

データの分析と知識発見

Introduction to Data Analysis

今回の構成

0101010101010101010101010101010101

因果関係を表すルールについて理解する

アソシエーションルールで用いる
指標を理解する

Rを用いて演習を行う

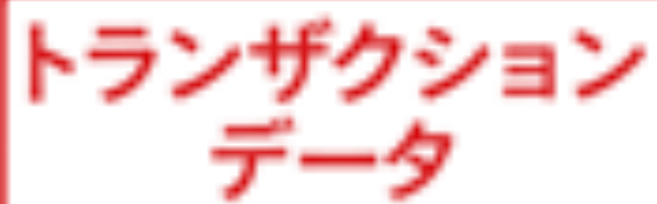
アソシエーションルール

01

100人の生徒が
A,B,C,D,E の5科目を受講する

学生	受講科目
1	{A}
2	{B}
3	{C,D}
4	{A,B,E}
5	{A,B}
⋮	⋮
100	{A,D}

トランザクション
データ



アイテム集合



アソシエーションルール

01

「もし A ならば B である」

$$A \implies B$$

A: 条件部 (ルールヘッド)

B: 結論部 (ルールボディ)

Aに占めるBの割合を求める

$n(A)$

$n(B)$



Aに占めるBの割合

$n(A)$



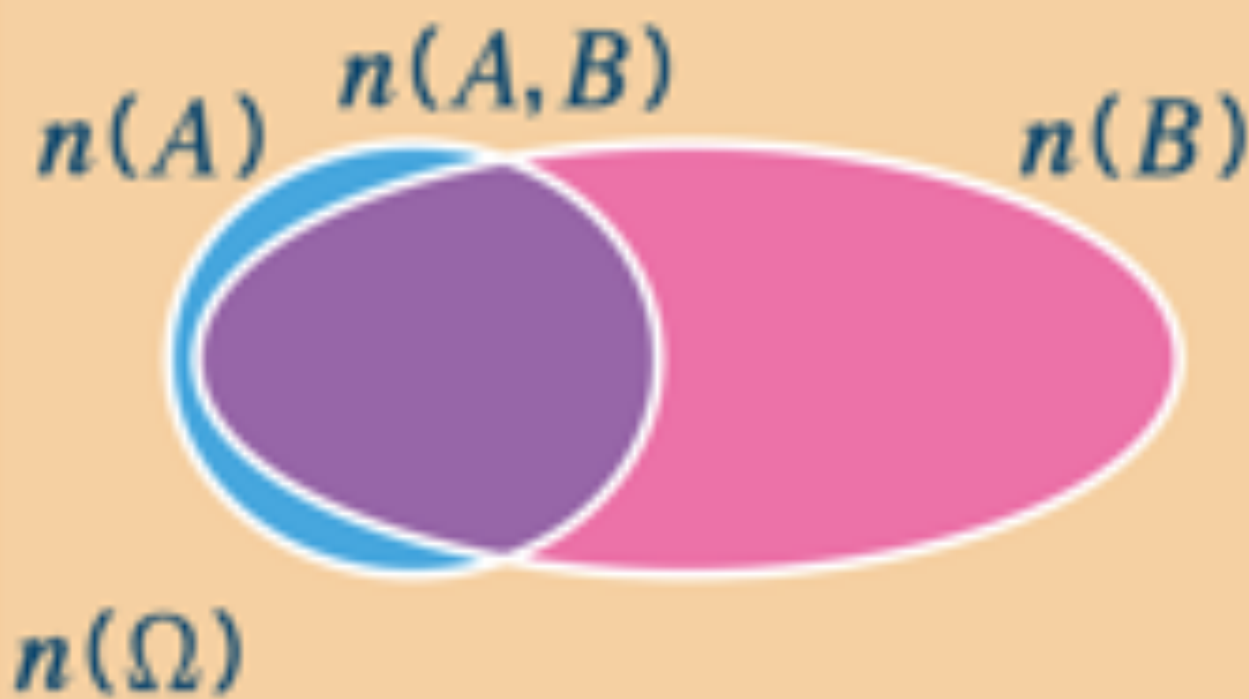
$n(A)$

$n(B)$

Bに占めるAの割合

アソシエーションルールの指標

アソシエーション分析
で用いる4つの指標



支持度 Support

$$P(A, B) = \frac{n(A, B)}{n(\Omega)}$$

期待信頼度 Expected Confidence

$$P(B) = \frac{n(B)}{n(\Omega)}$$

信頼度 Confidence

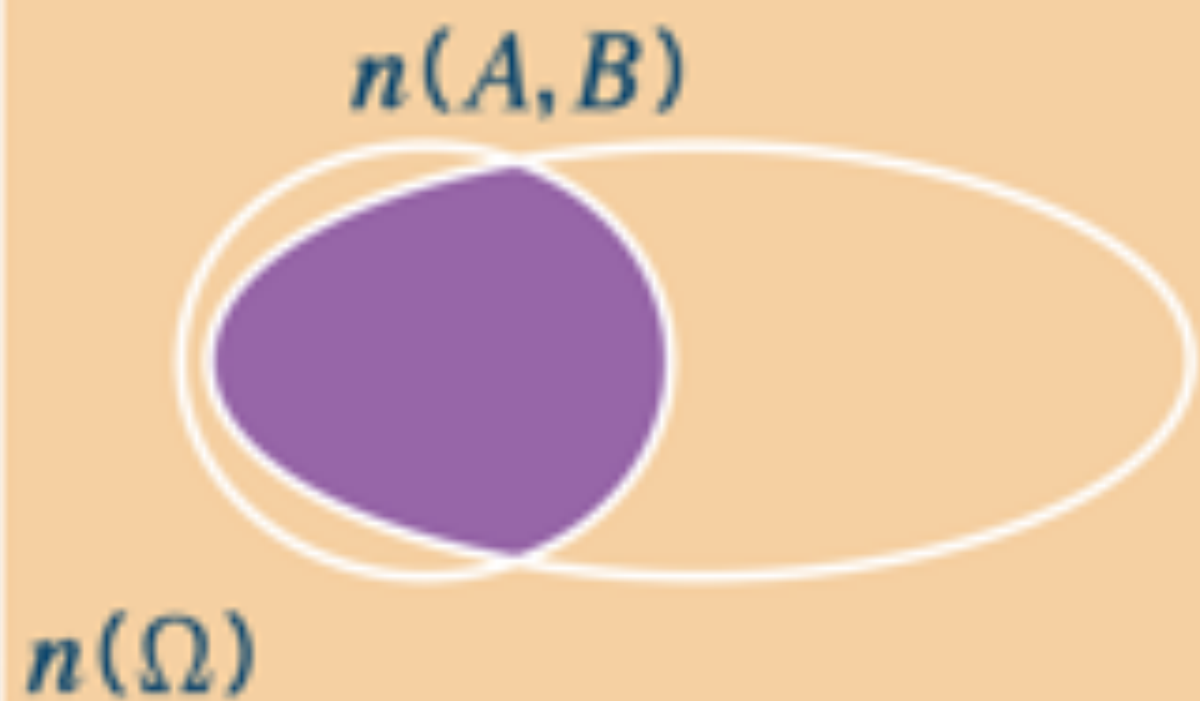
$$P(B|A) = \frac{n(A, B)}{n(A)} = \frac{P(A, B)}{P(A)}$$

