niedrigem Support für Anwender in der Praxis unter Umständen uninteressant, wenn sie in zu wenigen Transaktionen auftreten.

Bei dem Beispiel aus Tabelle 2.1 tritt die Menge $\{$Brot, Butter$\}$ in 3 von 5 Transaktionen auf. Damit ergibt sich $\text{supp}(\text{Brot} \cup \text{Butter}) = \frac{3}{5} = 0,6$.

(vgl. Hahsler u. a. 2005, Seite 2 und Tan u. a. 2005, Seite 330)

2.1.2. Confidence

Das zweite wichtige Maß, das im Rahmen des regelgenerierenden Prozesses verwendet wird, ist Confidence. Während Support ein Maß für die Bedeutung einer Assoziationsregel darstellt, ist Confidence ein Maß für die Treffsicherheit einer Assoziationsregel. Confidence entspricht der Wahrscheinlichkeit, dass die Konsequenz einer Regel (*hier:* Y) auftritt unter der Bedingung, dass eine Transaktion die Voraussetzungen (*hier:* X) erfüllt.

$$\mathrm{conf}(X \Rightarrow Y) = \frac{\mathrm{supp}(X \Rightarrow Y)}{\mathrm{supp}(X)} = \frac{\mathbb{P}(X \cup Y)}{\mathbb{P}(X)} = \mathbb{P}(Y|X) \tag{2.2}$$

Im Allgemeinen gilt nicht, dass $\mathbb{P}(Y|X) = \mathbb{P}(X|Y)$ bzw. $\mathrm{conf}(X \Rightarrow Y) = \mathrm{conf}(Y \Rightarrow X)$. Somit ergibt sich, dass Confidence ein asymmetrisches Maß ist. Ein daraus resultierendes Problem bei Confidence ist, dass es zu irreführenden Resultaten kommen kann, insbesondere dann, wenn der Support der Konsequenz einer Assoziationsregel höher ist als die Confidence der Assoziationsregel.

Bei dem Beispiel aus Tabelle 2.1 ergibt sich die Assoziationsregel $\{\mathrm{Brot}\} \Rightarrow \{\mathrm{Butter}\}$. Damit ergibt sich $\mathrm{conf}(\mathrm{Brot} \Rightarrow \mathrm{Butter}) = \frac{\mathrm{supp}(\mathrm{Brot} \cup \mathrm{Butter})}{\mathrm{supp}(\mathrm{Brot})} = \frac{0{,}6}{0{,}8} = 0,75$. Das bedeutet, dass für 75% der Transaktionen, die Brot beinhalten, diese Assoziationsregel zutrifft.

(vgl. Hahsler u. a. 2005, Seite 2 und Tan u. a. 2004, Seite 298)

2.1.3. Lift

Lift, in manchen, besonders älteren Publikationen, oft auch als Interest bezeichnet, wird häufig als Maß für die Abweichung von der statistischen Unabhängigkeit herangezogen.

$$\mathrm{lift}(X \Rightarrow Y) = \frac{\mathrm{conf}(X \Rightarrow Y)}{\mathrm{supp}(Y)} = \frac{\mathrm{supp}(X \Rightarrow Y)}{\mathrm{supp}(X)\mathrm{supp}(Y)} = \frac{\mathbb{P}(X \cup Y)}{\mathbb{P}(X)\mathbb{P}(Y)} \tag{2.3}$$

Man sieht, dass es sich beim Lift um ein symmetrisches Maß handelt. Ein Lift von 1 gibt an, dass die Items X und Y so zusammen auftreten, wie man es unter der Unabhängigkeitsannahme vermuten würde. Es wird allgemein angenommen, dass ein lift > 1 sich ergänzende (positiv korrelierte) Items und ein lift < 1 sich ersetzende (negativ korrelierte) Items angibt.

Das Problem des Lifts ist, dass er sensibel bezüglich $\mathrm{supp}(X)$ und $\mathrm{supp}(Y)$ ist. Seltene Mengen von Items (d. h., Mengen mit niedrigem Support) können sehr hohe Lift-Werte verursachen. Dieses Problem wird beim Apriori-Algorithmus insofern umgangen, da nur Mengen betrachtet werden, die über einem gewissen Schwellenwert des Supports liegen (siehe Punkt 2.2.2). Dennoch tendiert der Lift dazu, Assoziationsregeln, die knapp über dem Support-Schwellenwert liegen, höhere Werte zuzuordnen. Im Gegensatz dazu produzieren Mengen mit einem sehr hohen Support einen Lift nahe bei 1 (dies impliziert statistische Unabhängigkeit), obwohl die Items der Menge möglicherweise hoch korreliert sind. Ein Beispiel aus Tan u. a. (2004, Seite 304) verdeutlicht dies (siehe Tabelle 2.2).

	Y	$\bar{Y}$	
X	890	0	890
$\bar{X}$	0	10	10
	890	10	900

Tabelle 2.2.: $\text{lift}(X \Rightarrow Y) = 1.012$, $\phi = 1$

Aus dieser Eigenschaft ergibt sich, dass der Lift tendenziell Assoziationsregeln mit seltenen Items bevorzugt.

Bei dem Beispiel aus Tabelle 2.1 ergibt sich die Assoziationsregel {Brot} $\Rightarrow$ {Butter}. Damit ergibt sich $\text{lift}(\text{Brot} \Rightarrow \text{Butter}) = \frac{\text{supp}(\text{Brot} \cup \text{Butter})}{\text{supp}(\text{Brot})\text{supp}(\text{Butter})} = \frac{0{,}6}{0{,}8 \cdot 0{,}6} = 1{,}25$.

(vgl. Tan u. a. 2004, Seite 298 und Hahsler u. Hornik 2007, Seite 7–8)

2.1.4. ϕ-Koeffizient

Der ϕ-Koeffizient ist das Analogon zum Korrelationskoeffizienten nach Bravais und Pearson für binäre Merkmale.

$$\phi = \frac{\mathbb{P}(X,Y)\mathbb{P}(\bar{X},\bar{Y}) - \mathbb{P}(X,\bar{Y})\mathbb{P}(\bar{X},Y)}{\sqrt{\mathbb{P}(X)\mathbb{P}(Y)\mathbb{P}(\bar{X})\mathbb{P}(\bar{Y})}} \tag{2.4}$$

Der ϕ-Koeffizient ist ein symmetrisches Maß. Der ϕ-Koeffizient erscheint intuitiv als ein gutes Maß für Assoziation, jedoch hat dieses Maß den Nachteil, wie auch Support, Confidence und Lift, dass es nicht invariant gegenüber Änderungen von Zeilen- und Spaltenskalierungen ist. Ein Beispiel von Mosteller aus Tan u. a. (2004, Seite 294–295) verdeutlicht dies: In Tabelle 2.3 wird der Zusammenhang zwischen Geschlecht und einer Note untersucht. Man beachte, dass Tabelle (b) doppelt soviele Männer und 10-mal so viele Frauen beinhaltet, jedoch ist die relative Leistung beider Populationen in beiden Tabellen identisch. Das Maß sollte in beiden Fällen das Gleiche angeben und somit unabhängig von den relativen Häufigkeiten der Teilpopulationen sein. Vergleiche hierzu Punkt 2.1.6.

	Mann	Frau	
Gut	2	3	5
Schlecht	1	4	5
	3	7	10

(a)

	Mann	Frau	
Gut	4	30	34
Schlecht	2	40	42
	6	70	76

(b)

Tabelle 2.3.: Geschlecht-Noten Beispiel aus Tan u. a. (2004, Seite 295)

Für (a) entspricht $\phi = 0,218$ und für (b) $\phi = 0,129$. Die unterliegende Assoziation sollte intuitiv äquivalent sein, ist sie aber dem ϕ-Koeffizienten nach nicht.

Bei dem Beispiel aus Tabelle 2.1 ergibt sich die Assoziationsregel {Brot} ⇒ {Butter}. Damit ergibt sich $\phi = \frac{\mathbb{P}(\text{Brot, Butter})\mathbb{P}(\overline{\text{Brot}},\overline{\text{Butter}}) - \mathbb{P}(\text{Brot},\overline{\text{Butter}})\mathbb{P}(\overline{\text{Brot}},\text{Butter})}{\sqrt{\mathbb{P}(\text{Brot})\mathbb{P}(\text{Butter})\mathbb{P}(\overline{\text{Brot}})\mathbb{P}(\overline{\text{Butter}})}} = \frac{0,6 \cdot 0,2 - 0,2 \cdot 0}{\sqrt{0,6 \cdot 0,8 \cdot 0,4 \cdot 0,2}} = 0,612$.

(vgl. Tan u. a. 2004)

2.1.5. Odds-Ratio

Ein Maß, das hingegen invariant gegenüber Änderungen von Zeilen- und Spaltenskalierungen ist, ist das Odds-Ratio θ (siehe Punkt 2.1.6). Dieses Maß ist daher geeignet für Datensätze, die nominale Merkmale enthalten.

$$\theta = \frac{\mathbb{P}(X,Y)\mathbb{P}(\bar{X},\bar{Y})}{\mathbb{P}(X,\bar{Y})\mathbb{P}(\bar{X},Y)} = \frac{\mathbb{P}(X,Y)/\mathbb{P}(X,\bar{Y})}{\mathbb{P}(\bar{X},Y)/\mathbb{P}(\bar{X},\bar{Y})} \tag{2.5}$$

Das Odds-Ratio θ kann jeden nicht negativen Wert annehmen, wobei $\theta = 1$ der Unabhängigkeit von X und Y entspricht. Für $1 < \theta < \infty$ ist die Chance, Y in einer Transaktion zu finden, in der X auftritt, multiplikativ um θ größer. Für $0 < \theta < 1$ ist dementsprechend die Chance um den Faktor θ kleiner. Je weiter θ von 1 entfernt ist (in beiden Richtungen), desto stärker ist die Assoziation. Zwei Werte stehen für die gleiche Assoziation, aber in unterschiedliche Richtung, falls sie das Inverse voneinander sind. Sobald eine der Kombinationen von $X, \bar{X}$ und $Y, \bar{Y}$ nicht auftritt, nimmt θ den Wert 0 oder ∞ an. Da Assoziationsregeln im Allgemeinen auf große Datensätze angewendet werden, tritt dieser Fall eher selten auf.

Bei dem Beispiel aus Tabelle 2.1 ergibt sich die Assoziationsregel {Brot} ⇒ {Butter}. Damit ergibt sich $\theta = \frac{\mathbb{P}(\text{Brot, Butter})\mathbb{P}(\overline{\text{Brot}},\overline{\text{Butter}})}{\mathbb{P}(\text{Brot},\overline{\text{Butter}})\mathbb{P}(\overline{\text{Brot}},\text{Butter})} = \frac{0,6 \cdot 0,2}{0,2 \cdot 0} = \infty$.

(vgl. Agresti 2002, Seite 44–45 und Tan u. a. 2004, Seite 297)

2.1.6. Vergleich der Maße

In der Literatur ist man sich einig, dass es nicht das perfekte Maß $\mathcal{M}$ gibt. Jedoch hat Piatetsky-Shapiro (1991, Seite 232) 3 wünschenswerte Eigenschaften (P1–P3) vorgeschlagen. Im Artikel von Tan u. a. (2004) werden noch 5 weitere Eigenschaften (O1–O5) vorgeschlagen. Diese Eigenschaften lassen sich zum Vergleich der Maße heranziehen.

P1 $\mathcal{M} = 0$, falls X und Y statistisch unabhängig sind

P2 $\mathcal{M}$ steigt monoton mit $\mathbb{P}(X, Y)$, wenn $\mathbb{P}(X)$ und $\mathbb{P}(Y)$ konstant bleiben

P3 $\mathcal{M}$ sinkt monoton mit $\mathbb{P}(X)$ (oder $\mathbb{P}(Y)$), wenn die restlichen Parameter ($\mathbb{P}(X, Y)$ und $\mathbb{P}(Y)$ bzw. $\mathbb{P}(X)$) konstant bleiben

O1 Symmetrie bei Variablen-Permutation

O2 Zeilen- und Spaltenskalierungen-Invarianz

O3 Antisymmetrie bei Zeilen- oder Spalten-Permutation

O4 Inversions-Invarianz

O5 Null-Invarianz

Dies bedeutet, Maße, …

- die O1 nicht erfüllen, helfen die Stärke der Regeln $X \Rightarrow Y$ und $Y \Rightarrow X$ zu unterscheiden.
- die O2 erfüllen, sind - wie in Punkt 2.1.5 erwähnt - nützlich für Datensätze mit nominalen Variablen.
- die O3 nicht erfüllen, unterscheiden nicht zwischen positiver und negativer Korrelation für X und Y.
- die O4 erfüllen, sind symmetrische binäre Maße, d. h., Nullen („tritt nicht auf“) und Einser („tritt auf“) werden gleich gewichtet. Diese Eigenschaft ist z. B. in der Warenkorb-Analyse unerwünscht.
- die O5 erfüllen, sind z. B. für die Warenkorb-Analyse interessant, da sie gemeinsames Auftreten wichtiger werten als gemeinsames Nicht-Auftreten.

Keines der Maße erfüllt alle Eigenschaften, wie Tabelle 2.4 zeigt.

Maß	Wertebereich	P1	P2	P3	O1	O2	O3	O4	O5
Support	$0 \cdots 1$	Nein	Ja	Nein	Ja	Nein	Nein	Nein	Nein
Confidence	$0 \cdots 1$	Nein	Ja	Nein	Nein	Nein	Nein	Nein	Ja
Lift	$0 \cdots 1 \cdots \infty$	Nein	Ja	Ja	Ja	Nein	Nein	Nein	Nein
ϕ-Koeffizient	$-1 \cdots 0 \cdots 1$	Ja	Ja	Ja	Ja	Nein	Ja	Ja	Nein
Odds-Ratio θ	$0 \cdots 1 \cdots \infty$	Nein	Ja	Ja	Ja	Ja	Nein	Ja	Nein

Tabelle 2.4.: Eigenschaften der Maße (vgl. Tan u. a. 2004, Seite 299)

(vgl. Tan u. a. 2004)

2.2. Apriori-Algorithmus

Dieser Punkt basiert größtenteils auf dem Buch Tan u. a. (2005, Kapitel 6).

Neben dem Apriori-Algorithmus gibt es eine Vielzahl an Algorithmen (Eclat, FP-growth, ASSOC, OPUS, ...) als Lösung für die unten beschriebene Problemstellung (siehe Wikipedia 2011).

2.2.1. Problemstellung

Das Entdecken von Assoziationsregeln kann als folgendes Problem aufgefasst werden.

Definition 2 (Entdecken von Assoziationsregeln) *Gegeben eine Menge an Transaktionen T, finde alle Assoziationsregeln mit Support $S \geq \min(supp)$ und Confidence $C \geq \min(conf)$, wobei $\min(supp)$ bzw. $\min(conf)$ die Schwellenwerte für Support bzw. Confidence sind.*

Der Versuch, dieses Problem mit schierer Rechenleistung zu lösen („brute-force"), scheitert in der Regel an der Zahl aller möglichen Assoziationsregeln. Bei einem Datensatz mit d Items, beträgt die Anzahl an möglichen Assoziationsregeln:

$$R = 3^d - 2^{d+1} + 1. \tag{2.6}$$

Daher verwendet der Apriori-Algorithmus (wie nahezu alle obengenannten Algorithmen) die Strategie, das Problem in zwei Unteraufgaben aufzuteilen.

1. **Finden häufiger Mengen**, die den min(supp)-Schwellenwert erfüllen.
2. **Finden von starken Assoziationsregeln** aus Mengen von Schritt 1, die den min(conf)-Schwellenwert erfüllen.

Der Apriori-Algorithmus setzt dies um, wie in Abbildung 2.1 gezeigt.

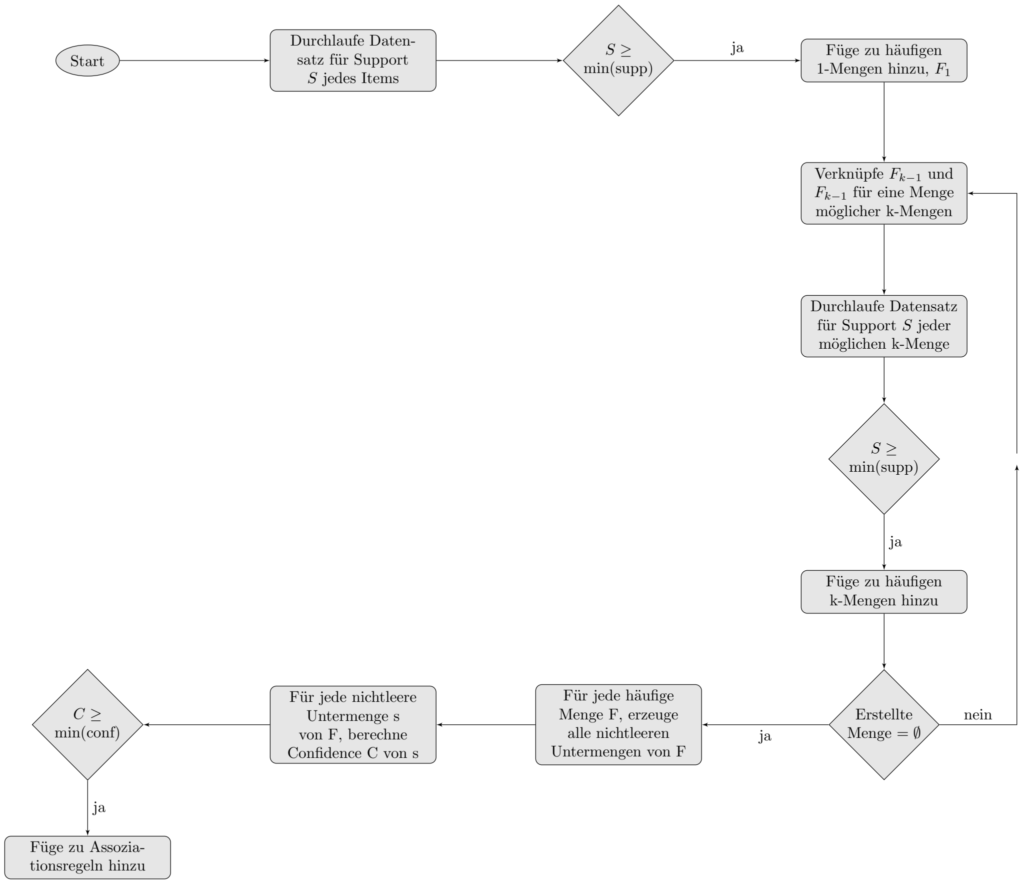

Abbildung 2.1.: Flowchart Apriori-Algorithmus (vgl. Gamil 2010)

2.2.2. Finden häufiger Mengen

Damit der Algorithmus nicht jede – der $2^d - 1$ möglichen – Menge in Schritt 1 überprüfen muss, nutzt der Apriori-Algorithmus die in Punkt 2.1.1 erwähnte Eigenschaft des Supports aus.

Definition 3 (Monotonie Eigenschaft) *Sei I eine Menge an Items, J die Potenzmenge von I und $|J| = 2^{|I|}$ die Mächtigkeit der Menge J. Ein Maß $\mathcal{M}$ ist monoton fallend (bzw. nach unten abgeschlossen), falls*

$$\forall X, Y \in J : (X \subseteq Y) \rightarrow \mathcal{M}(Y) \leq \mathcal{M}(X).$$

Dies bedeutet, falls X eine Untermenge von Y ist, darf $\mathcal{M}(Y)$ nicht größer als $\mathcal{M}(X)$ sein.

Auf dieser Eigenschaft beruht folgendes Theorem:

Theorem 1 (Apriori-Prinzip) *Wenn eine Menge häufig ist, sind auch alle ihre Untermengen häufig.*

Häufige Mengen werden als $F_{\#Elemente}$ bezeichnet.

Für den Support des Beispiels heißt das, wenn die Menge {Brot, Butter} häufig ist, sind auch die Untermengen {Brot} und {Butter} häufig. Ist aber die Menge {Bier, Brot} nicht häufig, sind auch alle Obermengen {Bier, Brot, Milch}, {Bier, Brot, Butter} und {Bier, Brot, Butter, Milch} nicht häufig.

Auf diesem Theorem beruht die Generierung möglicher häufiger Mengen, die als $C_{\#Elemente}$ bezeichnet werden.

Definition 4 ($F_{k-1} \times F_{k-1}$-Methode) *Sei $A = \{a_1, a_2, \ldots, a_{k-1}\}$ und $B = \{b_1, b_2, \ldots, b_{k-1}\}$ ein Paar häufiger $(k-1)$-Mengen. A und B werden vereinigt, falls sie folgende Bedingung erfüllen:*

$$a_i = b_i \, (\textit{für } i = 1, 2, \ldots, k-2) \textit{ and } a_{k-1} \neq b_{k-1}$$

D. h., zwei häufige $(k-1)$-Mengen werden nur vereinigt, wenn ihre ersten $(k-2)$ Elemente identisch sind. Um doppelte Mengen zu vermeiden, sollten die Elemente (lexikographisch) geordnet sein.

Zum Beispiel werden die häufigen Mengen {Brot, Butter} und {Brot, Milch} zu {Brot, Butter, Milch} vereinigt. Die Mengen {Brot, Butter} und {Butter, Milch} hingegen werden nicht vereinigt, da sie sich im ersten Element unterscheiden.

Das „Finden von häufigen Mengen“ ist für das Beispiel in Abbildung 2.2 illustriert.

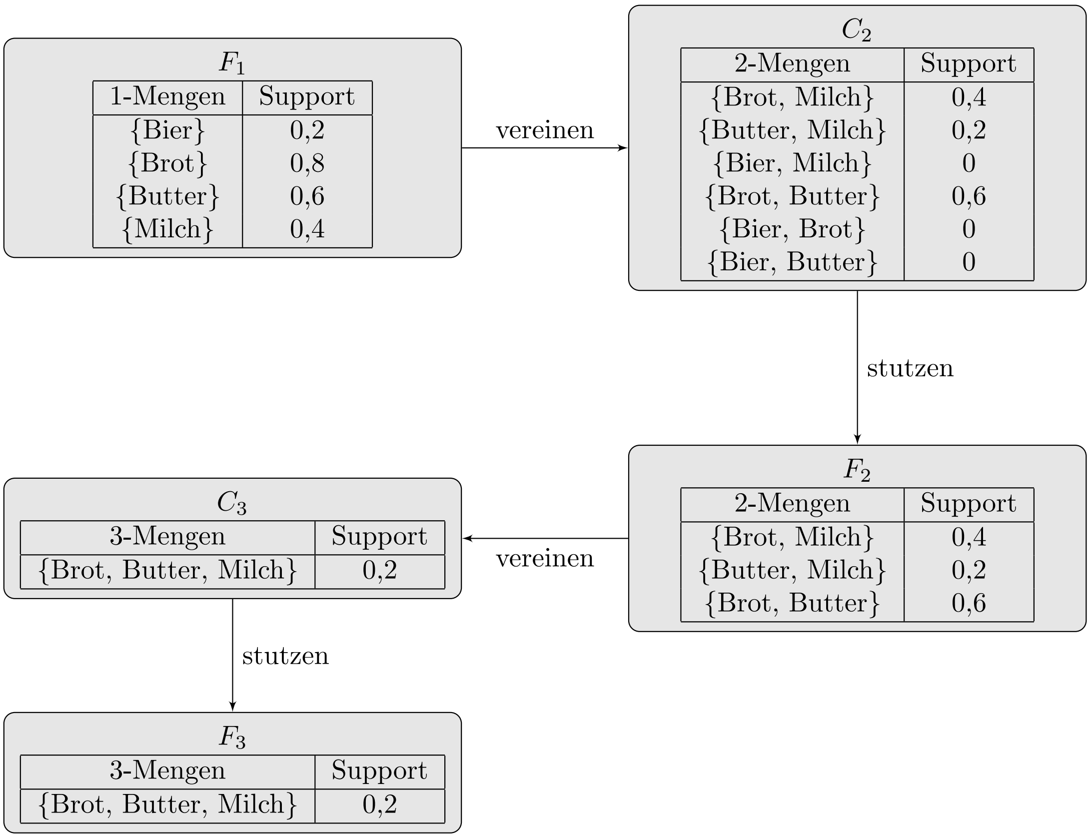

Abbildung 2.2.: Schritt 1 des Apriori-Algorithmus, mit min(supp) $= 0,2$ (vgl. Gamil 2010)

Daraus ergeben sich die häufigen Mengen: {Bier}, {Brot}, {Butter}, {Milch}, {Brot, Milch}, {Butter, Milch}, {Brot, Butter}, {Brot, Butter, Milch}.

2.2.3. Finden von starken Assoziationsregeln

Assoziationsregeln werden im 2. Schritt aus häufigen Mengen generiert. Jede häufige k-Menge F kann bis zu $2^k - 2$ Assoziationsregeln hervorbringen. Eine Assoziationsregel wird generiert, indem die Menge F in zwei nicht leere Untermengen X und $F - X = Y$ partitioniert wird, so dass $\text{conf}(X \Rightarrow Y) \geq \min(\text{conf})$ gilt.

Da Confidence im Gegensatz zu Support keine Monotonie-Eigenschaft besitzt (vgl. Definition 3), wird folgendes Theorem zum effizienten Stutzen von Assoziationsregeln verwendet.

Theorem 2 (Confidence basiertes Stutzen) *Falls eine Regel $X \Rightarrow F - X$ nicht den*

Confidence-Schwellenwert erfüllt, erfüllt auch keine Regel $X' \Rightarrow F - X'$, wobei X' eine Untermenge von X ist, den Confidence-Schwellenwert.

Anschaulich heißt dies in dem Beispiel, wenn aus der Menge {Brot, Butter, Milch}, die Regel {Brot, Butter} ⇒ {Milch} generiert wird, diese aber nicht den Confidence-Schwellenwert (d. h., conf(Brot, Butter ⇒ Milch) < min(conf)) erfüllt, können auch gleich die Regeln {Brot} ⇒ {Butter, Milch} und {Butter} ⇒ {Brot, Milch} weggeworfen werden.

Das „Finden von starken Assoziationsregeln“ ist für das Beispiel in Abbildung 2.3 illustriert.

Regel	supp($X \cup Y$)	supp(X)	conf($X \Rightarrow Y$)
Milch ⇒ Brot	0,4	0,4	1
Brot ⇒ Milch	0,4	0,8	0,5
Milch ⇒ Butter	0,2	0,4	0,5
Butter ⇒ Milch	0,2	0,6	0,33
Brot ⇒ Butter	0,6	0,8	0,75
Butter ⇒ Brot	0,6	0,6	1
Brot, Milch ⇒ Butter	0,2	0,4	0,5
Butter, Milch ⇒ Brot	0,2	0,2	1
Brot, Butter ⇒ Milch	0,2	0,6	0,33
Milch ⇒ Brot, Butter	0,2	0,4	0,5
Brot ⇒ Butter, Milch	0,2	0,8	0,25
Butter ⇒ Brot, Milch	0,2	0,6	0,33

stutzen

Regel	supp($X \cup Y$)	supp(X)	conf($X \Rightarrow Y$)
Milch ⇒ Brot	0,4	0,4	1
Brot ⇒ Butter	0,6	0,8	0,75
Butter ⇒ Brot	0,6	0,6	1
Butter, Milch ⇒ Brot	0,2	0,2	1

Abbildung 2.3.: Schritt 2 des Apriori Algorithmus, mit min(conf) $= 0,7$ (vgl. Gamil 2010)

3. Datensatz

Der Datensatz stammt wie unter Punkt 1.2 erwähnt aus dem Statistischen Praktikum des Wintersemesters 2010/11 am Institut für Statistik der LMU und enthält die Antworten der Studierendenbefragung der Universitätsbibliothek. Die Datenerhebung ist im dazugehörigen Bericht Eifler u. a. (2011) ausführlich beschrieben. Besonders wichtig ist, dass die Stichprobe, aufgrund ihrer nicht zufälligen Ziehung, nicht repräsentativ ist. Daher lassen sich die Ergebnisse nicht auf die Grundgesamtheit aller Studierenden anwenden.

3.1. Datenbeschreibung

Die Daten liegen in einem „data frame“ im wide-Format vor.

Von den ursprünglichen 1819 Beobachtungen wurden erst alle 217 unvollständigen Beobachtungen und im folgenden Schritt noch 6 Beobachtungen mit unsinnigen bzw. unmöglichen Antworten aus dem Datensatz entfernt. Dies bedeutet, dass alle Teilnehmer, die nicht zwischen 10 und 90 Jahren alt oder nicht zwischem dem 1. und 90. Semester sind, aus dem Datensatz gelöscht wurden. Diese Grenzen wurden aufgrund eines offiziellen Datensatzes der LMU gewählt, der im oben erwähnten Statistischen Praktikum Verwendung fand (siehe Eifler u. a. 2011). Der Datensatz enthält nach dieser Bereinigung noch 1596 Beobachtungen.

In der Originalform beinhaltet der Datensatz 188 Variablen. Von diesen 188 Variablen wurden die 4 Variablen `id`, `submitdate`, `startlanguage` und `refurl` entfernt, da sie für den Zweck dieser Arbeit keine relevanten Informationen bereitstellen. Zudem wurden noch alle Freitextantworten aus dem Datensatz genommen, da diese nicht quantitativ auswertbar sind. Im Gegenzug wurden 4 weitere Variablen aus den vorhandenen Variablen erstellt, wie im Folgenden beschrieben.
Die Variable `studienfach` gruppiert die Hauptfächer der Studierenden zu ihren Oberkategorien „Mathematik und Naturwissenschaften“, „Medizin und Gesundheitswissenschaften“, „Geisteswissenschaften“ und „Rechts-, Wirtschafts- und Sozialwissenschaften“.
Die Variable `nutzer` trennt die Studierenden danach, ob sie in einer Bibliothek der LMU

„fast immer“ oder „häufig“ lernen.
Die Variable anzahl_services summiert die Anzahl benutzter Services pro Student auf.
Die Variable zscore stellt einen Zufriedenheitsscore dar, und wird über die Variablen Zufriedenheit_kompe, Zufriedenheit_hilfe, Zufriedenheit_antwz, Zufriedenheit_supz, Services_Aush, Services_Flyer, Services_Infot, Services_HPInf, Services_Leits und Services_VerIn gebildet. Diese Variablen beschreiben die Zufriedenheit der Studierenden mit dem Support bzw. den Services der UB. Die Variable, die angibt, ob man Zusatzinformationen erhielt (Zufriedenheit_zusz), blieb bewusst unberücksichtigt, da eine Zustimmung nicht zwangsläufig positiv ist (vgl. Eifler u. a. 2011, Seite 38). Die Variable zscore verändert sich um +2 für jede „trifft immer zu“ bzw. „sehr gut“ Antwort, um +1 für jede „trifft meistens zu“ bzw. „gut“ Antwort, um −1 für jede „trifft selten zu“ bzw. „schlecht“ Antwort und um −2 für jede „trifft nie zu“ bzw. „sehr schlecht“ Antwort. Somit schwankt die Variable zscore zwischen −20 und +20.
Der Datensatz enthält somit schließlich 176 Variablen.
Eine nähere Beschreibung zu jeder einzelnen Variable ist in Eifler u. a. (2011, Seite 60–74) zu finden. Die in dieser Arbeit verwendeten Variablen werden im Abschnitt 3.4 nochmals gesondert erläutert.

3.2. Datenmanipulation

Generell werden zur Anwendung von Assoziationsregeln binäre Daten benötigt. In der klassischen Warenkorbanalyse steht eine 1 (Präsenz) für eine erworbene Ware und jede nicht erworbene Ware wird mit einer 0 (Abwesenheit) kodiert. Für den hier verwendeten Datensatz steht dementsprechend eine 1 für eine mit „ja“ beantwortete Frage und eine 0 für eine nicht oder mit „nein“ beantwortete Frage. Im Allgemeinen wird die Präsenz einer Variable (d. h., eine 1) als informativer als ihre Abwesenheit empfunden (siehe Tan u. a. 2005, Seite 328). Somit handelt es sich um asymmetrisch binäre Variablen. Dieser Umstand bevorzugt dementsprechend asymmetrisch binäre Maße, d. h., Maße, die die Eigenschaft O4 aus Abschnitt 2.1.6 nicht erfüllen (vgl. Tan u. a. 2004, Seite 301).

Da aber viele Fragen im Datensatz eine ordinale oder nominale Antwortskala vorgeben, muss beachtet werden, wie diese Variablen transformiert werden. Würde z. B. die nominale Variable sex für das Geschlecht als (asymmetrisch) binäre Variable aufgefasst werden, könnten keine Regeln mit männlichen Studenten gefunden werden, da diese als 0 kodiert sind. Um dies zu vermeiden, muss die Variable theoretisch in 2 Variablen geteilt werden, eine Variable für männlich: ja/nein und eine Variable weiblich: ja/nein. Praktisch wird dieser Schritt vom Programmpaket **arules** (Hahsler u. a. 2011 und Hahsler u. a. 2005) der freien Software R (R Development Core Team 2011) intern automatisch durchgeführt. Selbiges gilt für ordinale Variablen.

Problematischer sind hingegen metrische Variablen. In diesem Datensatz sind das `semester`, `alter`, `anzahl_services`, `zscore` und die 4 Variablen zu den Lernvolumina (`Lernvolumen_sbegi`, `Lernvolumen_stime`, `Lernvolumen_send` und `Lernvolumen_svac`). Diese Variablen müssen, damit sie mit **arules** weiterverarbeitet werden können, erst ordinalisiert werden.
Daher wird die Variable `semester` in die Gruppen „Studienanfänger" (1–2 Semester) (Anteil 12%), „Bachelor" (3–7) (49%), Master (8–12) (30%) und „Langzeitstudent" (13–90) (9%) aufgeteilt.
Die Variable `alter` erhält 3 Alterskategorien: „jung" (10–25 Jahre) (Anteil 74%), „mittel" (26–55) (23%) und „alt" (56–90) (1%).
Die Variable `anzahl_services` wird unterteilt in „wenige" (0–3 Services) (Anteil 9%), „eher wenige" (4–7) (34%), „eher viele" (8–12) (44%) und „viele" (13–20) (13%).
Die Variable `zscore` wird in die Gruppen „Unzufrieden" (`zscore`: -20–-1) (Anteil 11%), „Neutral" (0–5) (39%) und „Zufrieden" (6–20) (50%) aufgeteilt.
Die 4 Variablen zu den Lernvolumina werden nach ihren jeweiligen Quantilen mit den Ausprägungen „wenig", „eher wenig", „eher viel" und „viel" umkodiert.

Wie in Abschnitt 3.1 erwähnt, liegen die Daten in Form eines „data frame" vor. Das Paket **arules** benötigt als Eingabe aber Daten der Klasse „transactions". Sobald das „data frame" nur aus faktoriellen Variablen besteht, lässt sich das „data frame" einfach mit dem Aufruf `as(data, "transactions")` in die Klasse „transactions" zwängen. Dieser Aufruf muss nicht explizit ausgeführt werden, da jede Funktion des **arules**-Pakets diesen Aufruf, wenn nötig, ausführt.

3.3. Interne Speicherung

Intern werden Daten der Klasse „transactions" in einer dünnbesetzten Matrix („sparse matrix") gespeichert, da eine typische Transaktion nur wenige von vielen möglichen Elementen enthält. Diese komprimierte Darstellungsform bietet technisch den Vorteil, das weniger Speicher benötigt wird, da nur ein Vektor mit Indices der `1`-Einträge und ein Vektor, der angibt, wann eine Transaktion beginnt, gespeichert werden muss.

Für das Beispiel aus Tabelle 2.1 ergibt sich folgende (dünnbesetzte) Matrix (siehe Tabelle 3.1).

		Items			
		Bier	Brot	Butter	Milch
Transaktionen	1	0	1	0	1
	2	0	1	1	0
	3	1	0	0	0
	4	0	1	1	1
	5	0	1	1	0

Tabelle 3.1.: Beispiel: Datensatz aus Hahsler u. a. (2005, Seite 7)

Zur Speicherung dieser Tabelle 3.1 wird nun der Vektor der Indices der `1`-Einträge $(2, 4, 2, 3, 1, 2, 3, 4, 2, 3)$ und der Vektor, der angibt, wann eine Transaktion beginnt $(1, 3, 5, 6, 9)$ gespeichert. Dies bedeutet, dass die Transaktion 1 mit Element 1 des Indices-Vektors beginnt und mit Element 2 aufhört, da Transaktion 2 bereits mit dem dritten Element beginnt.

(vgl. Hahsler u. a. 2005, Seite 4–7)

3.4. Verwendete Variablen

Es werden insgesamt 45 verschiedene Variablen verwendet.
Für eine Tabelle aller hier verwendeten Variablen siehe Tabelle A.1 im Anhang A.1.

4. Anwendung und Ergebnisse

Die Auswertung des in Abschnitt 3 vorgestellten Datensatzes, erfolgt mit dem Paket **arules** (Hahsler u. a. 2011 und Hahsler u. a. 2005) für die freie Software R (R Development Core Team 2011). Es ist zu beachten, dass die dortige Implementation des Apriori-Algorithmus nur Assoziationsregeln mit einem Element in der Konsequenz bzw. RHS erstellt (siehe Hahsler u. a. 2011, Seite 10).

Die Idee, in dem kompletten Datensatz Assoziationsregeln finden zu wollen, musste schnell verworfen werden, da hier jede Transaktion (sprich: Beobachtung) zu viele Items enthält. Aus diesem Grund werden immer nur bestimmte Teile des Datensatzes betrachtet.
Die allgemeine Vorgehensweise in diesem Kapitel ist, dass nahezu alle möglichen Assoziationsregeln gewonnen werden. D. h., Support- und Confidence-Schwellenwerte werden sehr niedrig angesetzt (*hier:* beide $\leq 0,05$). Die gewonnenen Regeln werden zum einen nach ihren objektiven Bedeutungsmaßen, insbesondere dem Odds-Ratio θ entsprechend, als interessant oder eben uninteressant eingestuft (eine Einstufung nach Lift bzw. ϕ-Koeffizienten ändert hier praktisch nichts). Zum anderen werden subjektive Interessantheits-Maße im Data-Mining-Prozess verwendet (siehe Geng u. Hamilton 2006). D. h., es wird Hintergrundwissen angewendet, um z. B. die rechte Seite von Regeln von Anfang an festzulegen.

4.1. Geschlecht, Studienfach, Lernort

Wer lernt 'häufig' oder 'fast immer' an einem Lernort?

In diesem Abschnitt wird untersucht, ob sich aus dem Studienfach und dem Geschlecht eines Studierenden eine Aussage über den Lernort treffen lässt. Es werden nur universitätseigene Lernorte (d. h., Fakultät, Fachbibliothek, UB-Zentrale und Studentenbibliothek) betrachtet. Es werden nur Assoziationsregeln gesucht, die in der Konsequenz Lernorte mit „häufig“ oder „fast immer“ aufweisen.
Es ergibt sich der folgende ausgewählte Auszug an Regeln (vgl. Anhang B.1):

- {`studienfach=MaNa`, `sex=weiblich`} ⇒ {`Lernorte_Fak=häufig`} $\theta = 2,39$
- {`studienfach=MaNa`, `sex=männlich`} ⇒ {`Lernorte_Fak=häufig`} $\theta = 1,80$
- {`studienfach=MaNa`} ⇒ {`Lernorte_Fak=häufig`} $\theta = 2,30$
- {`studienfach=GeiWi`, `sex=weiblich`} ⇒ {`Lernorte_FachB=häufig`} $\theta = 1,43$
- {`studienfach=GeiWi`, `sex=männlich`} ⇒ {`Lernorte_FachB=häufig`} $\theta = 1,39$
- {`studienfach=GeiWi`} ⇒ {`Lernorte_FachB=häufig`} $\theta = 1,67$

Diese Regeln haben alle einen Odds-Ratio $\geq 1,39$ und legen die Vermutung nahe, dass nur das Studienfach eine Rolle bei der Wahl des Lernortes spielt. Die dazugehörige Analyse (siehe Anhang B.1.1) ergibt:

- {`studienfach=MaNa`} ⇒ {`Lernorte_Fak=häufig`} $\theta = 2,30$
- {`studienfach=MaNa`} ⇒ {`Lernorte_UBS=häufig`} $\theta = 1,42$
- {`studienfach=GeiWi`} ⇒ {`Lernorte_FachB=häufig`} $\theta = 1,67$
- {`studienfach=ReWiSo`} ⇒ {`Lernorte_UBZ=häufig`} $\theta = 1,54$
- {`studienfach=ReWiSo`} ⇒ {`Lernorte_UBS=häufig`} $\theta = 1,43$
- {`studienfach=ReWiSo`} ⇒ {`Lernorte_FachB=fast immer`} $\theta = 1,36$

D. h., Studierende, die ein mathematisches und/oder naturwissenschaftliches Studienfach studieren, haben eine $2,3$-mal so hohe Chance, „häufig“ in der Fakultät zu lernen. Die Interpretation ist für die anderen Regeln analog.
Untersucht man vollständigerweise, ob sich aus dem Geschlecht doch auf den bevorzugten Lernort zurückschließen lässt, ergeben sich als Regeln mit dem höchsten Odds-Ratio (siehe Anhang B.1.2):

- {`sex=männlich`} ⇒ {`Lernorte_FachB=fast immer`} $\theta = 1,35$
- {`sex=männlich `} ⇒ {`Lernorte_Fak=häufig`} $\theta = 1,13$
- {`sex=weiblich`} ⇒ {`Lernorte_FachB=häufig`} $\theta = 1,10$

Diese Regeln sollten nicht überinterpretiert werden, da zum einen die Bedeutungsmaße (Odds-Ratio, Lift und ϕ-Koeffizient) nicht stark vom Wert für statistische Unabhängigkeit abweichen und zum anderen die Unterscheidung bei der Fachbibliothek zwischen „fast immer“ und „häufig“ nicht klar definiert und somit sehr ungenau ist.

Der bevorzugte Lernort lässt sich somit nur gut aus dem Studienfach des Befragten vorhersagen.

4.2. Geschlecht, Studienfach

Wer studiert welches Fach?

Oft werden Studienrichtungen klischeehaft bestimmten Geschlechtern zugeordnet, daher wird hier untersucht, ob sich diese Klischees in diesem Datensatz wiederfinden lassen. Die Konsequenz der Regeln soll daher das Studienfach beinhalten. Anhang B.2 zeigt den zugehörigen R-Output. Es ergeben sich die interessanten Regeln:

- {`sex=männlich`} $\Rightarrow$ {`studienfach=MaNa`} $\theta = 2,30$
- {`sex=männlich`} $\Rightarrow$ {`studienfach=ReWiSo`} $\theta = 1,74$
- {`sex=weiblich`} $\Rightarrow$ {`studienfach=GeiWi`} $\theta = 2,28$
- {`sex=weiblich`} $\Rightarrow$ {`studienfach=Med`} $\theta = 1,13$

Die letzte Regel ist praktisch nur der Vollständigkeit halber aufgeführt, da vor allem Lift und ϕ-Koeffizient statistische Unabhängigkeit suggerieren. Die hier gewonnenen Assoziationsregeln zeigen insbesondere an, dass männliche Studenten eine um den Faktor $2,3$ erhöhte Chance haben, Mathematik und Naturwissenschaften zu studieren. Weibliche Studentinnen hingegen haben eine um den Faktor $2,28$ erhöhte Chance, eine Geisteswissenschaft zu studieren.
Abschließend lässt sich sagen, dass die vermuteten Klischees im Datensatz zu finden sind.

4.3. Alter, Geschlecht, Semester, Studienfach, Lernzeiten

Wer lernt zu welchen Zeiten?

Hier sollen Assoziationsregeln aus den personenbezogenen Variablen Alter, Geschlecht, Semester und Studienfach und allen Variablen zu Lernzeiten gefunden. Dieses Vorhaben erweist sich ohne weitere Restriktionen als sehr unübersichtlich. Deswegen werden die Variablen zu den Lernzeiten unterteilt nach „während des Semesters“ und „während der Ferien“. Zudem werden für die Konsequenz lediglich die Lernzeiten-Variablen verwendet. Bei diesen Variablen ist insbesondere interessant, wann Studierende lernen (und nicht, wann sie *nicht* lernen).
Zuerst werden nur die Lernzeiten „während des Semesters“ betrachtet. Alle Analysen, die mehr als eine Variable in der linken Seite beinhalteten, erwiesen sich hier nicht als aussagekräftiger, sondern nur als unübersichtlicher. Daher wird immer nur eine Variable

in der Voraussetzungsseite betrachtet.
Mit dem Studienfach als Voraussetzung ergeben sich als interessante Regeln (siehe Anhang B.3.1):

- {studienfach=Med} ⇒ {Lernzeiten_woche_abend_sem=1} $\theta = 3,51$
- {studienfach=Med} ⇒ {Lernzeiten_sonn_vorm_sem=1} $\theta = 2,38$
- {studienfach=Med} ⇒ {Lernzeiten_sams_abend_sem=1} $\theta = 1,94$
- {studienfach=Med} ⇒ {Lernzeiten_sams_vorm_sem=1} $\theta = 1,81$
- {studienfach=Med} ⇒ {Lernzeiten_sonn_nachm_sem=1} $\theta = 1,45$
- {studienfach=Med} ⇒ {Lernzeiten_sams_nachm_sem=1} $\theta = 1,35$
- {studienfach=MaNa} ⇒ {Lernzeiten_sams_abend_sem=1} $\theta = 1,47$
- {studienfach=MaNa} ⇒ {Lernzeiten_woche_abend_sem=1} $\theta = 1,44$
- {studienfach=MaNa} ⇒ {Lernzeiten_woche_nacht_sem=1} $\theta = 1,26$
- {studienfach=MaNa} ⇒ {Lernzeiten_sonn_nachm_sem=1} $\theta = 1,22$
- {studienfach=MaNa} ⇒ {Lernzeiten_sams_nachm_sem=1} $\theta = 1,12$
- {studienfach=GeiWi} ⇒ {Lernzeiten_woche_vorm_sem=1} $\theta = 1,36$
- {studienfach=GeiWi} ⇒ {Lernzeiten_woche_nachm_sem=1} $\theta = 1,27$
- {studienfach=ReWiSo} ⇒ {Lernzeiten_woche_nachm_sem=1} $\theta = 1,24$
- {studienfach=ReWiSo} ⇒ {Lernzeiten_woche_vorm_sem=1} $\theta = 1,11$

Diese Regeln lassen sich subjektiv so zusammenfassen, dass Studierende der Medizin und Gesundheitswissenschaften bzw. der Mathematik und Naturwissenschaften hauptsächlich am Wochenende oder unter der Woche abends lernen, währende Studierende der Geisteswissenschaften bzw. der Rechts-, Wirtschafts- und Sozialwissenschaften hauptsächlich unter der Woche vormittags und nachmittags lernen. Jedoch sind die Bedeutungsmaße nur für Studierende der Medizin und Gesundheitswissenschaften besonders groß.
Untersucht man, ob sich die Lernzeiten im Semester durch das Geschlecht vorhersagen lassen, ergeben sich als Regeln mit den höchsten Odds-Ratios (siehe Anhang B.3.2):

- {sex=weiblich} ⇒ {Lernzeiten_sonn_vorm_sem=1} $\theta = 1,47$
- {sex=weiblich} ⇒ {Lernzeiten_sams_vorm_sem=1} $\theta = 1,32$
- {sex=weiblich} ⇒ {Lernzeiten_woche_vorm_sem=1} $\theta = 1,19$
- {sex=männlich} ⇒ {Lernzeiten_woche_nacht_sem=1} $\theta = 1,46$
- {sex=männlich} ⇒ {Lernzeiten_sams_nacht_sem=1} $\theta = 1,31$
- {sex=männlich} ⇒ {Lernzeiten_sams_abend_sem=1} $\theta = 1,25$
- {sex=männlich} ⇒ {Lernzeiten_woche_nachm_sem=1} $\theta = 1,24$

Diese Regeln vermitteln als allgemeines Bild, dass weibliche Studierende eine erhöhte

Chance haben, vormittags zu lernen, wobei männliche Studenten eine erhöhte Chance haben, abends bzw. nachts zu lernen. Es ist jedoch zu beachten, dass keine der zugrunde liegenden Assoziationsregeln sonderlich stark vom Wert für statistische Unabhängigkeit bei den Bedeutungsmaßen Lift und ϕ-Koeffizienten abweicht. Dieselbe Aussage lässt sich für die Lernzeiten in den Ferien treffen (siehe Anhang B.3.3).
Für die Assoziationsstruktur zwischen dem Semester und den Lernzeiten während des Semesters erhält man die interessanten Regeln (siehe Anhang B.3.4):

- {`semester=Master`} $\Rightarrow$ {`Lernzeiten_woche_vorm_sem=1`} $\theta = 1,81$
- {`semester=Master`} $\Rightarrow$ {`Lernzeiten_woche_nachm_sem=1`} $\theta = 1,47$
- {`semester=Studienanfänger`} $\Rightarrow$ {`Lernzeiten_woche_abend_sem=1`} $\theta = 1,56$
- {`semester=Studienanfänger`} $\Rightarrow$ {`Lernzeiten_sams_nachm_sem=1`} $\theta = 1,13$
- {`semester=Studienanfänger`} $\Rightarrow$ {`Lernzeiten_sonn_nachm_sem=1`} $\theta = 1,11$
- {`semester=Bachelor`} $\Rightarrow$ {`Lernzeiten_sonn_nachm_sem=1`} $\theta = 1,40$
- {`semester=Bachelor`} $\Rightarrow$ {`Lernzeiten_sonn_vorm_sem=1`} $\theta = 1,37$
- {`semester=Bachelor`} $\Rightarrow$ {`Lernzeiten_woche_abend_sem=1`} $\theta = 1,28$
- {`semester=Bachelor`} $\Rightarrow$ {`Lernzeiten_sams_vorm_sem=1`} $\theta = 1,23$
- {`semester=Bachelor`} $\Rightarrow$ {`Lernzeiten_sams_nachm_sem=1`} $\theta = 1,17$

Diese Regeln kann man wieder zusammenfassen: Studierende, die sich in den ersten Semestern befinden, lernen eher am Wochenende und unter der Woche abends, und Studierende in fortgeschrittenen Semestern lernen eher unter der Woche vormittags und nachmittags. Hier gilt aber, dass die Zusammenfassung der Assoziationsregeln sehr subjektiv geprägt ist und zudem die Bedeutungsmaße keine großen Werte annehmen (vor allem für niedrige Semester).
Mit dem Alter als linke Seite, ließen sich keine wirklich interessanten Assoziationsregeln mit den Lernzeitem im Semester finden (siehe Anhang B.3.5).

Bei den Lernzeiten in den Ferien ergeben sich – für alle personenbezogenen Variablen in der linken Seite – als Regeln mit den höchsten Odds-Ratios (siehe Anhang B.3.6):

- {`alter=jung`} ⇒ {`Lernzeiten_woche_nie_fer=1`} $\theta = 2,31$
- {`semester=Bachelor`} ⇒ {`Lernzeiten_woche_nie_fer=1`} $\theta = 2,25$
- {`alter=jung`,
  `semester=Bachelor`} ⇒ {`Lernzeiten_woche_nie_fer=1`} $\theta = 2,21$
- {`sex=weiblich`,
  `studienfach=GeiWi`,
  `alter=jung`,
  `semester=Master`} ⇒ {`Lernzeiten_woche_vorm_fer=1`} $\theta = 2,19$

Diese Regeln legen dar, dass „junge“ und/oder Studierende in frühen Semestern nie in den Ferien (zumindest unter der Woche) lernen. Die 4. Regel zeigt zwar eine Regel für bestimmte „junge“ Studierende an, die unter der Woche in den Ferien liegen, jedoch zeigen weitere Untersuchungen, dass dies hauptsächlich an den Variablen `studienfach=GeiWi` und `semester=Master` hängt.
Die Analyse mit Studienfach und Semester als Voraussetzung ergibt (siehe Anhang B.3.7):

- {`semester=Bachelor`} ⇒ {`Lernzeiten_woche_nie_fer=1`} $\theta = 2,25$
- {`semester=Bachelor`} ⇒ {`Lernzeiten_sams_nie_fer=1`} $\theta = 1,79$
- {`studienfach=GeiWi`,
  `semester=Master`} ⇒ {`Lernzeiten_woche_vorm_fer=1`} $\theta = 2,03$
- {`studienfach=GeiWi`,
  `semester=Master`} ⇒ {`Lernzeiten_sams_vorm_fer=1`} $\theta = 1,98$
- {`studienfach=GeiWi`,
  `semester=Master`} ⇒ {`Lernzeiten_woche_nachm_fer=1`} $\theta = 1,83$
- {`semester=Langzeitstudent`} ⇒ {`Lernzeiten_woche_abend_fer=1`} $\theta = 2,02$

Wie in der vorherigen Analyse sieht man, dass Studierende im „Bachelor“ eine erhöhte Chance haben, in den Ferien praktisch nie zu lernen. Studierende im „Master“, die zu dem eine Geisteswissenschaft studieren, lernen eher schon in den Ferien. Es stellt sich die Frage, ob dies auch auf alle „Master“-Studierende übertragbar ist. Daher schließt sich eine Analyse nur mit dem Semester als linke Seite an (siehe Anhang B.3.8):

- {`semester=Bachelor`} $\Rightarrow$ {`Lernzeiten_woche_nie_fer=1`} $\theta = 2,25$
- {`semester=Bachelor`} $\Rightarrow$ {`Lernzeiten_sams_nie_fer=1`} $\theta = 1,79$
- {`semester=Master`} $\Rightarrow$ {`Lernzeiten_woche_vorm_fer=1`} $\theta = 1,63$
- {`semester=Master`} $\Rightarrow$ {`Lernzeiten_sams_vorm_fer=1`} $\theta = 1,62$
- {`semester=Master`} $\Rightarrow$ {`Lernzeiten_woche_abend_fer=1`} $\theta = 1,56$
- {`semester=Master`} $\Rightarrow$ {`Lernzeiten_woche_nachm_fer=1`} $\theta = 1,54$
- {`semester=Langzeitstudent`} $\Rightarrow$ {`Lernzeiten_woche_abend_fer=1`} $\theta = 2,02$
- {`semester=Langzeitstudent`} $\Rightarrow$ {`Lernzeiten_sams_nachm_fer=1`} $\theta = 1,48$

Zusammenfassend kann man sagen, dass Studierende in fortgeschrittenen Semestern eine erhöhte Chance (*oder Risiko*) haben, in der vorlesungsfreien Zeit zu lernen. Es kam bisher in keiner Regel die Ausprägung `semester=Studienanfänger` vor. Dieser Umstand ändert sich bei Betrachtung des Alters und des Semesters als Voraussetzung für die Lernzeit in den Ferien (siehe Anhang B.3.9):

- {`alter=alt`,
  `semester=Bachelor`} $\Rightarrow$ {`Lernzeiten_woche_nachm_fer=1`} $\theta = \infty$
- {`alter=mittel`,
  `semester=Studienanfänger`} $\Rightarrow$ {`Lernzeiten_sams_nacht_fer=1`} $\theta = 5,79$
- {`alter=mittel`,
  `semester=Studienanfänger`} $\Rightarrow$ {`Lernzeiten_woche_nacht_fer=1`} $\theta = 4,10$
- {`alter=mittel`,
  `semester=Studienanfänger`} $\Rightarrow$ {`Lernzeiten_sams_abend_fer=1`} $\theta = 3,37$

Es ist sehr auffällig, dass Studierende, die gerade mit dem Studium beginnen und schon ein mittleres Alter (26–55) erreicht haben, eine sehr stark erhöhte Chance haben, allgemein nachts und samstagabends zu lernen. Diese Assoziationsregeln treten auch für die Lernzeiten im Semester auf, aber mit niedrigeren Bedeutungsmaßen. Es gilt für die 4 oben dargestellten Regeln, dass sie alle einen Support $\leq 0,5\%$ haben. Es liegt die Vermutung nahe, dass es sich hierbei um berufstätige Studierende, die zur Weiterbildung studieren, handelt.

Die vorherige Kombination von „Master“ und „GeiWi“ legt auch noch eine Analyse der Form Studienfach $\Rightarrow$ Lernzeiten in den Ferien nahe (siehe Anhang B.3.10):

- {studienfach=GeiWi} ⇒ {Lernzeiten_woche_nachm_fer=1} $\theta = 1,60$
- {studienfach=GeiWi} ⇒ {Lernzeiten_sams_vorm_fer=1} $\theta = 1,50$
- {studienfach=GeiWi} ⇒ {Lernzeiten_woche_vorm_fer=1} $\theta = 1,44$
- {studienfach=GeiWi} ⇒ {Lernzeiten_sonn_nachm_fer=1} $\theta = 1,31$
- {studienfach=GeiWi} ⇒ {Lernzeiten_sams_nachm_fer=1} $\theta = 1,31$
- {studienfach=GeiWi} ⇒ {Lernzeiten_sonn_vorm_fer=1} $\theta = 1,28$
- {studienfach=ReWiSo} ⇒ {Lernzeiten_sams_nie_fer=1} $\theta = 1,38$
- {studienfach=MaNa} ⇒ {Lernzeiten_sams_abend_fer=1} $\theta = 1,33$
- {studienfach=MaNa} ⇒ {Lernzeiten_woche_abend_fer=1} $\theta = 1,12$

Studierende der Mathematik und Naturwissenschaften haben wie bei den Lernzeiten im Semester eine erhöhte Chance, unter der Woche und Samstag abends zu lernen. Auffällig ist, dass Studierende der Geisteswissenschaften eine erhöhte Chance bei vielen Lernzeiten in den Ferien haben. Intuitiv sollte daher im nächsten Punkt 4.4 die Assoziationsregel gefunden werden, dass Geisteswissenschaftler ein hohes Lernvolumen in den Ferien haben.

4.4. Alter, Geschlecht, Semester, Studienfach, Lernvolumen

Wer lernt wann wieviel?

Einen gemeinsamen Zusammenhang zwischen Alter, Geschlecht, Semester, Studienfach (alle LHS) und Lernvolumen (RHS) zu finden, erweist sich als sehr unübersichtlich. Daher wird versucht, aus einzelnen Variablen das Lernvolumen zu bestimmten Zeiten vorauszusagen.
Wie im vorigen Abschnitt 4.3 erwähnt, sollten Geisteswissenschaftler in den Ferien ein hohes Lernvolumen aufweisen (vgl. Anhang B.4.1):

- {studienfach=GeiWi} ⇒ {Lernvolumen_svac=eher viel} $\theta = 1,61$
- {studienfach=GeiWi} ⇒ {Lernvolumen_send=eher wenig} $\theta = 1,33$
- {studienfach=GeiWi} ⇒ {Lernvolumen_send=wenig} $\theta = 1,26$
- {studienfach=GeiWi} ⇒ {Lernvolumen_svac=viel} $\theta = 1,12$
- {studienfach=MaNa} ⇒ {Lernvolumen_send=viel} $\theta = 1,54$
- {studienfach=MaNa} ⇒ {Lernvolumen_svac=eher wenig} $\theta = 1,18$
- {studienfach=MaNa} ⇒ {Lernvolumen_send=eher viel} $\theta = 1,15$

Geisteswissenschaftler lernen somit im Vergleich in den Ferien mehr, dafür am Semesterende weniger. Für Studierende der Mathematik und Naturwissenschaften ergibt sich genau das gegensätzliche Bild. Interessanterweise ergibt sich für das Semester als linke

Seite ein relativ ähnlicher Output (siehe Anhang B.4.2):

- {semester=Master} ⇒ {Lernvolumen_svac=viel} $\theta = 1,88$
- {semester=Master} ⇒ {Lernvolumen_send=wenig} $\theta = 1,57$
- {semester=Master} ⇒ {Lernvolumen_svac=eher wenig} $\theta = 1,44$
- {semester=Bachelor} ⇒ {Lernvolumen_send=viel} $\theta = 1,37$
- {semester=Bachelor} ⇒ {Lernvolumen_svac=wenig} $\theta = 1,36$

D. h., die Chance für Studierende im „Bachelor“ zu lernen, ist am Ende des Semesters erhöht, dafür aber erniedrigt in der vorlesungsfreien Zeit. Für Studierende im „Master“ gilt genau das Gegenteil. Fasst man die Analysen mit dem Studienfach und Semester in eine Analyse, erhält man die gleichen Resultate (siehe Anhang B.4.3).
Bei Betrachtung des Alters ergibt sich vor allem *ein* interessanter Zusammenhang (vgl. Anhang B.4.4):

- {alter=mittel} ⇒ {Lernvolumen_svac=viel} $\theta = 2,02$
- {alter=mittel} ⇒ {Lernvolumen_sbegi=viel} $\theta = 1,54$
- {alter=mittel} ⇒ {Lernvolumen_stime=viel} $\theta = 1,36$
- {alter=mittel} ⇒ {Lernvolumen_send=viel} $\theta = 1,04$

D. h., Studierende im „mittleren“ Alter haben immer eine erhöhte Chance auf ein sehr hohes Lernvolumen.
Für das Geschlecht ergeben sich keine interessanten Assoziationsregeln (siehe Anhang B.4.5). D. h., Männer und Frauen unterscheiden sich in ihrem Lernvolumen zu jeder Zeit des akademischen Jahres kaum.

