

Energy Informatics

System Design — Data Analysis

Albert-Ludwigs-Universität Freiburg



UNI
FREIBURG

Peter Thiemann

11 Feb 2016

Second application

Analysis of voting patterns

Second application



Voting data

The German Bundestag runs two kinds of voting procedures

- anonymous vote
- vote by role call

Voting data

The German Bundestag runs two kinds of voting procedures

- anonymous vote
- vote by role call

Vote by role call (namentliche Abstimmung)

- for controversial topics
- the protocol registers votes along with the name of the voter
- procedure
 - there are voting cards in three colors, blue, red, and white
 - voting cards are imprinted with name and fraction
 - cards are dropped in an urn
 - outcome is published in a table

Procedure

- For the vote by role call procedure, we can download the outcome of the votes for last few years
- Use this data to demonstrate the steps
 - forming a hypothesis
 - obtaining the data
 - cleaning up the data
 - verifying the hypothesis

Voting by role call

- <https://www.bundestag.de/abstimmung>
- only accessible inside Germany
- only available in German
- Voting data from 2009-2016 in pdf and xls
- <https://www.bundestag.de/bundestag/plenum/abstimmung/2016>
- <https://www.bundestag.de/bundestag/plenum/abstimmung/2015>
- etc

Two formats

- pdf: unstructured — useless
https://www.bundestag.de/blob/404086/029a3812d1a1a63979de77b48fbbabc2/20160128_2-data.pdf
- xls: structured — can load into spreadsheet program or read with Python
https://www.bundestag.de/blob/404080/8d985dd7bac5ecff733d4b98d40a2c07/20160128_2_xls-data.xls

A look through the spreadsheet



- Is this data valid?
- What can go wrong?
- Consistency?

- Consider the columns
(ja, nein, Enthaltung, ungültig, nicht abgegeben)
that is: yes, no, abstain, invalid, not voted
- Each of them contains 0/1
- Internal consistency of one voting protocol
 - Each row should contain exactly one “1” entry in these columns
 - The columns (ja, nein, Enthaltung, ungültig, nicht abgegeben) should contain as many “1” as the summary count announces

Consistency check (spreadsheet)



- Load into spreadsheet program
- Make a copy (save with new name)
- Create a new work sheet
- On the new sheet create a formula that checks
 - whether each of the yes/no columns contains 0 or 1
 - that the yes/no columns contain exactly one 1
 - the above two points for each row

A sad truth of vote by role call ...

- Most of the time the fractions demand that their members vote according to the party line
 - some fractions do not enforce the whip
 - for some ballots, fractions do not impose restrictions
- As the voting is public, it is much tougher to “defect” or “betray” the party by voting differently

A sad truth of vote by role call ...

- Most of the time the fractions demand that their members vote according to the party line
 - some fractions do not enforce the whip
 - for some ballots, fractions do not impose restrictions
- As the voting is public, it is much tougher to “defect” or “betray” the party by voting differently

Task

- Write formulas to check that parliamentarians mostly adhere to the whip
- Which parties do / do not impose restrictions?

Second application

Analysis of **voting patterns**

- We just looked at one ballot

- We just looked at one ballot
- Can hardly be called a voting pattern!

- We just looked at one ballot
- Can hardly be called a voting pattern!
- Using a spreadsheet is ok to analyze one data set

- We just looked at one ballot
- Can hardly be called a voting pattern!
- Using a spreadsheet is ok to analyze one data set
- But in 2015, there were 56 such ballots

- We just looked at one ballot
- Can hardly be called a voting pattern!
- Using a spreadsheet is ok to analyze one data set
- But in 2015, there were 56 such ballots
- Solution: program!

Install Python module xlrd and pandas

- (from shell)
- `pip install xlrd`
- `pip install pandas`

Analyzing xls in Python

Prepare for processing with pandas

```
import pandas
import xlrd
wb = xlrd.open_workbook (
    '20160128_2_xls-data.xls')
sheet = pandas.read_excel(book, engine="xlrd")
```

Analyzing xls in Python

Prepare for processing with pandas

```
import pandas
import xlrd
wb = xlrd.open_workbook (
    '20160128_2_xls-data.xls')
sheet = pandas.read_excel(book, engine="xlrd")
```

Dataframe

- sheet is a pandas *dataframe*
- len (sheet) get # rows
- Columns are named according to first row in xls file
- sheet.columns lists columns
- sheet.index lists row indexes

Consistency

for a single row



```
def check_consistency(row):
    ja    = row['ja']
    nein  = row['nein']
    enth  = row['Enthaltung']
    ung   = row[u'ungueltig'] # should be u umlaut
    na    = row['nichtabgegeben']
    valid = ((ja == 0 or ja == 1) and
              (nein == 0 or nein == 1) and
              (enth == 0 or enth == 1) and
              (ung == 0 or ung == 1) and
              (na == 0 or na == 1) and
              (ja + nein + enth + ung + na == 1))
    return valid
```

Consistency

for all rows



Pattern for processing dataframe

```
for idx, row in sheet.iterrows():  
    # idx - row number  
    # row - representation of the row  
    #      indexed by column names
```

Consistency

for all rows



Pattern for processing dataframe

```
for idx, row in sheet.iterrows():
    # idx - row number
    # row - representation of the row
    #      indexed by column names
```

Example (continued)

```
invalid_set = set()
for idx, row in sheet.iterrows():
    if not check_consistency(row):
        invalid_set.add(idx)

print("invalid entries: " +
      str(len(invalid_set)))
print(invalid_set)
```