

Atherosclerosis prevention study

2nd Department of Medicine, 1st Faculty of Medicine of
Charles University and Charles University Hospital, U
nemocnice 2, Prague 2 (head. Prof. M. Aschermann, MD,
SDr, FESC)

Atherosclerosis prevention study:

- The STULONG 1 data set is a real database that keeps information about the study of the development of atherosclerosis risk factors in a population of middle aged men.
- Used for Discovery Challenge at PKDD 00-02-03-04

Atherosclerosis prevention study:

- Study on 1400 middle-aged men at Czech hospitals
 - Measurements concern development of cardiovascular disease and other health data in a series of exams
- The aim of this analysis is to look for **associations between medical characteristics of patients and death causes.**
- Four tables
 - Entry and subsequent exams, questionnaire responses, deaths

The input data

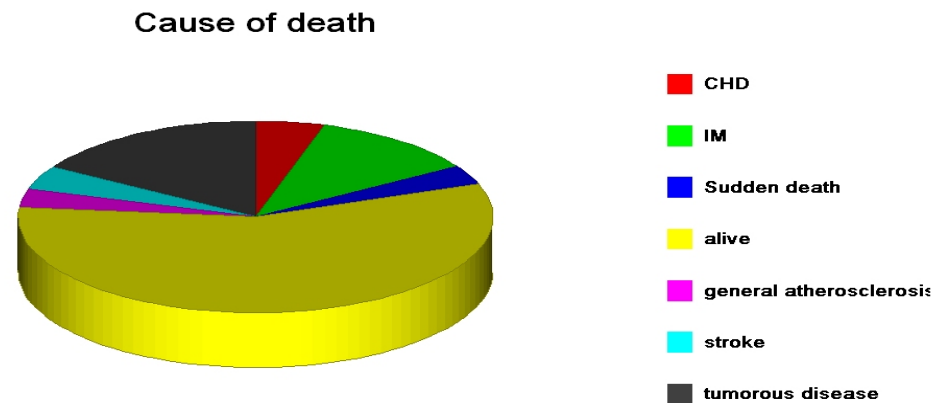
Data from Entry and Exams		
General characteristics	Examinations	habits
Marital status	Chest pain	Alcohol
Transport to a job	Breathlessness	Liquors
Physical activity in a job	Cholesterol	Beer 10
Activity after a job	Urine	Beer 12
Education	Subscapular	Wine
Responsibility	Triceps	Smoking
Age		Former smoker
Weight		Duration of smoking
Height		Tea
		Sugar
		Coffee

The input data

DEATH CAUSE	PATIENTS	%
myocardial infarction	80	20.6
coronary heart disease	33	8.5
stroke	30	7.7
other causes	79	20.3
sudden death	23	5.9
unknown	8	2.0
tumorous disease	114	29.3
general atherosclerosis	22	5.7
TOTAL	389	100.0

Data selection

- When joining “Entry” and “Death” tables we implicitly create a new attribute “Cause of death”, which is set to “alive” for subjects present in the “Entry” table but not in the “Death” table.
- We have only 389 subjects in death table.



The prepared data

Patient	General characteristics		Examinations		Habits		Cause of death
	Activity after work	Education	Chest pain	...	Alcohol	
1	moderate activity	university	not present		no		Stroke
2	great activity		not ischaemic		occasionally		myocardial infarction
3	he mainly sits		other pains		regularly		tumorous disease
.....	alive
389	he mainly sits		other pains		regularly		tumorous disease

Descriptive Analysis/ Subgroup Discovery / Association Rules

Are there strong relations concerning death cause?

General characteristics (?) \Rightarrow Death cause (?)

Examinations (?) \Rightarrow Death cause (?)

Habits (?) \Rightarrow Death cause (?)

Combinations (?) \Rightarrow Death cause (?)

Example of extracted rules

- Education(university) & Height<176-180>
⇒Death cause (tumouros disease), *16 ; 0.62*
- It means that on tumorous disease have died 16, i.e. 62% of patients with university education and with height 176-180 cm.

Example of extracted rules

- Physical activity in work(he mainly sits) & Height<176-180> \Rightarrow Death cause (tumouros disease), 24; 0.52
- It means that on tumorous disease have died 24 i.e. 52% of patients that mainly sit in the work and whose height is 176-180 cm.

Example of extracted rules

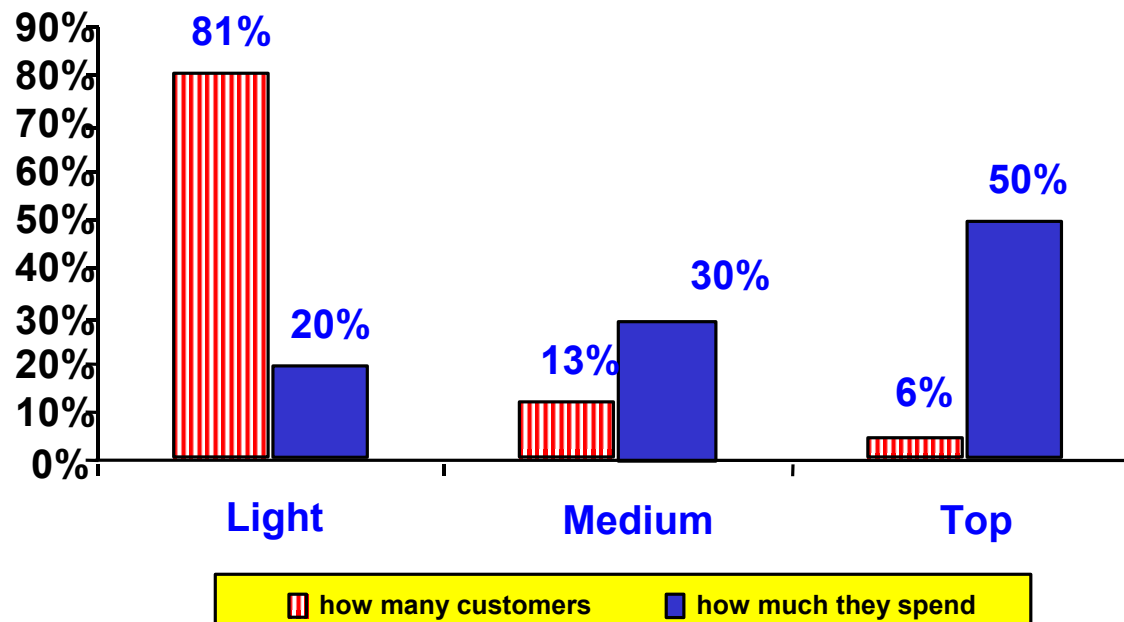
- Education(university) & Height<176-180>
⇒Death cause (tumorous disease),
16; 0.62; +1.1;
- the relative frequency of patients who died on tumorous disease among patients with university education and with height 176-180 cm is 110 per cent higher than the relative frequency of patients who died on tumorous disease among all the 389 observed patients

Conclusions

- Association rule mining
 - probably the most significant contribution from the database community to KDD
 - A large number of papers have been published
- Many interesting issues have been explored
- An interesting research direction
 - Association analysis in other types of data: spatial data, multimedia data, time series data, etc.

Conclusion (2)

- competition of supermarket retailers.
- Knowledge of customers MBA is a key factor of success in their purchasing behavior brings potentially huge added value.

