BYU
BRIGHAM YOUNG UNIVERSITY

Sample size matters when calculating Pillai scores

Joseph A. Stanley & Betsy Sneller

MICHIGAN STATE UNIVERSITY

Pillai Scores (Pillai 1955)

```
# R code
m <- manova(cbind(F1, F2) ~ vowel)
summary <- summary(m)
stats <- summary$stats
stats[1, "Pillai")
```

What is a Pillai score?

- Output of a MANOVA model ▶
  - 0 suggests complete overlap
  - 1 suggests complete separation

How are they used in sociophonetics?

- Quantify mergers (Hay, Warren, & Drager 2006, Fridland et al. 2014, Amengual & Chamorro 2015, Nadeu & Renwick 2016, Tse 2018, Gonzales & Starr 2020, Kettig 2021, etc.)
- Quantify splits (Babel et al. 2013, Fisher et al. 2015, Berry & Ernestus 2018)

Are they good measures?

- Often supported in meta-analyses (Nycz & Hall-Lew 2013, Kelley & Tucker 2020)
  - Appear to be better than Euclidean distance, mixed-effects regression, spectral overlap, SOAM, VOACH, APP, etc.
- However, Bhattacharyya's Affinity may be better suited for vowel data since it's robust to skewed data. (Johnson 2015)

Issues

1. What is considered merged?

- *p*-values from MANOVAs inconsistently reported.
  - (Especially in American English–based studies…)
- Ad hoc thresholds are inconsistent across studies.

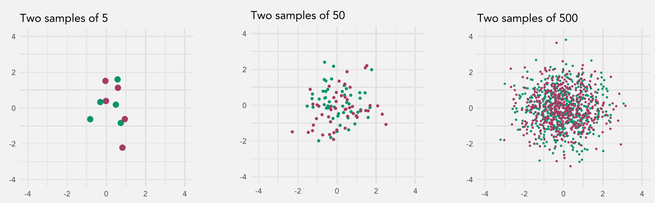
2. Sample size

- Typically not considered when comparing Pillai scores between studies or between individuals in the same study.

Methods

Data Generation

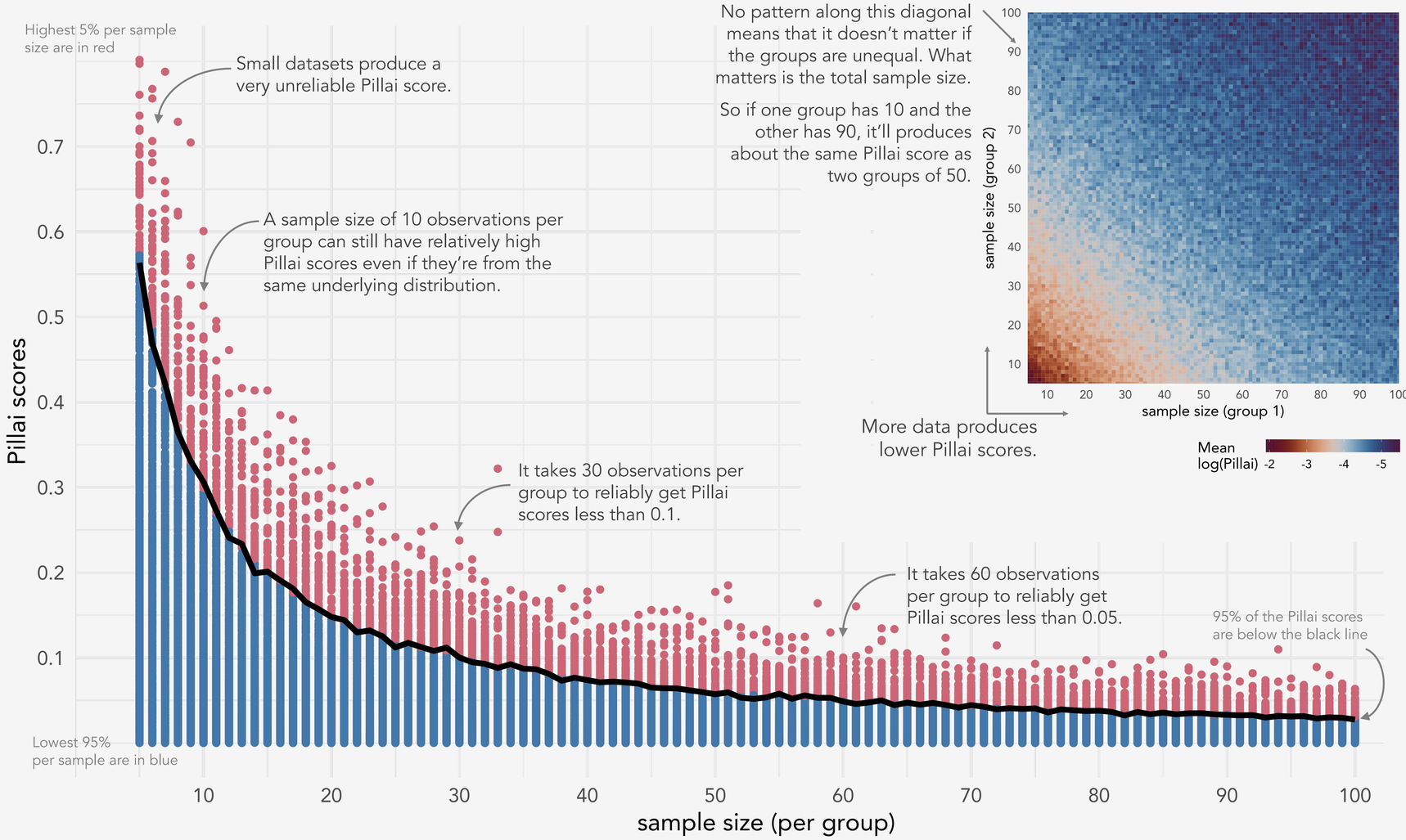
- Two samples drawn from the same bivariate normal distribution.
- In theory, the Pillai score should be 0 because they're pulled from **literally the same** underlying distribution.



