# Data Mining Homework 3

Madis Nõmme

March 6th, 2014

## 1 Task

List all the mentioned in the presentation common cases of misinterpretation in statistics (HIV, cot death case etc.), also provide the correct interpretations.

Misinterpretations:

1. Not realizing that patterns can overlap thus making them more frequent.
2. Taking into account only the probability of result being correct and not considering the chance that the test subject having the disease in first place.
3. Related to 2nd - not considering the external factors affecting the probability of outcome. So considering events independent where they are not.
4. Giving wrong meaning to numbers
    - likelyhood of 2 explainations: chance of being innocent vs very unlikely event happening (2 children in a row dying)
    - after taking a test with .99% accuracy your chance of actually being sick is .99% not considering whether you were sick at first place
5. Reasoning incorrectly with uncertainty and assuming it's correct. Trusting the 'expert' blindly.

## 6 Task

Examine the applicability of association rule mining in the following domains:

- Give an example of a hypothetical association rule that can be generated from the domain.
- Describe how you would construct the transaction matrix (table 1) in order to derive the example pattern given above. Specifically, what are the baskets and items?
- What are the limitations of your proposed transaction matrix? Specif- ically, will there be any interesting rules missed out by your choice of transaction matrix?

1. Text documents

- hypothetical association: {presence of exclamation mark} -> {article popularity}, {article length, inclusion of person's name in title} -> { article popularity}
- transaction matrix:
- limitations: none

2. Stock market data

- hypothetical association: {stock having closing price lower than opening price previous day} -> {same things on next day}
- transaction matrix:
- limitations:

3. Census data

- hypothetical association: { working hours per week more than average, university} -> { more than average income}
- transaction matrix:
- limitations: