university of groningen



Seminar in Methodology and Statistics

# Fisher's Exact Test

## as a measure of association strength

Jelke Bloem

# Contents

- Using Fisher's Exact Test
- Animacy classification
- Data
- Calculating association strength
- Results
- Evaluation
- Conclusion

# Fisher's Exact Test

- Used to test for associations between two variables
  - Identifying dependent bigrams

- Computes a p-value
- Calculates significance exactly (unlike $\chi^2$ test)
- Based on **hypergeometric** distribution
  - Drawing from a finite population without replacement

# Using Fisher's Exact Test

- Natural language data is skewed
  - Fisher's test does not require a normal distribution of data
- Sparse data problem
  - Fisher's test can be used with small sample sizes

However, Fisher's Exact Test is more computationally intensive

# Using Fisher's Exact Test

- Bigram association
  - Identifying dependent word pairs

    *white*, *snow*

- Standard test for collostructional analysis
  - Computing the association between words and syntactic frames (Stefanowitsch and Gries, 2003)

    [N *waiting to happen*]

# Problem: Determining animacy of nouns

- Human: doctor, player, photographer, Englishman
- Inanimate: banana, Netherlands, feeling, crime

- Automatically determine this based on co-occurrence with verbs

- The doctor thought John was right
- The banana thought John was right

# Animacy: Data

- A dictionary of words and their animacy property
- A list of verbs and their subject arguments, extracted from a corpus, with frequency counts

```
<noun animacy="nonanimate">gevoel</noun>
<noun animacy="nonanimate">IJsselmeer</noun>
<noun animacy="nonanimate">noord</noun>
<noun animacy="nonanimate">paasei</noun>
<noun animacy="human">doctor</noun>
<noun animacy="human">Engelsman</noun>
<noun animacy="human">roker</noun>
<noun animacy="human">symfonieorkest</noun>
<noun animacy="nonhuman">fuchsia</noun>
<noun animacy="nonhuman">pony</noun>
<noun animacy="nonhuman">yeti</noun>
```

From Cornetto lexical database

```
85#blijf|intransitive|su#gevoel
298#ontsta|intransitive|su#gevoel
1#schrijf|transitive|su#gevoel
8#rest|intransitive|su#gevoel

7#ontdek|transitive|su#Engelsman
4#ontwerp|transitive|su#Engelsman
3#overschat|transitive|su#Engelsman

1#besta|intransitive|su#Yeti
2#duik_op|part_intransitive(op)|su#Yeti
1#lijk|pc_pp(op)|su#Yeti
```

From Alpino Treebank

# Fisher's Exact Test on animacy data

- Hypothesis: Animate nouns are associated with different verbs than inanimate nouns

- Variables:
  1. Verb is "ontstaan" (*to start, to arise*)
  2. Subject is "gevoel" (*feeling*)
- Binary variables
- 4 classifications

# Contingency table

- The Fisher's exact test is calculated using tables
- Totals are fixed

The noun "gevoel" (*feeling*) as a subject of the verb "ontstaan" (*to start, to arise*)

| | gevoel | ¬gevoel | Row totals |
|---|---|---|---|
| ontstaan | **298** | 5927 | **6225** |
| ¬ontstaan | 405 | 111952 | 112357 |
| Column totals | **703** | 117879 | **118582** |

$p < 0.00001$

# Dependence and independence

- The p-value can go both ways: Association strength

The noun “gevoel” (*feeling*) as a subject of the verb “schrijven” (*to write*)

| | gevoel | ¬gevoel | Row totals |
|---|---|---|---|
| schrijven | **1** | 299 | **300** |
| ¬schrijven | 702 | 117578 | 118282 |
| Column totals | **703** | 117879 | **118582** |

$p > 0.99999$

# Association strength

- This p-value can be used as a measure of association strength
- A low value indicates a strong association, a high value indicates none

- Because the totals are fixed, you cannot compare p-values from samples of different sizes

# Hypothesis

- H0: The noun x and the verb y are independent in subject relations
- H1: The noun x occurs as a subject of the verb y more often than would be expected by chance