リフト値 Lift

$$\frac{P(B|A)}{P(B)} = \frac{P(A, B)}{P(A)P(B)}$$

アソシエーションルールの指標

アソシエーション分析
で用いる4つの指標



支持度 Support

$$P(A, B) = \frac{n(A, B)}{n(\Omega)}$$

期待信頼度 Expected Confidence

$$P(B) = \frac{n(B)}{n(\Omega)}$$

信頼度 Confidence

$$P(B|A) = \frac{n(A, B)}{n(A)} = \frac{P(A, B)}{P(A)}$$

リフト値 Lift

$$\frac{P(B|A)}{P(B)} = \frac{P(A, B)}{P(A)P(B)}$$

アソシエーションルールの指標

アソシエーション分析
で用いる4つの指標



支持度 Support

$$P(A, B) = \frac{n(A, B)}{n(\Omega)}$$

期待信頼度 Expected Confidence

$$P(B) = \frac{n(B)}{n(\Omega)}$$

信頼度 Confidence

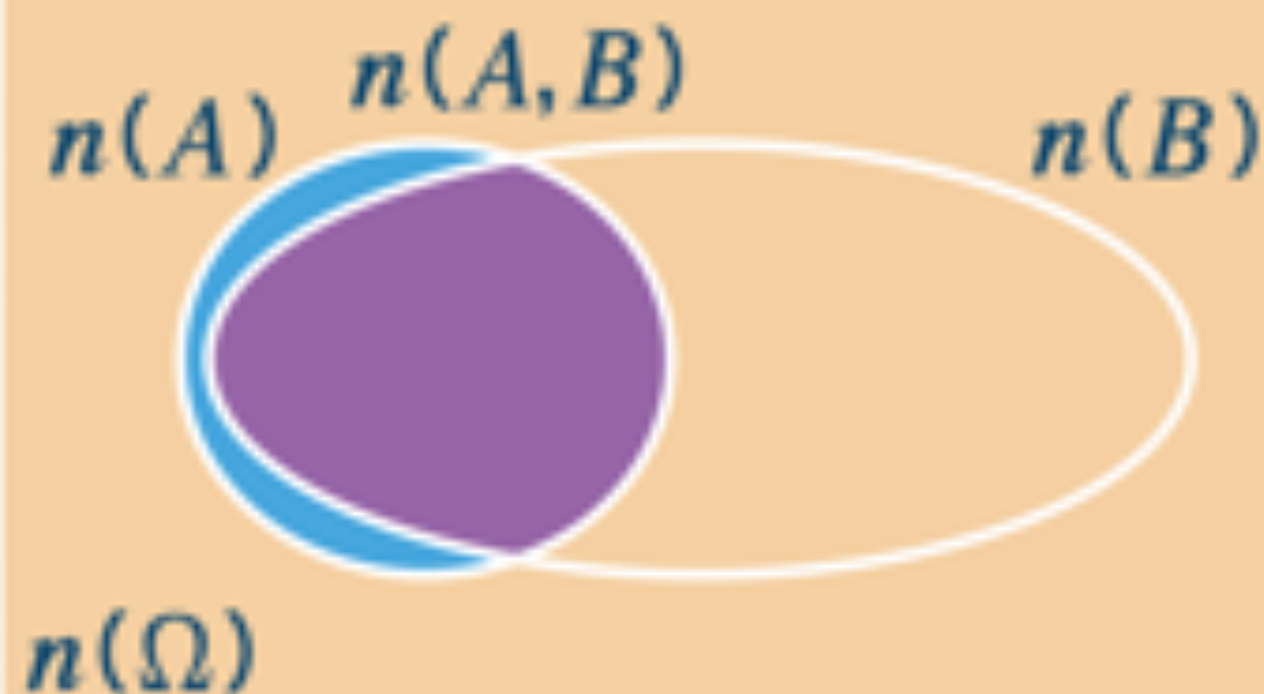
$$P(B|A) = \frac{n(A, B)}{n(A)} = \frac{P(A, B)}{P(A)}$$