4.5. Alter, Geschlecht, Semester, Studienfach, Nutzer, Anzahl Services, Zufriedenheits-Score

In diesem Punkt wird erst versucht aus personenbezogenen Daten vorherzusagen, ob Befragte Nutzer sind. Auf diesen Ergebnissen aufbauend soll schließlich eine Aussage zur Anzahl der benutzten Services getroffen werden. Letztendlich werden alle gerade genannten Variablen einbezogen, um einen Rückschluss auf den Zufriedenheits-Score zuzulassen.

4.5.1. Alter, Geschlecht, Semester, Studienfach, Nutzer

Wer nutzt die Universitätsbibliothek?

In diesem Abschnitt wird untersucht, ob sich aus den vier personenbezogenen Variablen `alter`, `sex`, `semester` und `studienfach` vorhersagen lässt, ob Befragte `nutzer` einer (oder mehrerer) Bibliotheken der LMU sind. Hier ist unbedingt zu beachten, dass die Stichprobe nicht repräsentativ ist, da Nutzer (im allgemeinen Sinne) stärker von der zugrunde liegenden Umfrage angesprochen wurden.
Der naheliegende Gedanke, dass Studierende, die schon länger studieren, Nutzer sind, zeigt sich auch in diesen Daten (siehe Anhang B.5.1):

- {semester=Studienanfänger} ⇒ {nutzer=Nichtnutzer} $\theta = 1,23$
- {semester=Bachelor} ⇒ {nutzer=Nichtnutzer} $\theta = 1,08$
- {semester=Master} ⇒ {nutzer=Nutzer} $\theta = 1,21$
- {semester=Langzeitstudent} ⇒ {nutzer=Nutzer} $\theta = 1,02$

Jedoch sind die resultierenden Bedeutungsmaße kaum vom Wert für statistische Unabhängigkeit verschieden.
Betrachtet man das Alter der Befragten, ergeben sich sich die Regeln (siehe Anhang B.5.2):

- {alter=alt} ⇒ {nutzer=Nutzer} $\theta = \infty$ $C = 1,00$
- {alter=mittel} ⇒ {nutzer=Nichtnutzer} $\theta = 1,17$ $C = 0,38$
- {alter=jung} ⇒ {nutzer=Nutzer} $\theta = 1,12$ $C = 0,66$

Hier sollte besonders auf die Confidence C der Assoziationsregeln eingegangen werden. So ist hier jeder „alte“ Studierende ein Nutzer der UB München. Man muss hier aber unbedingt beachten, dass nur 0,3% (bzw. in absoluten Zahlen: 5) der Befragten in die Kategorie „alt“ fallen (vgl. Anhang A.2), und es sich daher vermutlich um reinen Zufall handelt. Für Studierende im „jungen“ bzw. „mittleren“ Alter gilt, dass 66% bzw. 62% der Befragten Nutzer sind. Ein Schluss auf die Grundgesamtheit aller Studierenden sollte – besonders hier – auf Grund der angesprochenen Nicht-Repräsentativität der Stichprobe unbedingt unterlassen werden.
Am ehesten lässt sich vorhersagen, ob ein Befragter Nutzer ist, indem man das Studienfach betrachtet. Es ergeben sich die zwei interessanten Regeln (siehe Anhang B.5.3):

- {studienfach=GeiWi} ⇒ {nutzer=Nutzer} $\theta = 1,67$
- {studienfach=MaNa} ⇒ {nutzer=Nichtnutzer} $\theta = 1,64$

D. h., die Chance, dass ein Geisteswissenschaftler Nutzer ist, ist größer, während die Chance, dass ein Mathematiker bzw. Naturwissenschaftler Nutzer ist, kleiner ist (bzw. die Chance, dass ein Mathematiker bzw. Naturwissenschaftler Nichtnutzer ist, ist erhöht).
Für das Geschlecht ließen sich keine Regeln finden, die vom Wert für statistische Un-

abhängigkeit stark abweichen (siehe Anhang B.5.4).

4.5.2. Alter, Geschlecht, Semester, Studienfach, Nutzer, Anzahl Services

Wer nutzt wie viele Services?

In diesem Abschnitt wird die `anzahl_services` als Konsequenz von Assoziationsregeln betrachtet, wobei für die Voraussetzung alle im letzten Abschnitt 4.5.1 verwendeten Variablen zum Einsatz kommen.
Eine Vermutung ist, dass Nutzer der UB München auch eine Vielzahl an Services benutzen.

- {`nutzer=Nutzer`} ⇒ {`anzahl_services=viele`} $\theta = 1,71$
- {`nutzer=Nutzer`} ⇒ {`anzahl_services=eher viele`} $\theta = 1,29$
- {`nutzer=Nichtnutzer`} ⇒ {`anzahl_services=eher wenige`} $\theta = 1,34$

Die gewonnenen Regeln bestätigen dies (siehe Anhang B.6.1). Diese Fragestellung ließe sich auch gut mit einer Regression der Variable `nutzer` auf die (ungruppierte) Variable `anzahl_services` untersuchen.
Im vorigen Abschnitt 4.5.1 wurde ein Zusammenhang zwischen Studienfach und Nutzer festgestellt. Das gerade eben erzielte Ergebnis legt daher eine Untersuchung zwischen Studienfach und Anzahl der benutzten Services nahe (siehe Anhang B.6.2).

- {`studienfach=Med`} ⇒ {`anzahl_services=wenige`} $\theta = 2,86$
- {`studienfach=Med`} ⇒ {`anzahl_services=eher wenige`} $\theta = 1,47$
- {`studienfach=MaNa`} ⇒ {`anzahl_services=wenige`} $\theta = 2,13$
- {`studienfach=MaNa`} ⇒ {`anzahl_services=eher wenige`} $\theta = 1,77$
- {`studienfach=GeiWi`} ⇒ {`anzahl_services=viele`} $\theta = 2,07$
- {`studienfach=GeiWi`} ⇒ {`anzahl_services=eher viele`} $\theta = 1,50$

D. h., für Geisteswissenschaftler ist die Chance „eher viele“ oder sogar „viele“ Services zu benutzen erhöht, während die Assoziationsregeln für Studierende der Mathematik und Naturwissenschaften bzw. der Medizin und Gesundheitswissenschaften nahelegen, „eher wenige“ oder nur „wenige“ Services zu nutzen.
Untersucht man einen Zusammenhang von `nutzer` und `studienfach` mit `anzahl_services` simultan, erhält man ein ähnliches Ergebnis, das aber unübersichtlicher ist (siehe Anhang B.6.3).
Bei einer Analyse eines simultanen Zusammenhangs zwischen `alter` und `semester` mit `anzahl_services` ergibt sich (siehe B.6.4):

- {semester=Bachelor, alter=alt} $\Rightarrow$ {anzahl_services=eher wenige} $\theta = \infty$
- {semester=Studienanfänger} $\Rightarrow$ {anzahl_services=wenige} $\theta = 3,15$
- {semester=Studienanfänger, alter=jung} $\Rightarrow$ {anzahl_services=wenige} $\theta = 3,15$
- {alter=alt} $\Rightarrow$ {anzahl_services=eher wenige} $\theta = 2,96$
- {semester=Master} $\Rightarrow$ {anzahl_services=viele} $\theta = 2,32$
- {semester=Studienanfänger, alter=mittel} $\Rightarrow$ {anzahl_services=viele} $\theta = 2,27$
- {alter=jung} $\Rightarrow$ {anzahl_services=wenige} $\theta = 2,25$
- {semester=Langzeitstudent, alter=mittel} $\Rightarrow$ {anzahl_services=eher viele} $\theta = 2,25$
- {semester=Studienanfänger, alter=jung} $\Rightarrow$ {anzahl_services=eher wenige} $\theta = 2,21$
- {semester=Studienanfänger} $\Rightarrow$ {anzahl_services=eher wenige} $\theta = 2,17$
- {semester=Langzeitstudent} $\Rightarrow$ {anzahl_services=eher viele} $\theta = 2,10$
- {semester=Master, alter=jung} $\Rightarrow$ {anzahl_services=viele} $\theta = 2,09$
- {alter=mittel} $\Rightarrow$ {anzahl_services=viele} $\theta = 1,83$
- {semester=Bachelor, alter=mittel} $\Rightarrow$ {anzahl_services=viele} $\theta = 1,69$
- {semester=Master, alter=mittel} $\Rightarrow$ {anzahl_services=viele} $\theta = 1,67$
- {semester=Bachelor, alter=jung} $\Rightarrow$ {anzahl_services=wenige} $\theta = 1,54$
- {semester=Bachelor} $\Rightarrow$ {anzahl_services=wenige} $\theta = 1,49$
- {alter=jung} $\Rightarrow$ {anzahl_services=eher wenige} $\theta = 1,48$
- {alter=mittel} $\Rightarrow$ {anzahl_services=eher viele} $\theta = 1,46$
- {semester=Master, alter=mittel} $\Rightarrow$ {anzahl_services=eher viele} $\theta = 1,44$

Zusammenfassend kann man behaupten, dass Studierende in fortgeschrittenen Semestern, egal welchen Alters, eine höhere Chance haben „eher viele“ bis „viele“ Services zu nutzen. Für „Studienanfänger“ und „Bachelor“-Studenten unterscheidet sich das Bild je nach Alter der Studierenden; während junge (bzw. alte) Studierende in diesen Semestern eine erhöhte Chance besitzen, „eher wenige“ bis „wenige“ Services zu nutzen, haben „Studienfänger“ und „Bachelor“-Studenten im mittleren Alter eine erhöhte Chance,

„viele“ Services zu nutzen.

4.5.3. Alter, Geschlecht, Semester, Studienfach, Nutzer, Anzahl Services, Zufriedenheits-Score

Wer ist wie zufrieden?

In diesem Abschnitt wird als rechte Seite der Assoziationsregeln die künstlich erzeugte Variable `zscore` verwendet. Die Analyse mit allen aus dem vorherigen Punkt 4.5.2 verwendeten Variablen in der linken Seite liefert ein sehr einheitliches Bild (siehe Anhang B.7). Die ersten 16 Regeln mit dem höchsten Odds-Ratio zeigen in der Konsequenz nur die Ausprägung `zscore=Zufrieden`. Daher werden für die Voraussetzung der Regeln wieder nur Teilmengen betrachtet.
Betrachtet man als Teilmenge die Anzahl der benutzten Services und ob Studierende Nutzer sind, ergibt sich (siehe Anhang B.7.1):

- {`anzahl_services=viele`} ⇒ {`zscore=Zufrieden`} $\theta = 2,27$
- {`nutzer=Nutzer`, `anzahl_services=viele`} ⇒ {`zscore=Zufrieden`} $\theta = 2,19$
- {`nutzer=Nutzer`, `anzahl_services=wenige`} ⇒ {`zscore=Unzufrieden`} $\theta = 2,17$
- {`nutzer=Nichtnutzer`, `anzahl_services=viele`} ⇒ {`zscore=Zufrieden`} $\theta = 2,13$
- {`anzahl_services=wenige`} ⇒ {`zscore=Unzufrieden`} $\theta = 2,05$

Nutzer und Nichtnutzer, die viele Services nutzen, sind zufrieden. Nutzer, die hingegen nur wenige Services nutzen, sind unzufrieden. Dies legt die Vermutung nahe, dass nur die Anzahl der Services hier eine Rolle spielt und es egal für die Zufriedenheit ist, ob Studierende Nutzer sind oder nicht.
Daher wird erstmal nur die Anzahl der Services weiter betrachtet. Damit erhält man die Regeln (siehe Anhang B.7.2):

- {`anzahl_services=viele`} ⇒ {`zscore=Zufrieden`} $\theta = 2,27$
- {`anzahl_services=eher viele`} ⇒ {`zscore=Zufrieden`} $\theta = 1,34$
- {`anzahl_services=eher wenige`} ⇒ {`zscore=Neutral`} $\theta = 1,24$
- {`anzahl_services=eher wenige`} ⇒ {`zscore=Unzufrieden`} $\theta = 1,48$
- {`anzahl_services=wenige`} ⇒ {`zscore=Neutral`} $\theta = 1,75$
- {`anzahl_services=wenige`} ⇒ {`zscore=Unzufrieden`} $\theta = 2,05$

D. h., Studierende die „eher viele“ oder sogar „viele“ Services nutzen, haben eine erhöhte Chance, zufrieden zu sein. Studierende, die hingegen „eher wenige“ bis „wenige“ Services nutzen, haben eine erhöhte Chance, einen niedrigeren Zufriedenheitsscore zu haben.
Wird nun zudem noch das Alter in die Voraussetzung mit aufgenommen, ergeben sich die Regeln (siehe Anhang B.7.3):

- {`alter=jung`, `anzahl_services=viele`} ⇒ {`zscore=Zufrieden`} $\theta = 2,47$
- {`alter=jung`, `anzahl_services=wenige`} ⇒ {`zscore=Unzufrieden`} $\theta = 2,17$
- {`alter=jung`, `anzahl_services=wenige`} ⇒ {`zscore=Neutral`} $\theta = 1,88$
- {`alter=jung`, `anzahl_services=eher wenige`} ⇒ {`zscore=Unzufrieden`} $\theta = 1,46$
- {`alter=mittel`, `anzahl_services=viele`} ⇒ {`zscore=Zufrieden`} $\theta = 1,72$
- {`alter=mittel`, `anzahl_services=wenige`} ⇒ {`zscore=Zufrieden`} $\theta = 1,69$
- {`alter=mittel`, `anzahl_services=eher viele`} ⇒ {`zscore=Zufrieden`} $\theta = 1,68$

Dies bedeutet Studierende im „mittleren“ Alter haben, unabhängig von ihrer Anzahl an benutzten Services, eine höhere Chance auf einen hohen Zufriedenheitsscore. Für „junge“ Studierende ergibt sich das gleiche Bild wie im oberen Fall, als man nur die Anzahl an Services betrachtet hat.
Für die Untersuchung, ob die Studienfächer unterschiedlich zufrieden sind, wird in der linken Seite nur das Studienfach betrachtet. Man erhält die Regeln (siehe Anhang B.7.4):

- {`studienfach=Med`} ⇒ {`zscore=Unzufrieden`} $\theta = 1,87$
- {`studienfach=Med`} ⇒ {`zscore=Neutral`} $\theta = 1,45$
- {`studienfach=MaNa`} ⇒ {`zscore=Neutral`} $\theta = 1,51$
- {`studienfach=ReWiSo`} ⇒ {`zscore=Neutral`} $\theta = 1,28$
- {`studienfach=ReWiSo`} ⇒ {`zscore=Unzufrieden`} $\theta = 1,11$
- {`studienfach=GeiWi`} ⇒ {`zscore=Zufrieden`} $\theta = 1,84$

D. h., nur die Geisteswissenschaftler haben eine erhöhte Chance auf einen hohen Zufriedenheitsscore. Alle anderen Fächer, insbesondere die Medizin und Gesundheitswissenschaften, haben eine erhöhte Chance auf einen niedrigen Zufriedenheitsscore.

5. Schluss

5.1. Zusammenfassung

Das erste zu treffende Fazit ist, dass sich Assoziationsregeln nur bedingt zur Analyse dieses Datensatzes eignen. Die hier betrachteten Daten sind zu dicht, d. h., jede Transaktion (sprich: Beobachtung) enthält zu viele Items. Zudem enthält der Datensatz für die Analyse mit Assozationsregeln vergleichsweise wenig Beobachtungen.
Trotzdem können als wichtige und interessante Ergebnisse festgehalten werden:

- Jedes Studienfach bevorzugt andere Lernorte.
- Studienfächer unterscheiden sich in den Lernzeiten (Semester und Ferien).
- Frauen lernen bevorzugt tagsüber, Männer abends bzw. nachts.
- Jedes Studienfach lernt zu einer anderen Phase des akademischen Jahres.
- Geisteswissenschaftler sind verstärkt Nutzer der UB, Naturwissenschaftler nicht.
- Das Studienfach beeinflusst die Anzahl der benutzten Services.
- Mit der Anzahl der Services variiert der Zufriedenheitsscore.
- Nur Geisteswissenschaftler haben eine erhöhte Chance auf einen hohen Zufriedenheitsscore.

5.2. Ausblick

Man könnte Assoziationsregeln zwischen den einzelnen Services suchen. Dies würde am ehesten auch der klassischen Warenkorbanalyse entsprechen. Problematisch ist hierbei aber, dass Studierende angeben sollten, welche Services sie schon jemals benutzt haben. Somit hat man keine Information, ob die Services „gleichzeitig“ zusammen genutzt wurden oder mit großem zeitlichen Abstand und vollkommen unabhängig voneinander. Dies kann zu falschen Schlüssen bei der Interpretation der Ergebnisse führen.

Eine Erweiterung des **arules**-Packages zur Gewinnung von Assoziationsregeln mit mehrelementigen Konsequenzen könnte weitere interessante Regeln hervorbringen.

Desweiteren könnte man sich noch eine Vielzahl an weiteren Bedeutungsmaßen anschauen und die Ergebnisse der unterschiedlichen Maße vergleichen. Welche Maße suggerieren gleiche Ergebnisse und welche unterschiedliche? Dies ist mehr eine theoretische Fragestellung (siehe hierzu Tan u. a. 2004).

Literaturverzeichnis

Agresti 2002

Agresti, A.: *Categorical Data Analysis.* Zweite Edition. Wiley-Interscience, 2002 (Wiley Series in Probability and Statistics)

Bates u. Maechler 2011

Bates, Douglas ; Maechler, Martin: *Matrix: Sparse and Dense Matrix Classes and Methods*, 2011. http://CRAN.R-project.org/package=Matrix. – R package version 0.999375-50

Borgelt 1997

Borgelt, Christian: Einführung in Datenanalyse und Data Mining mit intelligenten Technologien. In: *Seminar zu Anwendungen in Datenanalyse und Data Mining (Bergholz-Rehbrücke, Germany).* Aachen, Germany : MIT GmbH, 1997

Eifler u. a. 2011

Eifler, Fabian ; Pietsch, Robert ; Schierholz, Malte: *Bericht zum Statistischen Praktikum: Studierendenbefragung der Universitätsbibliothek*. http://www.cip.ifi.lmu.de/~pietsch/bericht.pdf. Version: 2011

Gamil 2010

Gamil, Omar: *Apriori Algorithm*. Version: 25.04.2010, 2010. http://www.codeproject.com/KB/recipes/AprioriAlgorithm.aspx, Abruf: 13.07.2011

Geng u. Hamilton 2006

Geng, Liqiang ; Hamilton, Howard J.: Interestingness Measures for Data Mining: A Survey. In: *ACM Comput. Surv.* 38 (2006), September. http://dx.doi.org/10.1145/1132960.1132963. – DOI 10.1145/1132960.1132963. – ISSN 0360–0300

Hahsler u. a. 2011

Hahsler, Michael ; Buchta, Christian ; Grün, Bettina ; Hornik, Kurt: *arules: Mining Association Rules and Frequent Itemsets*, 2011. http://CRAN.R-project.org/. – R package version 1.0-6.