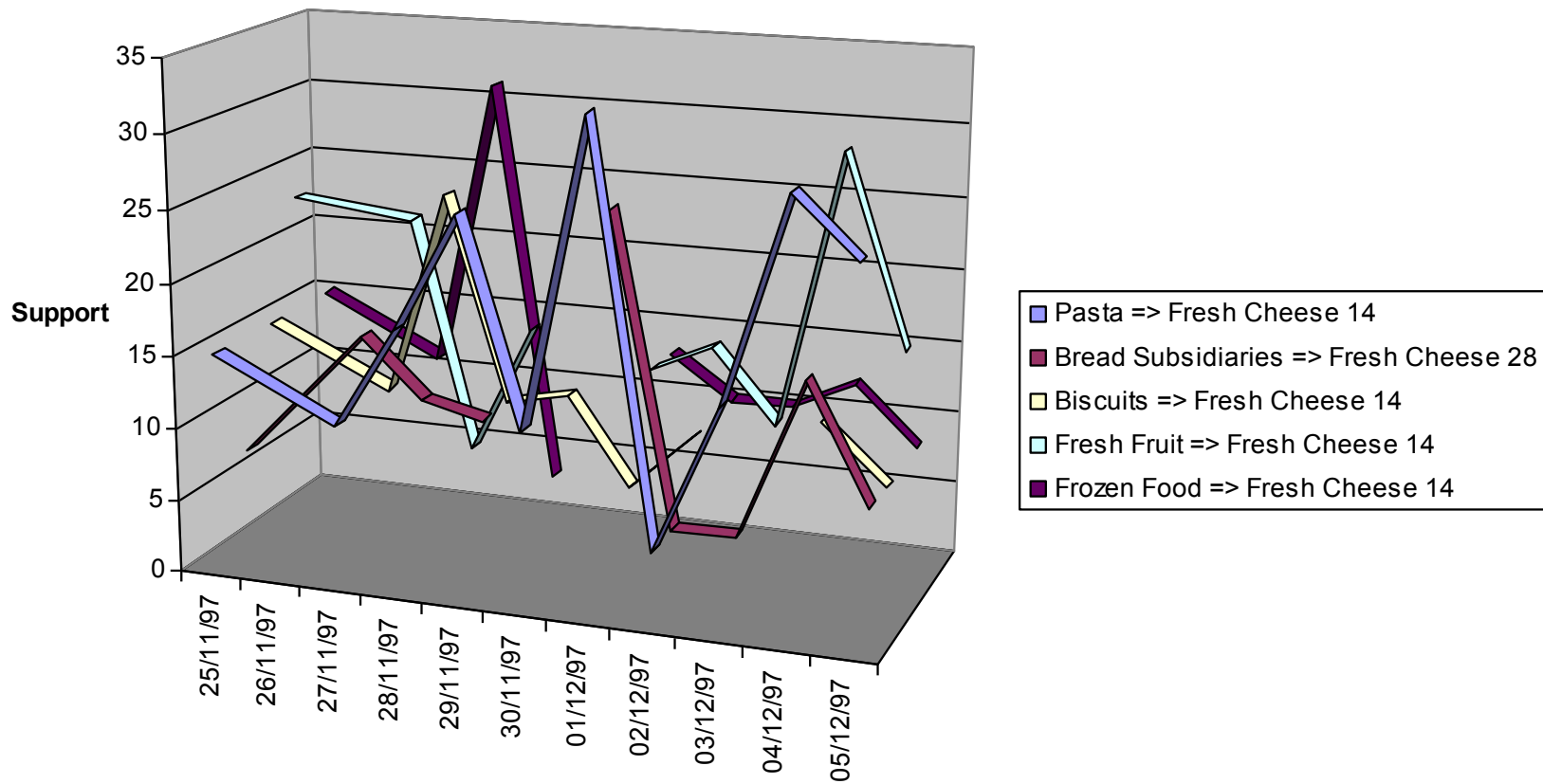


Which tools for market basket analysis?

- Association rules are needed but insufficient
- Market analysts ask for **business rules**:
 - Is supermarket assortment adequate for the company's target class of customers?
 - Is a promotional campaign effective in establishing a desired purchasing habit?

Business rules: temporal reasoning on AR

- Which rules are established by a promotion?
- How do rules change along time?



The KDD process