リフト値 Lift

$$\frac{P(B|A)}{P(B)} = \frac{P(A, B)}{P(A)P(B)}$$

アソシエーションルールの指標

アソシエーション分析
で用いる4つの指標



支持度 Support

$$P(A, B) = \frac{n(A, B)}{n(\Omega)}$$

期待信頼度 Expected Confidence

$$P(B) = \frac{n(B)}{n(\Omega)}$$

信頼度 Confidence

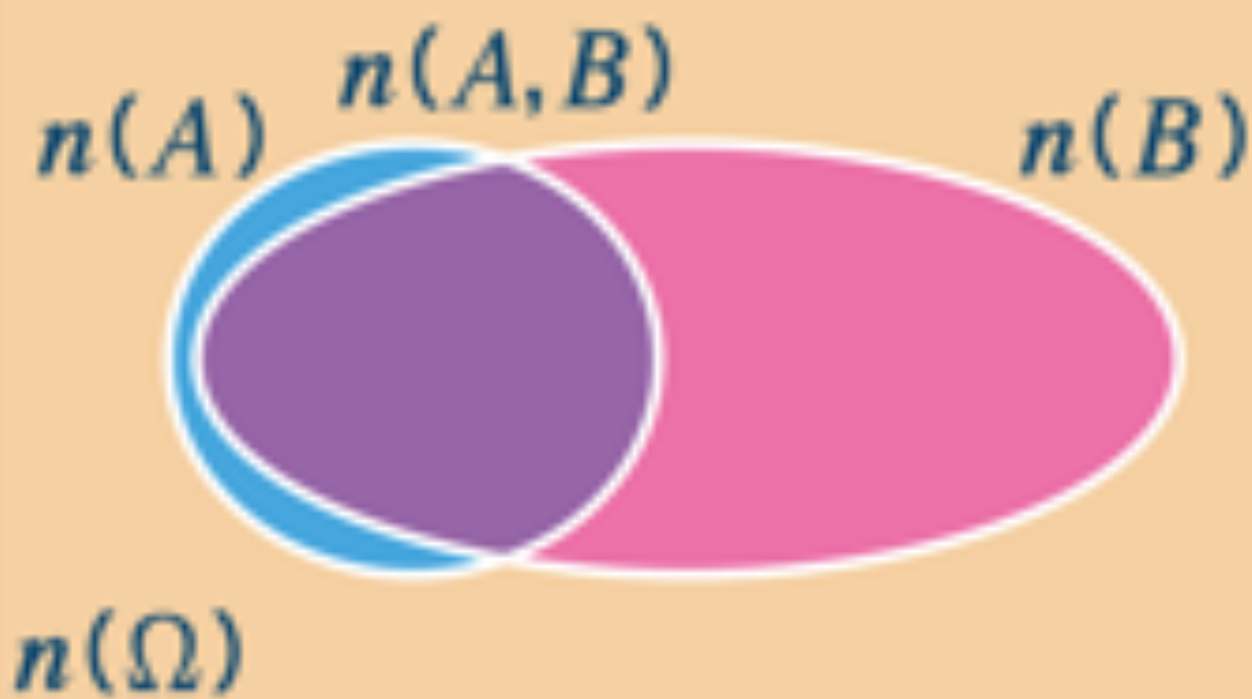
$$P(B|A) = \frac{n(A, B)}{n(A)} = \frac{P(A, B)}{P(A)}$$

リフト値 Lift

$$\frac{P(B|A)}{P(B)} = \frac{P(A, B)}{P(A)P(B)}$$

アソシエーションルールの指標

アソシエーション分析
で用いる4つの指標



支持度 Support

$$P(A, B) = \frac{n(A, B)}{n(\Omega)}$$

期待信頼度 Expected Confidence

$$P(B) = \frac{n(B)}{n(\Omega)}$$

信頼度 Confidence

$$P(B|A) = \frac{n(A, B)}{n(A)} = \frac{P(A, B)}{P(A)}$$

リフト値 Lift

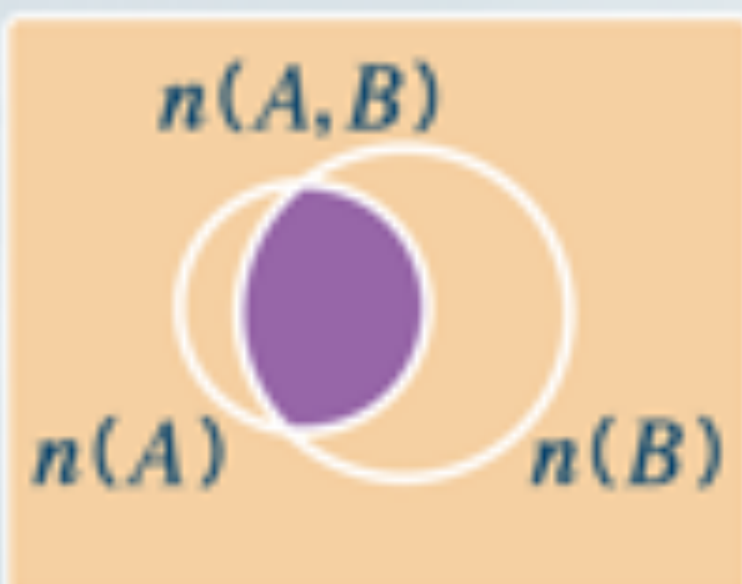
$$\frac{P(B|A)}{P(B)} = \frac{P(A, B)}{P(A)P(B)}$$

