Detecting linguistic variation with geographic sampling

Ezequiel Koile, George Moroz

Linguistic Convergence Laboratory, NRU HSE

... 2020

Outline of the talk

- introduction
- blar

Introduction

As Dorian (2010) mentioned

Simulated data

- total number of locations (N): $20, 40, 60, 80, 100, 120$
- type of spatial relations:
  - random
  - two more or less separable regions
  - central and periphery
- proportion of variation in the explored variable (p): $0.1, 0.2, \dots 0.5$
- amount of clusters (k): $0.1 \times N, 0.2 \times N, \dots 0.9 \times N$
- percentage of observations taken from each cluster (r): $0.1, 0.2, \dots 0.9$

From those values we could derive a number of sampled locations (n):

$$n = \frac{N \times r}{k}$$

Conclusion

References

Dorian, N. C. (2010). *Investigating variation: The effects of social organization and social setting.* Oxford University Press.