Hahsler u. a. 2005

Hahsler, Michael ; Grün, Bettina ; Hornik, Kurt: arules – A Computational Environment for Mining Association Rules and Frequent Item Sets. In: *Journal of Statistical Software* 14 (2005), October, Nr. 15, 1–25. http://www.jstatsoft.org/v14/i15/. – ISSN 1548–7660

Hahsler u. Hornik 2007

Hahsler, Michael ; Hornik, Kurt: New Probabilistic Interest Measures for Association Rules. In: *Intelligent Data Analysis* 11 (2007), Nr. 5, 437–455. http://iospress.metapress.com/openurl.asp?genre=article&issn=1088-467X&volume=11&issue=5&spage=437. – ISSN 1088–467X

Hastie u. a. 2001

Hastie, Trevor ; Tibshirani, Robert ; Friedman, Jerome: *The Elements of Statistical Learning (Data Mining, Inference and Prediction)*. Zweite Edition. Springer Verlag, 2001 http://www-stat.stanford.edu/~tibs/ElemStatLearn/

Naisbitt 1982

Naisbitt, John: *Megatrends*. Warner Books, 1982

Piatetsky-Shapiro 1991

Piatetsky-Shapiro, Gregory: Discovery, Analysis, and Presentation of Strong Rules. In: *Knowledge Discovery in Databases*. AAAI/MIT Press, 1991. – ISBN 0–262–62080–4, S. 229–248

R Development Core Team 2011

R Development Core Team: *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2011. http://www.R-project.org/. – ISBN 3-900051-07-0

Sarkar 2008

Sarkar, Deepayan: *Lattice: Multivariate Data Visualization with R*. New York : Springer, 2008 http://lmdvr.r-forge.r-project.org. – ISBN 978-0-387-75968-5

Tan u. a. 2004

Tan, Pang-Ning ; Kumar, Vipin ; Srivastava, Jaideep: Selecting the Right Objective Measure for Association Analysis. In: *Inf. Syst.* 29 (2004), June, 293–313. http://dx.doi.org/10.1016/S0306-4379(03)00072-3. – DOI 10.1016/S0306–4379(03)00072–3. – ISSN 0306–4379

Tan u. a. 2005

TAN, Pang-Ning ; STEINBACH, Michael ; KUMAR, Vipin: *Introduction to Data Mining, (First Edition)*. Boston, MA, USA : Addison-Wesley Longman Publishing Co., Inc., 2005. – ISBN 0321321367

Wikipedia 2011

WIKIPEDIA: *Association Rule Learning*. Version: 02.07.2011, 2011. http://en.wikipedia.org/wiki/Association_rule_learning, Abruf: 13.07.2011

A. Verwendete Variablen

A.1. Tabelle der verwendeten Variablen

Variablenname	basiert auf Frage	Ausprägungen	Anmerkungen
`alter`	Alter	• jung (10–25 Jahre) • mittel (26–55 Jahre) • alt (56–90 Jahre)	Gruppierung von ursprünglich metrischer Variable
`semester`	Hochschulsemester	• Studienanfänger (1–2 Semester) • Bachelor (3–7 Semester) • Master (8–12 Semester) • Langzeitstudent (13–90 Semester)	Gruppierung von ursprünglich metrischer Variable
`sex`	Geschlecht	• weiblich • männlich	

Fortsetzung nächste Seite

Variablenname	basiert auf Frage	Ausprägungen	Anmerkungen
`studienfach`	1. Studienfach	• MaNa (Mathematik und Naturwissenschaften) • GeiWi (Geisteswissenschaften) • ReWiSo (Rechts-, Wirtschaft- und Sozialwissenschaften) • Med (Medizin und Gesundheitswissenschaften)	Gruppierung von Variable `fach1` (siehe Eifler u. a. 2011)
`Lernorte_FachB`	Wo lernen Sie? Fachbibliothek	• fast immer • häufig • selten • nie	
`Lernorte_Fak`	Wo lernen Sie? Fakultät/Institut (ohne Bibliothek)	• fast immer • häufig • selten • nie	
`Lernorte_UBS`	Wo lernen Sie? Universitätsbibliothek/ Studentenbibliothek	• fast immer • häufig • selten • nie	

Fortsetzung nächste Seite

Variablenname	basiert auf Frage	Ausprägungen	Anmerkungen
`Lernorte_UBZ`	Wo lernen Sie? Universitätsbibliothek/ Zentrale	• fast immer • häufig • selten • nie	
`Lernzeiten`	Zu welchen Zeiten lernen Sie?		Mehrfachnennungen möglich
`_woche_vorm_sem`	unter der Woche vormittags (während des Semesters)	• ja (1) • nicht gewählt (0)	8–12 Uhr
`_woche_nachm_sem`	unter der Woche nachmittags (während des Semesters)	• ja (1) • nicht gewählt (0)	12–18 Uhr
`_woche_abend_sem`	unter der Woche abends (während des Semesters)	• ja (1) • nicht gewählt (0)	18–22 Uhr
`_woche_nacht_sem`	unter der Woche nachts (während des Semesters)	• ja (1) • nicht gewählt (0)	22–8 Uhr
`_woche_nie_sem`	unter der Woche nie (während des Semesters)	• ja (1) • nicht gewählt (0)	
`_sams_vorm_sem`	Samstag vormittags (während des Semesters)	• ja (1) • nicht gewählt (0)	8–12 Uhr
`_sams_nachm_sem`	Samstag nachmittags (während des Semesters)	• ja (1) • nicht gewählt (0)	12–18 Uhr
`_sams_abend_sem`	Samstag abends (während des Semesters)	• ja (1) • nicht gewählt (0)	18–22 Uhr
`_sams_nacht_sem`	Samstag nachts (während des Semesters)	• ja (1) • nicht gewählt (0)	22–8 Uhr

Fortsetzung nächste Seite

Variablenname	basiert auf Frage	Ausprägungen	Anmerkungen
`_sams_nie_sem`	Samstag nie (während des Semesters)	• ja (1) • nicht gewählt (0)	
`_sonn_vorm_sem`	Sonntag vormittags (während des Semesters)	• ja (1) • nicht gewählt (0)	8–12 Uhr
`_sonn_nachm_sem`	Sonntag nachmittags (während des Semesters)	• ja (1) • nicht gewählt (0)	12–18 Uhr
`_sonn_abend_sem`	Sonntag abends (während des Semesters)	• ja (1) • nicht gewählt (0)	18–22 Uhr
`_sonn_nacht_sem`	Sonntag nachts (während des Semesters)	• ja (1) • nicht gewählt (0)	22–8 Uhr
`_sonn_nie_sem`	Sonntag nie (während des Semesters)	• ja (1) • nicht gewählt (0)	
`_woche_vorm_fer`	unter der Woche vormittags (während der Ferien)	• ja (1) • nicht gewählt (0)	8–12 Uhr
`_woche_nachm_fer`	unter der Woche nachmittags (während der Ferien)	• ja (1) • nicht gewählt (0)	12–18 Uhr
`_woche_abend_fer`	unter der Woche abends (während der Ferien)	• ja (1) • nicht gewählt (0)	18–22 Uhr
`_woche_nacht_fer`	unter der Woche nachts (während der Ferien)	• ja (1) • nicht gewählt (0)	22–8 Uhr
`_woche_nie_fer`	unter der Woche nie (während der Ferien)	• ja (1) • nicht gewählt (0)	

Fortsetzung nächste Seite

Variablenname	basiert auf Frage	Ausprägungen	Anmerkungen
`_sams_vorm_fer`	Samstag vormittags (während der Ferien)	• ja (1) • nicht gewählt (0)	8–12 Uhr
`_sams_nachm_fer`	Samstag nachmittags (während der Ferien)	• ja (1) • nicht gewählt (0)	12–18 Uhr
`_sams_abend_fer`	Samstag abends (während der Ferien)	• ja (1) • nicht gewählt (0)	18–22 Uhr
`_sams_nacht_fer`	Samstag nachts (während der Ferien)	• ja (1) • nicht gewählt (0)	22–8 Uhr
`_sams_nie_fer`	Samstag nie (während der Ferien)	• ja (1) • nicht gewählt (0)	
`_sonn_vorm_fer`	Sonntag vormittags (während der Ferien)	• ja (1) • nicht gewählt (0)	8–12 Uhr
`_sonn_nachm_fer`	Sonntag nachmittags (während der Ferien)	• ja (1) • nicht gewählt (0)	12–18 Uhr
`_sonn_abend_fer`	Sonntag abends (während der Ferien)	• ja (1) • nicht gewählt (0)	18–22 Uhr
`_sonn_nacht_fer`	Sonntag nachts (während der Ferien)	• ja (1) • nicht gewählt (0)	22–8 Uhr
`_sonn_nie_fer`	Sonntag nie (während der Ferien)	• ja (1) • nicht gewählt (0)	

Fortsetzung nächste Seite

Variablenname	**basiert auf Frage**	**Ausprägungen**	**Anmerkungen**
`Lernvolumen_sbegi`	Wie viele Stunden lernen Sie durchschnittlich pro Woche? Semesterbeginn	• wenig (0–5 Stunden) • eher wenig (6–10 Stunden) • eher viel (11–20 Stunden) • viel (20–80 Stunden)	Gruppierung von ursprünglich metrischer Variable basierend auf den Quantilen
`Lernvolumen_stime`	Wie viele Stunden lernen Sie durchschnittlich pro Woche? Semesterverlauf	• wenig (0–8 Stunden) • eher wenig (9–15 Stunden) • eher viel (16–23 Stunden) • viel (24–80 Stunden)	Gruppierung von ursprünglich metrischer Variable basierend auf den Quantilen
`Lernvolumen_send`	Wie viele Stunden lernen Sie durchschnittlich pro Woche? Semesterende	• wenig (0–16 Stunden) • eher wenig (17–25 Stunden) • eher viel (26–39 Stunden) • viel (40–80 Stunden)	Gruppierung von ursprünglich metrischer Variable basierend auf den Quantilen
`Lernvolumen_svac`	Wie viele Stunden lernen Sie durchschnittlich pro Woche? Semesterferien	• wenig (0–2 Stunden) • eher wenig (3–8 Stunden) • eher viel (9–20 Stunden) • viel (21–80 Stunden)	Gruppierung von ursprünglich metrischer Variable basierend auf den Quantilen

Fortsetzung nächste Seite

Variablenname	basiert auf Frage	Ausprägungen	Anmerkungen
`nutzer`	Wo lernen Sie?	• Nutzer • Nichtnutzer	siehe Abschnitt 3.1
`anzahl_services`	Welche Services haben Sie schon einmal genutzt?	• wenige (0–3 Services) • eher wenige (4–7 Services) • eher viele (8–12 Services) • viele (13–20 Services)	siehe Abschnitt 3.1
`zscore`		• Unzufrieden (−20–−1) • Neutral (0–5) • Zufrieden (6–20)	siehe Abschnitt 3.1

Tabelle A.1.: In Kapitel 4 verwendete Variablen

A.2. Zusammenfassung der verwendeten Variablen

```
 alter             semester             sex               studienfach
 jung  :1181       Studienanfänger:190  männlich: 502     GeiWi :852
 mittel: 388       Bachelor       :780  weiblich:1068     MaNa  :265
 alt   :   5       Master         :482  NA's    :  26     Med   :153
 NA's  :  22       Langzeitstudent:144                    ReWiSo:326

 Lernorte_FachB    Lernorte_Fak         Lernorte_UBS      Lernorte_UBZ
 fast immer:321    fast immer: 31       fast immer: 49    fast immer: 61
 häufig    :629    häufig    :303       häufig    :193    häufig    :161
 selten    :484    selten    :796       selten    :624    selten    :655
 nie       :162    nie       :466       nie       :730    nie       :719

Lernzeiten
 _woche_vorm_sem   _woche_nachm_sem     _woche_abend_sem  _woche_nacht_sem  _woche_nie_sem
 0:879             0: 565               0: 391            0:1159            0:1588
 1:717             1:1031               1:1205            1: 437            1:   8

 _sams_vorm_sem    _sams_nachm_sem      _sams_abend_sem   _sams_nacht_sem   _sams_nie_sem
 0:911             0: 375               0:905             0:1324            0:1509
 1:685             1:1221               1:691             1: 272            1:  87

 _sonn_vorm_sem    _sonn_nachm_sem      _sonn_abend_sem   _sonn_nacht_sem   _sonn_nie_sem
 0:1033            0: 457               0:860             0:1343            0:1453
 1: 563            1:1139               1:736             1: 253            1: 143
```

```
_woche_vorm_fer     _woche_nachm_fer    _woche_abend_fer    _woche_nacht_fer    _woche_nie_fer
0:913               0: 527              0:939               0:1356              0:1432
1:683               1:1069              1:657               1: 240              1: 164

_sams_vorm_fer      _sams_nachm_fer     _sams_abend_fer     _sams_nacht_fer     _sams_nie_fer
0:1169              0:737               0:1124              0:1417              0:1192
1: 427              1:859               1: 472              1: 179              1: 404

_sonn_vorm_fer      _sonn_nachm_fer     _sonn_abend_fer     _sonn_nacht_fer     _sonn_nie_fer
0:1266              0:852               0:1147              0:1422              0:1080
1: 330              1:744               1: 449              1: 174              1: 516

Lernvolumen_sbegi   Lernvolumen_stime   Lernvolumen_send    Lernvolumen_svac
wenig     :392      wenig     :338      wenig     :388      wenig     :316
eher wenig:330      eher wenig:400      eher wenig:350      eher wenig:459
eher viel :435      eher viel :458      eher viel :456      eher viel :384
viel      :439      viel      :400      viel      :402      viel      :437

nutzer              anzahl_services     zscore
Nichtnutzer: 561    wenige     :154     Unzufrieden:182
Nutzer     :1035    eher wenige:538     Neutral    :620
                    eher viele :698     Zufrieden  :794
                    viele      :206
```

B. R-Code und Output

R-Code und Output zum Kapitel 4. Anwendung und Ergebnisse.

Verwendete Software

- R 2.13.0 (R Development Core Team 2011)

mit Paketen

- **arules** Version 1.0-6 (Hahsler u. a. 2011, 2005)
- **Matrix** Version 0.999375-50 (Bates u. Maechler 2011)
- **lattice** Version 0.19-23 (Sarkar 2008)

B.1. Geschlecht, Studienfach, Lernort

```
> StudSexOrte <- apriori(data[c("Lernorte_Fak","Lernorte_FachB","Lernorte_UBZ","Lernorte_UBS",
+ "studienfach", "sex")], parameter = list(supp = 0.01, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(lhs = c(paste("studienfach=",levels(data$studienfach)[1:4],sep=""),
+ paste("sex=",levels(data$sex)[1:2],sep="")),
+ rhs = c(paste("Lernorte_Fak=",levels(data$Lernorte_Fak)[1:2],sep=""),
+ paste("Lernorte_FachB=",levels(data$Lernorte_FachB)[1:2],sep=""),
+ paste("Lernorte_UBZ=",levels(data$Lernorte_UBZ)[1:2],sep=""),
+ paste("Lernorte_UBS=",levels(data$Lernorte_UBS)[1:2],sep="")), default = "none"))
> quality(StudSexOrte) <- cbind(quality(StudSexOrte),
+ interestMeasure(StudSexOrte, c("oddsRatio", "phi"), data))
> inspect(head(sort(StudSexOrte, by="oddsRatio"),13))

  lhs                       rhs                          support confidence     lift oddsRatio        phi
1 {studienfach=MaNa,
   sex=weiblich}         => {Lernorte_Fak=häufig}   0.02882206  0.3382353 1.781596  2.392477 0.11547748
2 {studienfach=MaNa}     => {Lernorte_Fak=häufig}   0.05200501  0.3132075 1.649766  2.303022 0.14035028
3 {studienfach=MaNa,
   sex=weiblich}         => {Lernorte_UBS=häufig}   0.01629073  0.1911765 1.580920  1.830049 0.06575965
4 {studienfach=MaNa,
   sex=männlich}         => {Lernorte_Fak=häufig}   0.02255639  0.2857143 1.504950  1.802247 0.07156447
5 {studienfach=GeiWi}    => {Lernorte_FachB=häufig} 0.24060150  0.4507042 1.143599  1.671167 0.12393605
6 {studienfach=ReWiSo}   => {Lernorte_UBZ=häufig}   0.02756892  0.1349693 1.337957  1.537613 0.05735286
7 {studienfach=GeiWi,
   sex=männlich}         => {Lernorte_UBZ=häufig}   0.01629073  0.1340206 1.328552  1.452469 0.04093721
```

```
8  {studienfach=GeiWi,
    sex=weiblich}       => {Lernorte_FachB=häufig} 0.17919799  0.4454829 1.130351  1.431077 0.08624206
9  {studienfach=ReWiSo} => {Lernorte_UBS=häufig}   0.03132832  0.1533742 1.268317  1.427739 0.05042034
10 {studienfach=ReWiSo,
    sex=weiblich}       => {Lernorte_UBS=häufig}   0.01817043  0.1576087 1.303334  1.423761 0.04061277
11 {studienfach=MaNa}   => {Lernorte_UBS=häufig}   0.02568922  0.1547170 1.279421  1.419731 0.04624270
12 {studienfach=ReWiSo,
    sex=weiblich}       => {Lernorte_UBZ=häufig}   0.01503759  0.1304348 1.293006  1.395985 0.03542865
13 {studienfach=GeiWi,
    sex=männlich}       => {Lernorte_FachB=häufig} 0.05639098  0.4639175 1.177126  1.385579 0.05314002
```

B.1.1. Studienfach und Lernort

```
> StudOrte <- apriori(data[c("Lernorte_Fak","Lernorte_FachB","Lernorte_UBZ","Lernorte_UBS", "studienfach")],
+ parameter = list(supp = 0.01, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(lhs = c(paste("studienfach=",levels(data$studienfach)[1:4],sep="")),
+ rhs = c(paste("Lernorte_Fak=",levels(data$Lernorte_Fak)[1:2],sep=""),
+ paste("Lernorte_FachB=",levels(data$Lernorte_FachB)[1:2],sep=""),
+ paste("Lernorte_UBZ=",levels(data$Lernorte_UBZ)[1:2],sep=""),
+ paste("Lernorte_UBS=",levels(data$Lernorte_UBS)[1:2],sep="")), default = "none"))
> quality(StudOrte) <- cbind(quality(StudOrte), interestMeasure(StudOrte, c("oddsRatio", "phi"), data))
> inspect(head(sort(StudOrte, by="oddsRatio"),6))

  lhs                       rhs                            support confidence     lift oddsRatio        phi
1 {studienfach=MaNa}    => {Lernorte_Fak=häufig}        0.05200501  0.3132075 1.649766  2.303022 0.14035028
2 {studienfach=GeiWi}   => {Lernorte_FachB=häufig}      0.24060150  0.4507042 1.143599  1.671167 0.12393605
3 {studienfach=ReWiSo} => {Lernorte_UBZ=häufig}        0.02756892  0.1349693 1.337957  1.537613 0.05735286
4 {studienfach=ReWiSo} => {Lernorte_UBS=häufig}        0.03132832  0.1533742 1.268317  1.427739 0.05042034
5 {studienfach=MaNa}    => {Lernorte_UBS=häufig}        0.02568922  0.1547170 1.279421  1.419731 0.04624270
6 {studienfach=ReWiSo} => {Lernorte_FachB=fast immer} 0.04949875  0.2423313 1.204862  1.358651 0.05207940
```

B.1.2. Geschlecht und Lernort

```
> SexOrte <- apriori(data[c("Lernorte_Fak","Lernorte_FachB","Lernorte_UBZ","Lernorte_UBS", "sex")],
+ parameter = list(supp = 0.01, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(lhs = c(paste("sex=",levels(data$sex)[1:2],sep="")),
+ rhs = c(paste("Lernorte_Fak=",levels(data$Lernorte_Fak)[1:2],sep=""),
+ paste("Lernorte_FachB=",levels(data$Lernorte_FachB)[1:2],sep=""),
+ paste("Lernorte_UBZ=",levels(data$Lernorte_UBZ)[1:2],sep=""),
+ paste("Lernorte_UBS=",levels(data$Lernorte_UBS)[1:2],sep="")), default = "none"))
> quality(SexOrte) <- cbind(quality(SexOrte), interestMeasure(SexOrte, c("oddsRatio", "phi"), data))
> inspect(head(sort(SexOrte, by="oddsRatio"),3))

  lhs                rhs                           support confidence     lift oddsRatio        phi
1 {sex=männlich} => {Lernorte_FachB=fast immer} 0.07393484  0.2350598 1.168708  1.348753 0.05734257
2 {sex=männlich} => {Lernorte_Fak=häufig}       0.06390977  0.2031873 1.070254  1.132910 0.02303743
3 {sex=weiblich} => {Lernorte_FachB=häufig}     0.26879699  0.4016854 1.019221  1.101033 0.02204715
```

B.2. Geschlecht, Studienfach

```
> StudSex <- apriori(data[c("sex", "studienfach")],
+ parameter = list(supp = 0.05, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste("studienfach=",levels(data$studienfach)[1:4],sep="")), default = "lhs"))
> quality(StudSex) <- cbind(quality(StudSex), interestMeasure(StudSex, c("oddsRatio", "phi"), data))
> inspect(head(sort(StudSex, by="oddsRatio"),4))

  lhs                rhs                        support confidence     lift oddsRatio        phi
1 {sex=männlich} => {studienfach=MaNa}   0.07894737 0.25099602 1.511659  2.302350 0.15465278
2 {sex=weiblich} => {studienfach=GeiWi}  0.40225564 0.60112360 1.126048  2.282093 0.19184011
3 {sex=männlich} => {studienfach=ReWiSo} 0.08458647 0.26892430 1.316574  1.739090 0.10864898
4 {sex=weiblich} => {studienfach=Med}    0.06641604 0.09925094 1.035323  1.127660 0.01635855
```

B.3. Alter, Geschlecht, Semester, Studienfach, Lernzeiten

B.3.1. Studienfach und Lernzeiten im Semester

```
> StudLernZeit_sem <- apriori(data[c(names(LZsem), "studienfach")],
+ parameter = list(supp = 0.05, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste(names(LZsem)[1:12],"=1",sep="")),
+ lhs = c(paste("studienfach=",levels(data$studienfach)[1:4],sep="")), default = "none"))
> quality(StudLernZeit_sem) <- cbind(quality(StudLernZeit_sem),
+ interestMeasure(StudLernZeit_sem, c("oddsRatio", "phi"), data))
> inspect(head(sort(StudLernZeit_sem, by="oddsRatio"),15))

   lhs                          rhs                               support confidence     lift oddsRatio        phi
1  {studienfach=Med}    => {Lernzeiten_woche_abend_sem=1} 0.08709273  0.9084967 1.203287  3.511324 0.11620564
2  {studienfach=Med}    => {Lernzeiten_sonn_vorm_sem=1}   0.05200501  0.5424837 1.537840  2.378839 0.12929141
3  {studienfach=Med}    => {Lernzeiten_sams_abend_sem=1}  0.05576441  0.5816993 1.343549  1.942717 0.09774974
4  {studienfach=Med}    => {Lernzeiten_sams_vorm_sem=1}   0.05388471  0.5620915 1.309632  1.808586 0.08742683
5  {studienfach=MaNa}   => {Lernzeiten_sams_abend_sem=1}  0.08521303  0.5132075 1.185353  1.474069 0.07226855
6  {studienfach=Med}    => {Lernzeiten_sonn_nachm_sem=1}  0.07456140  0.7777778 1.089845  1.451471 0.04618593
7  {studienfach=MaNa}   => {Lernzeiten_woche_abend_sem=1} 0.13408521  0.8075472 1.069581  1.439623 0.05450428
8  {studienfach=GeiWi}  => {Lernzeiten_woche_vorm_sem=1}  0.25877193  0.4847418 1.079007  1.361647 0.07635949
9  {studienfach=Med}    => {Lernzeiten_sams_nachm_sem=1}  0.07769424  0.8104575 1.059370  1.348631 0.03488336
10 {studienfach=GeiWi}  => {Lernzeiten_woche_nachm_sem=1} 0.35839599  0.6713615 1.039275  1.268441 0.05677528
11 {studienfach=MaNa}   => {Lernzeiten_woche_nacht_sem=1} 0.05200501  0.3132075 1.143888  1.258630 0.03942383
12 {studienfach=ReWiSo} => {Lernzeiten_woche_nachm_sem=1} 0.13972431  0.6840491 1.058916  1.237936 0.04032226
13 {studienfach=MaNa}   => {Lernzeiten_sonn_nachm_sem=1}  0.12406015  0.7471698 1.046956  1.224801 0.03307729
```

```
14 {studienfach=MaNa}   => {Lernzeiten_sams_nachm_sem=1}  0.12969925  0.7811321 1.021038  1.115742 0.01693832
15 {studienfach=ReWiSo} => {Lernzeiten_woche_vorm_sem=1}  0.09586466  0.4693252 1.044690  1.107059 0.02044962
```

B.3.2. Geschlecht und Lernzeiten im Semester

```
> SexLernZeit_sem <- apriori(data[c(names(LZsem), "sex")],
+ parameter = list(supp = 0.05, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste(names(LZsem)[1:12],"=1",sep="")),
+ lhs = c(paste("sex=",levels(data$sex)[1:2],sep="")), default = "none"))
> quality(SexLernZeit_sem) <- cbind(quality(SexLernZeit_sem),
+ interestMeasure(SexLernZeit_sem, c("oddsRatio", "phi"), data))
> inspect(head(sort(SexLernZeit_sem, by="oddsRatio"),7))

  lhs               rhs                               support confidence     lift oddsRatio        phi
1 {sex=weiblich} => {Lernzeiten_sonn_vorm_sem=1}    0.25501253  0.3810861 1.080308  1.468288 0.08432033
2 {sex=männlich} => {Lernzeiten_woche_nacht_sem=1} 0.10275689  0.3266932 1.193140  1.459176 0.08033698
3 {sex=weiblich} => {Lernzeiten_sams_vorm_sem=1}    0.30200501  0.4513109 1.051521  1.316851 0.06353926
4 {sex=männlich} => {Lernzeiten_sams_nacht_sem=1}  0.06203008  0.1972112 1.157165  1.307807 0.04825483
5 {sex=männlich} => {Lernzeiten_sams_abend_sem=1}  0.14786967  0.4701195 1.085833  1.246005 0.05080581
6 {sex=männlich} => {Lernzeiten_woche_nachm_sem=1} 0.21365915  0.6792829 1.051538  1.240112 0.04716003
7 {sex=weiblich} => {Lernzeiten_woche_vorm_sem=1}  0.31015038  0.4634831 1.031686  1.190746 0.04070107
```

B.3.3. Geschlecht und Lernzeiten in den Ferien

```
> SexLernZeit_fer <- apriori(data[c(names(LZfer), "sex")],
+ parameter = list(supp = 0.05, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste(names(LZfer)[1:12],"=1",sep="")),
+ lhs = c(paste("sex=",levels(data$sex)[1:2],sep="")), default = "none"))
> quality(SexLernZeit_fer) <- cbind(quality(SexLernZeit_fer),
+ interestMeasure(SexLernZeit_fer, c("oddsRatio", "phi"), data))
> inspect(head(sort(SexLernZeit_fer, by="oddsRatio"),6))

  lhs               rhs                               support confidence     lift oddsRatio        phi
1 {sex=weiblich} => {Lernzeiten_woche_vorm_fer=1}  0.30576441  0.4569288 1.067728  1.436817 0.08331319
2 {sex=männlich} => {Lernzeiten_woche_nacht_fer=1} 0.05701754  0.1812749 1.205478  1.404252 0.05855776
3 {sex=männlich} => {Lernzeiten_sams_abend_fer=1}  0.10776942  0.3426295 1.158552  1.379475 0.06959906
4 {sex=weiblich} => {Lernzeiten_sonn_vorm_fer=1}   0.14786967  0.2209738 1.068710  1.309636 0.04989144
5 {sex=männlich} => {Lernzeiten_woche_abend_fer=1} 0.14285714  0.4541833 1.103313  1.289878 0.05853925
6 {sex=weiblich} => {Lernzeiten_sams_vorm_fer=1}   0.18609023  0.2780899 1.039418  1.179348 0.03388202
```

B.3.4. Semester und Lernzeiten im Semester

```
> SemLernZeit_sem <- apriori(data[c(names(LZsem), "semester")],+
 parameter = list(supp = 0.05, conf = 0.05, target = "rules", minlen=2),
 + appearance = list(rhs = c(paste(names(LZsem)[1:12],"=1",sep="")),
 + lhs = c(paste("semester=",levels(data$semester)[1:4],sep="")), default = "none"))
> quality(SemLernZeit_sem) <- cbind(quality(SemLernZeit_sem),
+ interestMeasure(SemLernZeit_sem, c("oddsRatio", "phi"), data))
> inspect(head(sort(SemLernZeit_sem, by="oddsRatio"),10))

   lhs                          rhs                              support confidence     lift oddsRatio        phi
1  {semester=Master}     => {L.zeiten_woche_vorm_sem=1}  0.16666667  0.5518672 1.228424  1.810360 0.13570255
2  {semester=S.anfänger} => {L.zeiten_woche_abend_sem=1} 0.09774436  0.8210526 1.087469  1.561487 0.05644726
3  {semester=Master}     => {L.zeiten_woche_nachm_sem=1} 0.21303258  0.7053942 1.091958  1.465726 0.08171043
4  {semester=Bachelor}   => {L.zeiten_sonn_nachm_sem=1}  0.36591479  0.7487179 1.049125  1.401213 0.07582496
5  {semester=Bachelor}   => {L.zeiten_sonn_vorm_sem=1}   0.19047619  0.3897436 1.104850  1.373479 0.07567916
6  {semester=Bachelor}   => {L.zeiten_woche_abend_sem=1} 0.38032581  0.7782051 1.030718  1.279081 0.05272330
7  {semester=Bachelor}   => {L.zeiten_sams_vorm_sem=1}   0.22243108  0.4551282 1.060415  1.230160 0.05121966
8  {semester=Bachelor}   => {L.zeiten_sams_nachm_sem=1}  0.38095238  0.7794872 1.018887  1.170606 0.03332092
9  {semester=S.anfänger} => {L.zeiten_sams_nachm_sem=1}  0.09335840  0.7842105 1.025061  1.132281 0.01662387
10 {semester=S.anfänger} => {L.zeiten_sonn_nachm_sem=1}  0.08709273  0.7315789 1.025110  1.106549 0.01457237
```

B.3.5. Alter und Lernzeiten im Semester

```
> AlterLernZeit_sem <- apriori(data[c(names(LZsem), "alter")],
+ parameter = list(supp = 0.001, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste(names(LZsem)[1:12],"=1",sep="")),
+ lhs = c(paste("alter=",levels(data$alter)[1:3],sep="")), default = "none"))
> quality(AlterLernZeit_sem) <- cbind(quality(AlterLernZeit_sem),
+ interestMeasure(AlterLernZeit_sem, c("oddsRatio", "phi"), data))
> inspect(head(sort(AlterLernZeit_sem, by="oddsRatio"),12))

   lhs               rhs                              support confidence     lift oddsRatio         phi
1  {alter=jung}   => {Lernzeiten_sams_nachm_sem=1}  0.582080201  0.7866215 1.028213  1.552878 0.085879698
2  {alter=jung}   => {Lernzeiten_sonn_nachm_sem=1}  0.545112782  0.7366638 1.032235  1.518306 0.085848029
3  {alter=mittel} => {Lernzeiten_woche_vorm_sem=1}  0.122807018  0.5051546 1.124445  1.346089 0.063697695
4  {alter=jung}   => {Lernzeiten_sams_vorm_sem=1}   0.330827068  0.4470787 1.041661  1.328742 0.060941774
5  {alter=jung}   => {Lernzeiten_sonn_vorm_sem=1}   0.273182957  0.3691787 1.046553  1.327147 0.057976153
6  {alter=mittel} => {Lernzeiten_sams_nacht_sem=1}  0.046992481  0.1932990 1.134210  1.229708 0.034475307
7  {alter=alt}    => {Lernzeiten_sonn_vorm_sem=1}   0.001253133  0.4000000 1.133925  1.224005 0.005542643
8  {alter=mittel} => {Lernzeiten_woche_nacht_sem=1} 0.073308271  0.3015464 1.101300  1.198063 0.035252528
9  {alter=mittel} => {Lernzeiten_sams_abend_sem=1}  0.112155388  0.4613402 1.065556  1.164249 0.032464401
10 {alter=jung}   => {Lernzeiten_woche_abend_sem=1} 0.561403509  0.7586791 1.004856  1.078476 0.014381692
11 {alter=jung}   => {Lernzeiten_sams_nie_sem=1}    0.040726817  0.0550381 1.009665  1.040445 0.003914678
12 {alter=mittel} => {Lernzeiten_woche_nachm_sem=1} 0.158521303  0.6520619 1.009399  1.035799 0.007195897
```

B.3.6. Personenbezogene Variablen und Lernzeiten in den Ferien

```
> PersLernZeit_fer <- apriori(data[c(names(LZfer), "sex", "studienfach", "alter", "semester")],
+ parameter = list(supp = 0.05, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste(names(LZfer)[1:12],"=1",sep="")),
+ lhs = c(paste("sex=",levels(data$sex)[1:2],sep=""),
+ paste("studienfach=",levels(data$studienfach)[1:4],sep=""),
+ paste("alter=",levels(data$alter)[1:3],sep=""),
+ paste("semester=",levels(data$semester)[1:4],sep="")), default = "none"))
> quality(PersLernZeit_fer) <- cbind(quality(PersLernZeit_fer),
+ interestMeasure(PersLernZeit_fer, c("oddsRatio", "phi"), data))
> inspect(head(sort(PersLernZeit_fer, by="oddsRatio"),4))

  lhs                      rhs                              support confidence     lift oddsRatio        phi
1 {alter=jung}          => {Lernzeiten_woche_nie_fer=1}  0.08834586  0.1193903 1.161872  2.310702 0.09241082
2 {semester=Bachelor} => {Lernzeiten_woche_nie_fer=1}  0.06829574  0.1397436 1.359944  2.247636 0.11909327
3 {alter=jung,
   semester=Bachelor} => {Lernzeiten_woche_nie_fer=1}  0.06203008  0.1447368 1.408537  2.205207 0.11973249
4 {sex=weiblich,
   studienfach=GeiWi,
   alter=jung,
   semester=Master}   => {Lernzeiten_woche_vorm_fer=1} 0.05451128  0.6041667 1.411786  2.192158 0.11216174
```

B.3.7. Semester, Studienfach und Lernzeiten in den Ferien

```
> StudSemLernZeit_fer <- apriori(data[c(names(LZfer), "studienfach", "semester")],
+ parameter = list(supp = 0.05, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste(names(LZfer)[1:12],"=1",sep="")),
+ lhs = c(paste("studienfach=",levels(data$studienfach)[1:4],sep=""),
+ paste("semester=",levels(data$semester)[1:4],sep="")), default = "none"))
> quality(StudSemLernZeit_fer) <- cbind(quality(StudSemLernZeit_fer),
+ interestMeasure(StudSemLernZeit_fer, c("oddsRatio", "phi"), data))
> inspect(head(sort(StudSemLernZeit_fer, by="oddsRatio"),6))
```

```
  lhs                      rhs                                support confidence     lift oddsRatio       phi
1 {semester=Bachelor} => {Lernzeiten_woche_nie_fer=1}   0.06829574  0.1397436 1.359944  2.247636 0.1190933
2 {studienfach=GeiWi,
   semester=Master}   => {Lernzeiten_woche_vorm_fer=1}  0.10588972  0.5709459 1.334158  2.034897 0.1379116
3 {semester=Langzeit} => {Lernzeiten_woche_abend_fer=1} 0.05137845  0.5694444 1.383308  2.017223 0.1009708
4 {studienfach=GeiWi,
   semester=Master}   => {Lernzeiten_sams_vorm_fer=1}   0.07142857  0.3851351 1.439521  1.975178 0.1267538
5 {studienfach=GeiWi,
   semester=Master}   => {Lernzeiten_woche_nachm_fer=1} 0.14285714  0.7702703 1.150001  1.829964 0.1019419
6 {semester=Bachelor} => {Lernzeiten_sams_nie_fer=1}    0.15100251  0.3089744 1.220602  1.791240 0.1255635
```

B.3.8. Semester und Lernzeiten in den Ferien

```
> SemLernZeit_fer <- apriori(data[c(names(LZfer), "semester")],
+ parameter = list(supp = 0.05, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste(names(LZfer)[1:12],"=1",sep="")),
+ lhs = c(paste("semester=",levels(data$semester)[1:4],sep="")), default = "none"))
> quality(SemLernZeit_fer) <- cbind(quality(SemLernZeit_fer),
+ interestMeasure(SemLernZeit_fer, c("oddsRatio", "phi"), data))
> inspect(head(sort(SemLernZeit_fer, by="oddsRatio"),8))

  lhs                       rhs                             support confidence     lift oddsRatio        phi
1 {semester=Bachelor} => {L.zeiten_woche_nie_fer=1}   0.06829574  0.1397436 1.359944  2.247636 0.11909327
2 {semester=Langzeit} => {L.zeiten_woche_abend_fer=1} 0.05137845  0.5694444 1.383308  2.017223 0.10097077
3 {semester=Bachelor} => {L.zeiten_sams_nie_fer=1}    0.15100251  0.3089744 1.220602  1.791240 0.12556355
4 {semester=Master}   => {L.zeiten_woche_vorm_fer=1}  0.15476190  0.5124481 1.197463  1.634452 0.11234185
5 {semester=Master}   => {L.zeiten_sams_vorm_fer=1}   0.10150376  0.3360996 1.256241  1.621910 0.10186766
6 {semester=Master}   => {L.zeiten_woche_abend_fer=1} 0.14724311  0.4875519 1.184373  1.560144 0.10144418
7 {semester=Master}   => {L.zeiten_woche_nachm_fer=1} 0.22180451  0.7344398 1.096507  1.543335 0.09041141
8 {semester=Langzeit} => {L.zeiten_sams_nachm_fer=1}  0.05639098  0.6250000 1.161234  1.480277 0.05481727
```

B.3.9. Alter, Semester und Lernzeiten in den Ferien

```
> AlterSemLernZeit_fer <- apriori(data[c(names(LZfer), "alter", "semester")],
+ parameter = list(supp = 0.001, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste(names(LZfer)[1:12],"=1",sep="")),
+ lhs = c(paste("alter=",levels(data$alter)[1:3],sep=""),
+ paste("semester=",levels(data$semester)[1:4],sep="")), default = "none"))
> quality(AlterSemLernZeit_fer) <- cbind(quality(AlterSemLernZeit_fer),
+ interestMeasure(AlterSemLernZeit_fer, c("oddsRatio", "phi"), data))
> inspect(head(sort(AlterSemLernZeit_fer, by="oddsRatio"),4))

  lhs                        rhs                             support confidence     lift oddsRatio        phi
1 {alter=alt,
   semester=Bachelor}   => {L.zeiten_woche_nachm_fer=1} 0.001253133  1.0000000 1.492984       Inf 0.02487066
2 {alter=mittel,
   semester=S.anfänger} => {L.zeiten_sams_nacht_fer=1}  0.003132832  0.4166667 3.715084  5.788177 0.08399199
3 {alter=mittel,
   semester=S.anfänger} => {L.zeiten_woche_nacht_fer=1} 0.003132832  0.4166667 2.770833  4.100304 0.06484351
4 {alter=mittel,
   semester=S.anfänger} => {L.zeiten_sams_abend_fer=1}  0.004385965  0.5833333 1.972458  3.369032 0.05484936
```