アソシエーションルールの指標の例(1)

例1

- 全部で100人の生徒がいる
- Aという科目の受講者が25人
- Bという科目の受講者が40人
- AとBの両方の受講者は20人だった

$n(\Omega)$	100
$n(A)$	25
$n(B)$	40
$n(A, B)$	20



支持度 $\frac{n(A, B)}{n(\Omega)} = \frac{20}{100} = 0.2$

期待信頼度 $\frac{n(B)}{n(\Omega)} = \frac{40}{100} = 0.4$

信頼度 $\frac{n(A, B)}{n(A)} = \frac{20}{25} = 0.8$

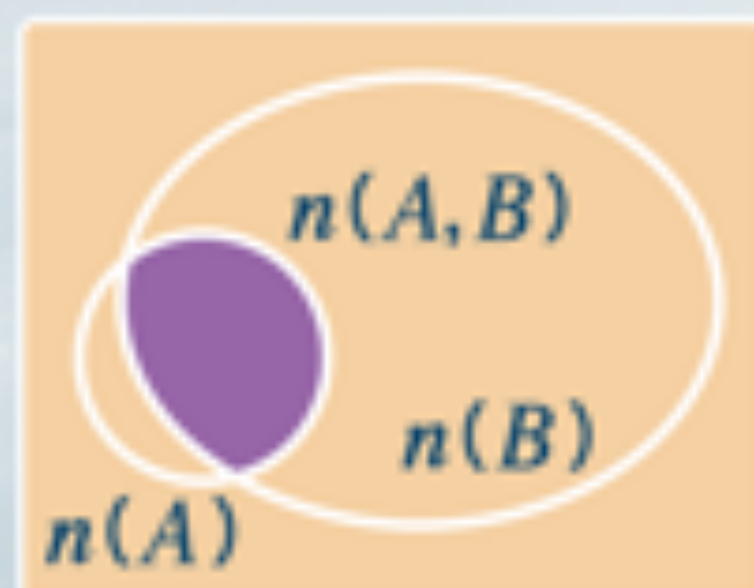
リフト値 $\frac{P(B|A)}{P(B)} = \frac{0.8}{0.4} = 2$

アソシエーションルールの指標の例(2)

例2

- 全部で100人の生徒がいる
- Aという科目の受講者が25人
- Bという科目の受講者が80人
- AとBの両方の受講者は20人だった

$n(\Omega)$	100
$n(A)$	25
$n(B)$	80
$n(A, B)$	20



支持度 $\frac{n(A, B)}{n(\Omega)} = \frac{20}{100} = 0.2$

期待信頼度 $\frac{n(B)}{n(\Omega)} = \frac{80}{100} = 0.8$

信頼度 $\frac{n(A, B)}{n(A)} = \frac{20}{25} = 0.8$

リフト値 $\frac{P(B|A)}{P(B)} = \frac{0.8}{0.8} = 1$

計算の工夫(アプリアリ)

100人の生徒が, A,B,C,D,E の5科目を受講する
「Xならば Yである」となる集合の候補

C,D の支持度が最低支持度よりも小さい場合

{A}	{A,B} {B,D}	{A,B,C} {A,D,E}	{A,B,C,D}
{B}	{A,C} {B,E}	{A,B,D} {B,C,D}	{A,B,C,E}
{C}	{A,D} {C,D}	{A,B,E} {B,C,E}	{A,B,D,E}
{D}	{A,E} {C,E}	{A,C,D} {B,D,E}	{A,C,D,E}
{E}	{B,C} {D,E}	{A,C,E} {C,D,E}	{B,C,D,E}

Rによる演習 (apriori)

```
> g0 <- Groceries
```

```
> gfrm0 <- as (g0, "data.frame")
```

```
> gdat1 <- itemFrequency (g0)
```

```
> grule2 <- apriori (g0, parameter=list (confidence=0.5, support=0.01))
```

```
> grule3 <- sort (grule2, d=T, by="confidence")
```

```
> inspect (grule3)
```