# Calculating the value

› The p-value expresses the total probability of the observed distribution (table) and all the more extreme ones

| | gevoel | ¬gevoel |
|---|---|---|
| ontstaan | **298** | 5927 |
| ¬ontstaan | 405 | 111952 |

| | gevoel | ¬gevoel |
|---|---|---|
| ontstaan | **300** | 5925 |
| ¬ontstaan | 403 | 111950 |

| | gevoel | ¬gevoel |
|---|---|---|
| ontstaan | **299** | 5926 |
| ¬ontstaan | 404 | 111951 |

| | gevoel | ¬gevoel |
|---|---|---|
| ontstaan | **301** | 5924 |
| ¬ontstaan | 402 | 111949 |

# Calculating the value

| | gevoel | ¬gevoel | totals |
|---|---|---|---|
| ontstaan | 298 | 5927 | 6225 |
| ¬ontstaan | 405 | 111952 | 112357 |
| totals | 703 | 117879 | 118582 |

- $P(n) = \frac{6225!*112357!*703!*117879!}{298!*5927!*405!*111952!*118582!}$
- $P(n+1) = \frac{6225!*112357!*703!*117879!}{299!*5926!*404!*111951!*118582!}$
- etc
- $p = P(n) + P(n+1) + P(n+2) + \ldots$
- A and B are associated more strongly than would be expected by chance ($\alpha = 0.001$)

# Association strength

## "gevoel" subject relations (inanimate)

| | | |
|---|---|---|
| 0.000000000000000 | ontsta | *arise* |
| 0.000000000000830 | heb | *have* |
| 0.000000000002380 | speel | *play* |
| 0.000000000501125 | ben | *be* |
| 0.000000003404273 | zeg | *say* |
| | | |
| 0.731409478841741 | krijg | *get* |
| 0.823487761949459 | spreek | *speak* |
| 0.853510038160385 | neem | *take* |
| 0.902189553992116 | ken | *know* |
| 1.000000000002866 | schrijf | *write* |

# Association strength

"hippie" subject relations (human)

| | | |
|---|---|---|
| 0.001468162077883 | ga | *go* |
| 0.019216198962412 | kom | *come* |
| 0.048523337414639 | noem | *call, name* |
| 0.053750193619017 | zeg | *say* |
| 0.101731760645688 | vind | *think, find* |
| | | |
| 0.847872307894773 | heb | *have* |
| 1.000000000000009 | maak | *make* |

# Association strength

"haai" (*shark*) subject relations (nonhuman)

```
0.000512195152676  heb (have)
0.049576264380802  zit (sit)
0.067926380355665  verschijn (appear)
0.184926869408314  besta (exist)

1.000000000000016  sta (stand)
1.000000000000070  doe (do)
1.000000000000196  maak (make)
1.000000000000472  lig (lie, be situated)
```

'Zit' is also a position verb

# Classification task

- For a list of nouns, decide whether they refer to something human, nonhuman animate or inanimate, using these subject-verb associations

- Train a classifier on the weighted associations
- Test it on novel data (unseen nouns)

# Evaluation

| Measure of association | Correctly classified |
|---|---|
| Pointwise Mutual Information | 90.64% |
| Fisher's Exact Test | 88.85% |
| Frequency | 83.09% |
| None (Baseline) | 81.81% |

# Fisher's Exact Test for association strength

- Fisher's Exact Test is a very robust measure
- It is computationally intensive
- Cannot compare data from samples of different sizes

- For this animacy classification task, PMI seems to perform better

# References

L. Øvrelid. Empirical evaluations of animacy annotation. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 2009.

T. Pedersen. Fishing for exactness, 1996.

A. Stefanowitsch and S.T. Gries. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8 (2):209-243, 2003.

D. Wiechmann. On the computation of collostruction strength: Testing measures of association as expressions of lexical bias. *Corpus Linguistics and Linguistic Theory*, 4(2):253-290, 2008.

A. Zaenen, J. Carletta, G. Garretson, J. Bresnan, A. Koontz-Garboden, T. Nikitina, M.C. O'Connor, and T. Wasow. Animacy Encoding in English: why and how. In *Proceedings of the 2004 ACL Workshop on Discourse Annotation*, pages 118-125. Association for Computational Linguistics, 2004.

Cornetto lexical database: http://www2.let.vu.nl/oz/cltl/cornetto/

Alpino Treebank: http://odur.let.rug.nl/~vannoord/trees/

# Questions?

university of groningen



# Thank you for your attention