B.3.10. Studienfach und Lernzeiten in den Ferien

```
> StudLernZeit_fer <- apriori(data[c(names(LZfer), "studienfach")],
+ parameter = list(supp = 0.05, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste(names(LZfer)[1:12],"=1",sep="")),
+ lhs = c(paste("studienfach=",levels(data$studienfach)[1:4],sep="")), default = "none"))
> quality(StudLernZeit_fer) <- cbind(quality(StudLernZeit_fer),
+ interestMeasure(StudLernZeit_fer, c("oddsRatio", "phi"), data))
> inspect(head(sort(StudLernZeit_fer, by="oddsRatio"),9))

  lhs                        rhs                                support confidence     lift oddsRatio        phi
1 {studienfach=GeiWi}  => {Lernzeiten_woche_nachm_fer=1} 0.38345865  0.7183099 1.072425  1.601422 0.11038411
2 {studienfach=GeiWi}  => {Lernzeiten_sams_vorm_fer=1}   0.16228070  0.3039906 1.136227  1.497470 0.08810567
3 {studienfach=GeiWi}  => {Lernzeiten_woche_vorm_fer=1}  0.25062657  0.4694836 1.097066  1.441571 0.08984076
4 {studienfach=ReWiSo} => {Lernzeiten_sams_nie_fer=1}    0.06203008  0.3036810 1.199690  1.379866 0.05890013
5 {studienfach=MaNa}   => {Lernzeiten_sams_abend_fer=1}  0.05764411  0.3471698 1.173905  1.330879 0.05028438
6 {studienfach=GeiWi}  => {Lernzeiten_sonn_nachm_fer=1}  0.26566416  0.4976526 1.067545  1.312617 0.06754506
7 {studienfach=GeiWi}  => {Lernzeiten_sams_nachm_fer=1}  0.30388471  0.5692488 1.057650  1.307392 0.06660296
8 {studienfach=GeiWi}  => {Lernzeiten_sonn_vorm_fer=1}   0.12030075  0.2253521 1.089885  1.277470 0.04910882
9 {studienfach=MaNa}   => {Lernzeiten_woche_abend_fer=1} 0.07205514  0.4339623 1.054191  1.116052 0.02022621
```

B.4. Alter, Geschlecht, Semester, Studienfach, Lernvolumen

B.4.1. Studienfach und Lernvolumen

```
> StudLernVol <- apriori(data[c("Lernvolumen_sbegi","Lernvolumen_stime","Lernvolumen_send",
+ "Lernvolumen_svac","studienfach")], parameter = list(supp = 0.05, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste("Lernvolumen_sbegi=",levels(data$Lernvolumen_sbegi)[1:4],sep=""),
+ paste("Lernvolumen_stime=",levels(data$Lernvolumen_stime)[1:4],sep=""),
+ paste("Lernvolumen_send=",levels(data$Lernvolumen_send)[1:4],sep=""),
+ paste("Lernvolumen_svac=",levels(data$Lernvolumen_svac)[1:4],sep="")),
+ lhs = c(paste("studienfach=",levels(data$studienfach)[1:4],sep="")), default = "none"))
> quality(StudLernVol) <- cbind(quality(StudLernVol),
+ interestMeasure(StudLernVol, c("oddsRatio", "phi"), data))
> inspect(head(sort(StudLernVol, by="oddsRatio"),11))

   lhs                     rhs                              support confidence     lift oddsRatio        phi
1  {studienfach=GeiWi}  => {Lernvolumen_svac=eher viel}  0.14974937  0.2805164 1.165896  1.610632 0.09992741
2  {studienfach=ReWiSo} => {Lernvolumen_sbegi=wenig}     0.06453634  0.3159509 1.286372  1.567846 0.08278786
3  {studienfach=MaNa}   => {Lernvolumen_send=viel}       0.05388471  0.3245283 1.288426  1.543208 0.07467563
4  {studienfach=GeiWi}  => {Lernvolumen_send=eher wenig} 0.12907268  0.2417840 1.102535  1.328689 0.05815424
5  {studienfach=ReWiSo} => {Lernvolumen_stime=wenig}     0.05075188  0.2484663 1.173231  1.303153 0.04549368
6  {studienfach=GeiWi}  => {Lernvolumen_stime=eher viel} 0.16604010  0.3110329 1.083861  1.288849 0.05693197
7  {studienfach=GeiWi}  => {Lernvolumen_send=wenig}      0.14035088  0.2629108 1.081458  1.261457 0.04940245
8  {studienfach=MaNa}   => {Lernvolumen_svac=eher wenig} 0.05263158  0.3169811 1.102183  1.183116 0.02896928
9  {studienfach=MaNa}   => {Lernvolumen_send=eher viel}  0.05137845  0.3094340 1.083019  1.146577 0.02342830
10 {studienfach=GeiWi}  => {Lernvolumen_sbegi=viel}      0.15288221  0.2863850 1.041163  1.129858 0.02713331
```

```
11 {studienfach=GeiWi}  => {Lernvolumen_svac=viel}       0.15162907  0.2840376 1.037355  1.116923 0.02454573
```

B.4.2. Semester und Lernvolumen

```
> SemLernVol <- apriori(data[c("Lernvolumen_sbegi","Lernvolumen_stime","Lernvolumen_send","Lernvolumen_svac",
+ "semester")], parameter = list(supp = 0.05, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste("Lernvolumen_sbegi=",levels(data$Lernvolumen_sbegi)[1:4],sep=""),
+ paste("Lernvolumen_stime=",levels(data$Lernvolumen_stime)[1:4],sep=""),
+ paste("Lernvolumen_send=",levels(data$Lernvolumen_send)[1:4],sep=""),
+ paste("Lernvolumen_svac=",levels(data$Lernvolumen_svac)[1:4],sep="")),
+ lhs = c(paste("semester=",levels(data$semester)[1:4],sep="")), default = "none"))
> quality(SemLernVol) <- cbind(quality(SemLernVol),
+ interestMeasure(SemLernVol, c("oddsRatio", "phi"), data))
> inspect(head(sort(SemLernVol, by="oddsRatio"),5))

  lhs                      rhs                               support confidence     lift oddsRatio        phi
1 {semester=Master}   => {Lernvolumen_svac=viel}       0.1102757  0.3651452 1.333574  1.879749 0.13473237
2 {semester=Master}   => {Lernvolumen_send=wenig}      0.0914787  0.3029046 1.245968  1.565722 0.09169431
3 {semester=Bachelor} => {Lernvolumen_svac=eher wenig} 0.1591479  0.3256410 1.132294  1.439247 0.08218062
4 {semester=Bachelor} => {Lernvolumen_send=viel}       0.1378446  0.2820513 1.119786  1.368524 0.06795447
5 {semester=Bachelor} => {Lernvolumen_svac=wenig}      0.1090226  0.2230769 1.126680  1.362850 0.06153860
```

B.4.3. Semester, Studienfach und Lernvolumen

```
> StudSemLernVol <- apriori(data[c("Lernvolumen_sbegi","Lernvolumen_stime","Lernvolumen_send",
+ "Lernvolumen_svac", "studienfach", "semester")],
+ parameter = list(supp = 0.05, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste("Lernvolumen_sbegi=",levels(data$Lernvolumen_sbegi)[1:4],sep=""),
+ paste("Lernvolumen_stime=",levels(data$Lernvolumen_stime)[1:4],sep=""),
+ paste("Lernvolumen_send=",levels(data$Lernvolumen_send)[1:4],sep=""),
+ paste("Lernvolumen_svac=",levels(data$Lernvolumen_svac)[1:4],sep="")),
+ lhs = c(paste("studienfach=",levels(data$studienfach)[1:4],sep=""),
+ paste("semester=",levels(data$semester)[1:4],sep="")), default = "none"))
> quality(StudSemLernVol) <- cbind(quality(StudSemLernVol),
+ interestMeasure(StudSemLernVol, c("oddsRatio", "phi"), data))
> inspect(head(sort(StudSemLernVol, by="oddsRatio"),11))

  lhs                       rhs                             support confidence     lift oddsRatio        phi
1 {semester=Master}      => {Lernvolumen_svac=viel}      0.11027569  0.3651452 1.333574  1.879749 0.13473237
2 {studienfach=GeiWi,
   semester=Master}      => {Lernvolumen_svac=viel}      0.07080201  0.3817568 1.394242  1.860082 0.11551440
3 {studienfach=GeiWi,
   semester=Master}      => {Lernvolumen_send=wenig}     0.06140351  0.3310811 1.361870  1.723790 0.09786078
4 {studienfach=GeiWi}    => {Lernvolumen_svac=eher viel} 0.14974937  0.2805164 1.165896  1.610632 0.09992741
5 {studienfach=ReWiSo}   => {Lernvolumen_sbegi=wenig}    0.06453634  0.3159509 1.286372  1.567846 0.08278786
6 {semester=Master}      => {Lernvolumen_send=wenig}     0.09147870  0.3029046 1.245968  1.565722 0.09169431
7 {studienfach=MaNa}     => {Lernvolumen_send=viel}      0.05388471  0.3245283 1.288426  1.543208 0.07467563
8 {studienfach=GeiWi,
```

```
    semester=Bachelor}  => {Lernvolumen_svac=eher viel}  0.07142857  0.2961039 1.230682  1.466093 0.07321262
9  {semester=Bachelor}  => {Lernvolumen_svac=eher wenig} 0.15914787  0.3256410 1.132294  1.439247 0.08218062
10 {semester=Bachelor}  => {Lernvolumen_send=viel}       0.13784461  0.2820513 1.119786  1.368524 0.06795447
11 {semester=Bachelor}  => {Lernvolumen_svac=wenig}      0.10902256  0.2230769 1.126680  1.362850 0.06153860
```

B.4.4. Alter und Lernvolumen

```
> AlterLernVol <- apriori(data[c("Lernvolumen_sbegi","Lernvolumen_stime","Lernvolumen_send",
+ "Lernvolumen_svac", "alter")],
+ parameter = list(supp = 0.05, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste("Lernvolumen_sbegi=",levels(data$Lernvolumen_sbegi)[1:4],sep=""),
+ paste("Lernvolumen_stime=",levels(data$Lernvolumen_stime)[1:4],sep=""),
+ paste("Lernvolumen_send=",levels(data$Lernvolumen_send)[1:4],sep=""),
+ paste("Lernvolumen_svac=",levels(data$Lernvolumen_svac)[1:4],sep="")),
+ lhs = c(paste("alter=",levels(data$alter)[1:3],sep="")), default = "none"))
> quality(AlterLernVol) <- cbind(quality(AlterLernVol),
+ interestMeasure(AlterLernVol, c("oddsRatio", "phi"), data))
> inspect(head(sort(AlterLernVol, by="oddsRatio"),11))

   lhs               rhs                               support confidence     lift oddsRatio         phi
1  {alter=mittel} => {Lernvolumen_svac=viel}        0.09398496  0.3865979 1.411923  2.022516 0.143349881
2  {alter=jung}   => {Lernvolumen_svac=eher wenig}  0.22932331  0.3099069 1.077585  1.554878 0.083157748
3  {alter=mittel} => {Lernvolumen_sbegi=viel}       0.08333333  0.3427835 1.246202  1.537434 0.085948586
4  {alter=jung}   => {Lernvolumen_sbegi=wenig}      0.19611529  0.2650296 1.079049  1.533687 0.076090133
5  {alter=jung}   => {Lernvolumen_svac=wenig}       0.15852130  0.2142252 1.081973  1.523262 0.068708481
6  {alter=mittel} => {Lernvolumen_stime=viel}       0.07205514  0.2963918 1.182603  1.364244 0.059848729
7  {alter=jung}   => {Lernvolumen_stime=eher viel}  0.22117794  0.2988992 1.041579  1.258684 0.044497600
8  {alter=jung}   => {Lernvolumen_sbegi=eher viel}  0.20551378  0.2777307 1.018984  1.106858 0.019603269
9  {alter=jung}   => {Lernvolumen_send=eher viel}   0.21365915  0.2887384 1.010584  1.059006 0.011292524
10 {alter=jung}   => {Lernvolumen_stime=wenig}      0.15852130  0.2142252 1.011549  1.058443 0.010098416
11 {alter=jung}   => {Lernvolumen_stime=eher wenig} 0.18734336  0.2531753 1.010169  1.053928 0.009921068
```

```
12 {alter=mittel} => {Lernvolumen_send=viel}        0.06265664  0.2577320 1.023234  1.041667 0.007640533
```

B.4.5. Geschlecht und Lernvolumen

```
> SexLernVol <- apriori(data[c("Lernvolumen_sbegi","Lernvolumen_stime","Lernvolumen_send","Lernvolumen_svac",
+ "sex")], parameter = list(supp = 0.05, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste("Lernvolumen_sbegi=",levels(data$Lernvolumen_sbegi)[1:4],sep=""),
+ paste("Lernvolumen_stime=",levels(data$Lernvolumen_stime)[1:4],sep=""),
+ paste("Lernvolumen_send=",levels(data$Lernvolumen_send)[1:4],sep=""),
+ paste("Lernvolumen_svac=",levels(data$Lernvolumen_svac)[1:4],sep="")),
+ lhs = c(paste("sex=",levels(data$sex)[1:2],sep="")), default = "none"))
> quality(SexLernVol) <- cbind(quality(SexLernVol),
+ interestMeasure(SexLernVol, c("oddsRatio", "phi"), data))
> inspect(head(sort(SexLernVol, by="oddsRatio"),12))

   lhs                rhs                               support confidence     lift oddsRatio         phi
1  {sex=männlich} => {Lernvolumen_svac=wenig}        0.07142857  0.2270916 1.146956  1.297438 0.049461865
2  {sex=weiblich} => {Lernvolumen_svac=eher viel}    0.16917293  0.2528090 1.050737  1.228730 0.040617279
3  {sex=männlich} => {Lernvolumen_stime=wenig}       0.07393484  0.2350598 1.109927  1.220786 0.038598126
4  {sex=weiblich} => {Lernvolumen_stime=eher wenig} 0.17418546  0.2602996 1.038596  1.171073 0.031744596
5  {sex=weiblich} => {Lernvolumen_sbegi=eher wenig} 0.14348371  0.2144195 1.037011  1.153931 0.026874128
6  {sex=männlich} => {Lernvolumen_sbegi=viel}        0.09273183  0.2948207 1.071831  1.153668 0.029972396
7  {sex=männlich} => {Lernvolumen_send=wenig}        0.08020050  0.2549801 1.048836  1.097820 0.018748304
8  {sex=weiblich} => {Lernvolumen_send=eher viel}    0.19486216  0.2911985 1.019195  1.085164 0.017265584
9  {sex=weiblich} => {Lernvolumen_send=eher wenig}   0.14974937  0.2237828 1.020449  1.083070 0.015414344
10 {sex=männlich} => {Lernvolumen_stime=eher viel}   0.09335840  0.2968127 1.034308  1.072316 0.014743565
11 {sex=weiblich} => {Lernvolumen_svac=eher wenig}   0.19548872  0.2921348 1.015789  1.069647 0.014267629
12 {sex=weiblich} => {Lernvolumen_sbegi=eher viel}   0.18421053  0.2752809 1.009996  1.042553 0.008702196
```

B.5. Alter, Geschlecht, Semester, Studienfach, Nutzer

B.5.1. Semester und Nutzer

```
> SemNutz<- apriori(data[c("nutzer", "semester")], parameter = list(supp = 0.001, conf = 0.05,
+ target = "rules", minlen=2), appearance = list(rhs = c(paste("nutzer=",levels(data$nutzer)[1:2],sep="")),
+ default = "lhs"))
> quality(SemNutz) <- cbind(quality(SemNutz), interestMeasure(SemNutz, c("oddsRatio", "phi"), data))
> inspect(head(sort(SemNutz, by="oddsRatio"),4))

  lhs                               rhs                     support confidence     lift oddsRatio         phi
1 {semester=Studienanfänger} => {nutzer=Nichtnutzer} 0.04699248  0.3947368 1.122995  1.234568 0.033287534
2 {semester=Master}          => {nutzer=Nutzer}      0.20488722  0.6784232 1.046148  1.209787 0.041231139
3 {semester=Bachelor}        => {nutzer=Nichtnutzer} 0.17606516  0.3602564 1.024901  1.077985 0.017923512
4 {semester=Langzeitstudent} => {nutzer=Nutzer}      0.05889724  0.6527778 1.006602  1.020914 0.002824093
```

B.5.2. Alter und Nutzer

```
> AlterNutz<- apriori(data[c("nutzer", "alter")], parameter = list(supp = 0.001, conf = 0.05,
+ target = "rules", minlen=2), appearance = list(rhs = c(paste("nutzer=",levels(data$nutzer)[1:2],sep="")),
+ default = "lhs"))
> quality(AlterNutz) <- cbind(quality(AlterNutz), interestMeasure(AlterNutz, c("oddsRatio", "phi"), data))
> inspect(head(sort(AlterNutz, by="oddsRatio"),3))

  lhs               rhs                     support confidence     lift oddsRatio        phi
1 {alter=alt}    => {nutzer=Nutzer}      0.003132832  1.0000000 1.542029       Inf 0.04127254
2 {alter=mittel} => {nutzer=Nichtnutzer} 0.092105263  0.3788660 1.077843  1.169824 0.03247992
3 {alter=jung}   => {nutzer=Nutzer}      0.484962406  0.6553768 1.010610  1.122088 0.02431117
```

B.5.3. Studienfach und Nutzer

```
> StudNutz<- apriori(data[c("nutzer", "studienfach")],
+ parameter = list(supp = 0.05, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste("nutzer=",levels(data$nutzer)[1:2],sep="")), default = "lhs"))
> quality(StudNutz) <- cbind(quality(StudNutz), interestMeasure(StudNutz, c("oddsRatio", "phi"), data))
> inspect(head(sort(StudNutz, by="oddsRatio"),2))

  lhs                      rhs                      support confidence     lift oddsRatio        phi
1 {studienfach=GeiWi} => {nutzer=Nutzer}      0.3753133  0.7030516 1.084126  1.672517 0.12227914
2 {studienfach=MaNa}  => {nutzer=Nichtnutzer} 0.0745614  0.4490566 1.277530  1.639357 0.09117067
```

B.5.4. Geschlecht und Nutzer

```
> SexNutz<- apriori(data[c("nutzer", "sex")],
+ parameter = list(supp = 0.05, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste("nutzer=",levels(data$nutzer)[1:2],sep="")), default = "lhs"))
> quality(SexNutz) <- cbind(quality(SexNutz),
+ interestMeasure(SexNutz, c("oddsRatio", "phi"), data))
> inspect(head(sort(SexNutz, by="oddsRatio"),5))

  lhs               rhs                     support confidence      lift oddsRatio         phi
1 {sex=weiblich} => {nutzer=Nichtnutzer} 0.2399749  0.3586142 1.0202287 1.0994013  0.02118110
2 {sex=männlich} => {nutzer=Nutzer}      0.2073935  0.6593625 1.0167562 1.0723186  0.01541724
3 {sex=männlich} => {nutzer=Nichtnutzer} 0.1071429  0.3406375 0.9690862 0.9325587 -0.01541724
4 {sex=weiblich} => {nutzer=Nutzer}      0.4291980  0.6413858 0.9890354 0.9095860 -0.02118110
```

B.6. Alter, Geschlecht, Semester, Studienfach, Nutzer, Anzahl Services

B.6.1. Nutzer und Anzahl Services

```
> NutzServ <- apriori(data[c("nutzer", "anzahl_services")], parameter = list(supp = 0.05, conf = 0.05,
+ target = "rules", minlen=2), appearance = list(
+ rhs = c(paste("anzahl_services=",levels(data$anzahl_services)[1:4],sep="")), default = "lhs"))
> quality(NutzServ) <- cbind(quality(NutzServ), interestMeasure(NutzServ, c("oddsRatio", "phi"), data))
> inspect(head(sort(NutzServ, by="oddsRatio"),3))

  lhs                        rhs                              support confidence     lift oddsRatio        phi
1 {nutzer=Nutzer}         => {anzahl_services=viele}       0.09649123  0.1487923 1.152779  1.711036 0.07988743
2 {nutzer=Nichtnutzer} => {anzahl_services=eher wenige} 0.13345865  0.3796791 1.126334  1.337135 0.06632561
3 {nutzer=Nutzer}         => {anzahl_services=eher viele}  0.29761905  0.4589372 1.049375  1.285634 0.05912697
```

B.6.2. Studienfach und Anzahl Services

```
> StudServ <- apriori(data[c("studienfach", "anzahl_services")], parameter = list(supp = 0.01, conf = 0.05,
+ target = "rules", minlen=2), appearance = list(
+ rhs = c(paste("anzahl_services=",levels(data$anzahl_services)[1:4],sep="")), default = "lhs"))
> quality(StudServ) <- cbind(quality(StudServ), interestMeasure(StudServ, c("oddsRatio", "phi"), data))
> inspect(head(sort(StudServ, by="oddsRatio"),6))

  lhs                     rhs                                support confidence     lift oddsRatio        phi
1 {studienfach=Med}   => {anzahl_services=wenige}      0.02005013  0.2091503 2.167558  2.863569 0.12424208
2 {studienfach=MaNa}  => {anzahl_services=wenige}      0.02694236  0.1622642 1.681647  2.128886 0.09939647
3 {studienfach=GeiWi} => {anzahl_services=viele}       0.08834586  0.1654930 1.282169  2.071600 0.11624344
4 {studienfach=MaNa}  => {anzahl_services=eher wenige} 0.07456140  0.4490566 1.332146  1.774087 0.10568436
5 {studienfach=GeiWi} => {anzahl_services=eher viele}  0.25814536  0.4835681 1.105694  1.499491 0.09971824
6 {studienfach=Med}   => {anzahl_services=eher wenige} 0.04010025  0.4183007 1.240907  1.470061 0.05593838
```

B.6.3. Studienfach, Nutzer und Anzahl Services

```
> StudNutzServ <- apriori(data[c("studienfach", "nutzer", "anzahl_services")],
+ parameter = list(supp = 0.01, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste("anzahl_services=",levels(data$anzahl_services)[1:4],sep="")),
+ default = "lhs"))
> quality(StudNutzServ) <- cbind(quality(StudNutzServ),
+ interestMeasure(StudNutzServ, c("oddsRatio", "phi"), data))
> inspect(head(sort(StudNutzServ, by="oddsRatio"),10))

   lhs                         rhs                                support confidence     lift oddsRatio        phi
1  {studienfach=Med,
    nutzer=Nichtnutzer} => {anzahl_services=wenige}      0.01065163  0.2328767 2.413450  3.071168 0.10112756
2  {studienfach=Med}    => {anzahl_services=wenige}      0.02005013  0.2091503 2.167558  2.863569 0.12424208
3  {studienfach=MaNa,
    nutzer=Nichtnutzer} => {anzahl_services=wenige}      0.01378446  0.1848739 1.915966  2.310997 0.08496509
4  {studienfach=GeiWi,
    nutzer=Nutzer}      => {anzahl_services=viele}       0.07080201  0.1886477 1.461562  2.260100 0.13772800
5  {studienfach=MaNa}   => {anzahl_services=wenige}      0.02694236  0.1622642 1.681647  2.128886 0.09939647
6  {studienfach=GeiWi}  => {anzahl_services=viele}       0.08834586  0.1654930 1.282169  2.071600 0.11624344
7  {studienfach=MaNa,
    nutzer=Nichtnutzer} => {anzahl_services=eher wenige} 0.03508772  0.4705882 1.396020  1.834947 0.08015835
8  {studienfach=MaNa}   => {anzahl_services=eher wenige} 0.07456140  0.4490566 1.332146  1.774087 0.10568436
9  {nutzer=Nichtnutzer} => {anzahl_services=wenige}      0.04573935  0.1301248 1.348566  1.761840 0.08386367
10 {nutzer=Nutzer}      => {anzahl_services=viele}       0.09649123  0.1487923 1.152779  1.711036 0.07988743
```

B.6.4. Alter, Semester und Anzahl Services

```
> AlterSemServ <- apriori(data[c("semester", "alter", "anzahl_services")],
+ parameter = list(supp = 0.001, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste("anzahl_services=",levels(data$anzahl_services)[1:4],sep="")),
+ default = "lhs"))
> quality(AlterSemServ) <- cbind(quality(AlterSemServ),
+ interestMeasure(AlterSemServ, c("oddsRatio", "phi"), data))
> inspect(head(sort(AlterSemServ, by="oddsRatio"),20))

   lhs                        rhs                                 support confidence     lift oddsRatio        phi
1  {semester=Bachelor,
    alter=alt}             => {anzahl_services=eher wenige} 0.00125313  1.0000000 2.966543       Inf 0.04967325
2  {semester=S.anfänger} => {anzahl_services=wenige}      0.02568922  0.2157895 2.236364  3.148601 0.14852792
3  {semester=S.anfänger,
    alter=jung}            => {anzahl_services=wenige}      0.02380952  0.2183908 2.263323  3.145791 0.14441653
4  {alter=alt}             => {anzahl_services=eher wenige} 0.00187970  0.6000000 1.779926  2.960748 0.03117823
5  {semester=Master}       => {anzahl_services=viele}       0.06077694  0.2012448 1.559159  2.323007 0.14159318
6  {semester=S.anfänger,
    alter=mittel}          => {anzahl_services=viele}       0.00187970  0.2500000 1.936893  2.267652 0.03139275
7  {alter=jung}            => {anzahl_services=wenige}      0.08270677  0.1117697 1.158340  2.247855 0.08729118
8  {semester=Langzeit,
    alter=mittel}          => {anzahl_services=eher viele}  0.04824561  0.6209677 1.419863  2.245075 0.10743710
9  {semester=S.fänger,
    alter=jung}            => {anzahl_services=eher wenige} 0.05513785  0.5057471 1.500320  2.210233 0.12480196
10 {semester=S.anfänger} => {anzahl_services=eher wenige} 0.05952381  0.5000000 1.483271  2.173815 0.12668451
```

```
11 {semester=Langzeit.}  => {anzahl_services=eher viele}  0.05451128  0.6041667 1.381447  2.100870 0.10590627
12 {semester=Master,
    alter=jung}          => {anzahl_services=viele}       0.03947368  0.2065574 1.600318  2.089927 0.11232970
13 {alter=mittel}        => {anzahl_services=viele}       0.04511278  0.1855670 1.437694  1.826186 0.09549467
14 {semester=Bachelor,
    alter=mittel}        => {anzahl_services=viele}       0.01002506  0.1951220 1.511722  1.689314 0.04584626
15 {semester=Master,
    alter=mittel}        => {anzahl_services=viele}       0.02005013  0.1882353 1.458367  1.668499 0.06092618
16 {semester=Bachelor,
    alter=jung}          => {anzahl_services=wenige}      0.05075188  0.1184211 1.227273  1.543856 0.06432143
17 {semester=Bachelor}   => {anzahl_services=wenige}      0.05576441  0.1141026 1.182517  1.488122 0.05831556
18 {alter=jung}          => {anzahl_services=eher wenige} 0.26566416  0.3590178 1.065042  1.478876 0.07824208
19 {alter=mittel}        => {anzahl_services=eher viele}  0.12343358  0.5077320 1.160946  1.455508 0.08041772
20 {semester=Master,
    alter=mittel}        => {anzahl_services=eher viele}  0.05513785  0.5176471 1.183617  1.435586 0.05589427
```

B.7. Alter, Geschlecht, Semester, Studienfach, Nutzer, Anzahl Services, Zufriedenheits-Score

```
> AlterSemSexStudNutzServZS <- apriori(data[c("nutzer", "sex", "alter", "studienfach", "semester",
+ "anzahl_services", "zscore")], parameter = list(supp = 0.1, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste("zscore=",levels(data$zscore)[1:3],sep="")), default = "lhs"))
> quality(AlterSemSexStudNutzServZS) <- cbind(quality(AlterSemSexStudNutzServZS),
+ interestMeasure(AlterSemSexStudNutzServZS, c("oddsRatio", "phi"), data))
> inspect(head(sort(AlterSemSexStudNutzServZS, by="oddsRatio"),16))

  lhs                               rhs                   support confidence     lift oddsRatio        phi
1 {nutzer=Nutzer,
   studienfach=GeiWi}            => {zscore=Zufrieden} 0.2230576  0.5943239 1.194636  1.869741 0.15011076
2 {studienfach=GeiWi}            => {zscore=Zufrieden} 0.3032581  0.5680751 1.141874  1.841304 0.15106329
3 {nutzer=Nutzer,
   sex=weiblich,
   studienfach=GeiWi}            => {zscore=Zufrieden} 0.1685464  0.6031390 1.212355  1.809255 0.13158429
4 {nutzer=Nutzer,
   sex=weiblich,
   alter=jung,
   studienfach=GeiWi}            => {zscore=Zufrieden} 0.1271930  0.6023739 1.210817  1.712301 0.10852526
5 {nutzer=Nutzer,
   alter=jung,
   studienfach=GeiWi}            => {zscore=Zufrieden} 0.1585213  0.5897436 1.185429  1.663355 0.11186504
6 {nutzer=Nutzer,
   studienfach=GeiWi,
   anzahl_services=eher viele} => {zscore=Zufrieden} 0.1109023  0.6000000 1.206045  1.662885 0.09762437
```

```
7  {sex=weiblich,
    studienfach=GeiWi}          => {zscore=Zufrieden} 0.2299499  0.5716511 1.149062  1.647085 0.12166986
8  {sex=weiblich,
    studienfach=GeiWi,
    anzahl_services=eher viele} => {zscore=Zufrieden} 0.1140351  0.5928339 1.191641  1.610641 0.09305806
9  {studienfach=GeiWi,
    anzahl_services=eher viele} => {zscore=Zufrieden} 0.1503759  0.5825243 1.170918  1.586769 0.10031899
10 {studienfach=GeiWi,
    semester=Master}            => {zscore=Zufrieden} 0.1090226  0.5878378 1.181598  1.564252 0.08622028
11 {nutzer=Nutzer}              => {zscore=Zufrieden} 0.3477444  0.5362319 1.077867  1.557793 0.10523569
12 {nutzer=Nutzer,
    sex=weiblich}               => {zscore=Zufrieden} 0.2399749  0.5591241 1.123882  1.542837 0.10688486
13 {alter=jung,
    studienfach=GeiWi}          => {zscore=Zufrieden} 0.2105263  0.5647059 1.135101  1.538062 0.10363935
14 {sex=weiblich,
    alter=jung,
    studienfach=GeiWi}          => {zscore=Zufrieden} 0.1679198  0.5690021 1.143737  1.503418 0.09253939
15 {nutzer=Nutzer,
    sex=weiblich,
    anzahl_services=eher viele} => {zscore=Zufrieden} 0.1140351  0.5741325 1.154050  1.469305 0.07630944
16 {nutzer=Nutzer,
    anzahl_services=eher viele} => {zscore=Zufrieden} 0.1679198  0.5642105 1.134106  1.464521 0.08685896
```

B.7.1. Nutzer, Anzahl Services und Zufriedenheits-Score

```
> NutzServZS <- apriori(data[c("nutzer", "anzahl_services", "zscore")],
+ parameter = list(supp = 0.01, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste("zscore=",levels(data$zscore)[1:3],sep="")), default = "lhs"))
> quality(NutzServZS) <- cbind(quality(NutzServZS),
+ interestMeasure(NutzServZS, c("oddsRatio", "phi"), data))
> inspect(head(sort(NutzServZS, by="oddsRatio"),5))

  lhs                            rhs                       support confidence     lift oddsRatio        phi
1 {anzahl_services=viele}  => {zscore=Zufrieden}   0.08646617  0.6699029 1.346555  2.270714 0.13274618
2 {nutzer=Nutzer,
   anzahl_services=viele}  => {zscore=Zufrieden}   0.06453634  0.6688312 1.344401  2.194972 0.11198644
3 {nutzer=Nutzer,
   anzahl_services=wenige} => {zscore=Unzufrieden} 0.01065163  0.2098765 1.840456  2.173295 0.06972076
4 {nutzer=Nichtnutzer,
   anzahl_services=viele}  => {zscore=Zufrieden}   0.02192982  0.6730769 1.352935  2.129350 0.06444607
5 {anzahl_services=wenige} => {zscore=Unzufrieden} 0.01879699  0.1948052 1.708292  2.053268 0.08304259
```

B.7.2. Anzahl Services und Zufriedenheits-Score

```
> ServZS <- apriori(data[c("anzahl_services", "zscore")],
+ parameter = list(supp = 0.01, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste("zscore=",levels(data$zscore)[1:3],sep="")), default = "lhs"))
> quality(ServZS) <- cbind(quality(ServZS),
+ interestMeasure(ServZS, c("oddsRatio", "phi"), data))
> inspect(head(sort(ServZS, by="oddsRatio"),6))

  lhs                                rhs                      support confidence     lift oddsRatio        phi
1 {anzahl_services=viele}       => {zscore=Zufrieden}   0.08646617  0.6699029 1.346555  2.270714 0.13274618
2 {anzahl_services=wenige}      => {zscore=Unzufrieden} 0.01879699  0.1948052 1.708292  2.053268 0.08304259
3 {anzahl_services=wenige}      => {zscore=Neutral}     0.04949875  0.5129870 1.320528  1.754258 0.08348615
4 {anzahl_services=eher wenige} => {zscore=Unzufrieden} 0.04761905  0.1412639 1.238776  1.477416 0.06108719
5 {anzahl_services=eher viele}  => {zscore=Zufrieden}   0.23558897  0.5386819 1.082791  1.340902 0.07262698
6 {anzahl_services=eher wenige} => {zscore=Neutral}     0.14223058  0.4219331 1.086137  1.235079 0.04895659
```

B.7.3. Alter, Anzahl Services und Zufriedenheits-Score

```
> AlterServZS <- apriori(data[c("alter","anzahl_services", "zscore")],
+ parameter = list(supp = 0.005, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste("zscore=",levels(data$zscore)[1:3],sep="")), default = "lhs"))
> quality(AlterServZS) <- cbind(quality(AlterServZS),
+ interestMeasure(AlterServZS, c("oddsRatio", "phi"), data))
> inspect(head(sort(AlterServZS, by="oddsRatio"),11))

   lhs                                rhs                         support confidence     lift oddsRatio        phi
1  {alter=jung,
    anzahl_services=viele}       => {zscore=Zufrieden}    0.057017544  0.6946565 1.396312  2.465932 0.11791720
2  {anzahl_services=viele}       => {zscore=Zufrieden}    0.086466165  0.6699029 1.346555  2.270714 0.13274618
3  {alter=jung,
    anzahl_services=wenige}      => {zscore=Unzufrieden} 0.016917293  0.2045455 1.793706  2.171613 0.08550416
4  {anzahl_services=wenige}      => {zscore=Unzufrieden} 0.018796992  0.1948052 1.708292  2.053268 0.08304259
5  {alter=jung,
    anzahl_services=wenige}      => {zscore=Neutral}      0.043859649  0.5303030 1.365103  1.876246 0.08737805
6  {anzahl_services=wenige}      => {zscore=Neutral}      0.049498747  0.5129870 1.320528  1.754258 0.08348615
7  {alter=mittel,
    anzahl_services=viele}       => {zscore=Zufrieden}    0.028195489  0.6250000 1.256297  1.724522 0.05542947
8  {alter=mittel,
    anzahl_services=wenige}      => {zscore=Zufrieden}    0.006265664  0.6250000 1.256297  1.692177 0.02566247
9  {alter=mittel,
    anzahl_services=eher viele}  => {zscore=Zufrieden}    0.075187970  0.6091371 1.224412  1.676365 0.08379005
10 {anzahl_services=eher wenige} => {zscore=Unzufrieden} 0.047619048  0.1412639 1.238776  1.477416 0.06108719
```

```
11 {alter=jung,
    anzahl_services=eher wenige} => {zscore=Unzufrieden} 0.038220551  0.1438679 1.261611  1.459623 0.05645289
```

B.7.4. Studienfach und Zufriedenheits-Score

```
> StudZS <- apriori(data[c("studienfach", "zscore")],
+ parameter = list(supp = 0.01, conf = 0.05, target = "rules", minlen=2),
+ appearance = list(rhs = c(paste("zscore=",levels(data$zscore)[1:3],sep="")), default = "lhs"))
> quality(StudZS) <- cbind(quality(StudZS),
+ interestMeasure(StudZS, c("oddsRatio", "phi"), data))
> inspect(head(sort(StudZS, by="oddsRatio"),6))

  lhs                        rhs                        support confidence     lift oddsRatio        phi
1 {studienfach=Med}    => {zscore=Unzufrieden} 0.01754386  0.1830065 1.604827  1.874909 0.07065687
2 {studienfach=GeiWi}  => {zscore=Zufrieden}   0.30325815  0.5680751 1.141874  1.841304 0.15106329
3 {studienfach=MaNa}   => {zscore=Neutral}     0.07832080  0.4716981 1.214242  1.507937 0.07619210
4 {studienfach=Med}    => {zscore=Neutral}     0.04511278  0.4705882 1.211385  1.451744 0.05486029
5 {studienfach=ReWiSo} => {zscore=Neutral}     0.08897243  0.4355828 1.121274  1.278697 0.04897194
6 {studienfach=ReWiSo} => {zscore=Unzufrieden} 0.02506266  0.1226994 1.075979  1.111002 0.01381061
```

C. Beiliegende CD

Am Ende dieser Bachelorarbeit befindet sich eine CD mit folgendem Inhalt:

- Inhaltsverzeichnis.txt
- Readme.txt
- Bachelorarbeit
  - BARobertPietsch.pdf
- Data
  - survey_54795_data_file.csv
  - Surveydata_syntax.r
  - RCodeRobertPietsch.r
- Literatur
  - Agresti – Categorical Data Analysis
    - Agresti – Categorical Data Analysis Seite 44.pdf
    - Agresti – Categorical Data Analysis Seite 45.pdf
  - Bates et al. – Matrix – Sparse and Dense Matrix Classes and Methods.pdf
  - Borgelt – Einführung in Datenanalyse und Data Mining.pdf
  - Eifler et al. – Bericht zum Statistischen Praktikum Studierendenbefragung der Universitätsbibliothek.pdf
  - Geng et al. – Interestingness Measures for Data Mining A Survey.pdf
  - Hahsler et al. – arules – A Computational Environment for Mining Association Rules and Frequent Item Sets.pdf
  - Hahsler et al. – arules – Mining Association Rules and Frequent Itemsets.pdf
  - Hahsler et al. – New Probabilistic Interest Measures for Association Rules.pdf

- Hastie et al. – The Elements of Statistical Learning Data Mining, Inference, and Prediction.pdf
- Link Liste.txt
- Tan et al. – Selecting the Right Objective Measure for Association Analysis.pdf
- Tan et al. – Introduction to Data Mining Kapitel 6.pdf

Danksagung

An dieser Stelle möchte ich mich nochmals für das sehr gelungene Statistische Praktikum, in dem der Datensatz entstand, bei Dr. Antje Michel und Medea Seyder von der UB München bedanken. Derselbe Dank geht auch an meine Praktikumspartner Fabian Eifler und Malte Schierholz, die mich auch während der Bachelorarbeit immer unterstützten. Ein Dank auch an alle weiteren Mitarbeiter der UB München für die Erlaubnis, den vorliegenden Datensatz weiter verwenden zu dürfen.

Ein besonderer Dank geht an meine Betreuerin Prof. Dr. Bettina Grün für ihre freundliche und sehr engagierte Betreuung.

Ein letzter Dank geht an meine Eltern, die meine Launen ertragen haben und sicherer in der Komma-Setzung sind als ich.

Erklärung zur Urheberschaft

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.

München, den 3. August 2011

....

(Robert Pietsch)

CD (siehe Anhang C)