Two Experiments

- Simulated sample sizes from 5 to 100
  - Experiment 1: two distributions were equally sized
  - Experiment 2: two distributions were unequally sized
- Repeated this 1000 times for each sample size.
  - Produced 921,600 pairs of distributions total.
  - Calculated the Pillai scores of each pair.

Experiment 1: Equal Sample Sizes



Experiment 2: Unequal Sample Sizes

No pattern along this diagonal means that it doesn't matter if the groups are unequal. What matters is the total sample size.

So if one group has 10 and the other has 90, it'll produces about the same Pillai score as two groups of 50.

More data produces lower Pillai scores.

Implications

Mergers are probably underreported and separation is probably overreported

- It takes a lot of data (more than many studies use) to get reliably low Pillai scores
- Mergers may be more common previously reported

Statistical significance should be reported

- Reporting *p*-values from MANOVA tests removes the need for ad hoc thresholds. puts less weight on Pillai scores, and makes interpretation more objective.

Comparison across studies

- Analyses of speakers with less data will look *less merged*
- Analysis of speakers with more data will look *more merged*

Comparison within studies

- Speakers with less data will have inflated Pillai scores.
- **Reading tasks** will have higher scores than **conversational data** (which will be interpreted as style differences)

References

Amengual, Mark & Pilar Chamorro. 2015. The Effects of Language Dominance in the Perception and Production of the Galician Mid Vowel Contrasts. *Phonetica* 72.

Babel, Molly, McAuliffe Michael, & Graham Haber. 2013. Can mergers-in-progress be unmerged in speech accommodation? *Frontiers in Psychology*.

Berry, Grant M. & Ernestus Mirjam. 2018. Phonetic alignment in English as a *lingua franca*: Coming together while splitting apart. *Second Language Research* 34(3).

Fisher, Sabriya, Hilary Prichard, & Betsy Sneller. 2015. The Apple Doesn't Fall Far From the Tree: Incremental Change in Philadelphia Families. *PWPL* 21(2).

Fridland, Valerie, Tyler Kendall, & Charlie Farrington. 2014. Durational and Spectral Differences in American English Vowels: Dialect Variation within and across Regions. *JASA* 136(1).

Gonzales, Wilkinson Daniel Wong & Rebecca Lurie Starr. 2020. Vowel System or Vowel Systems? Variation in the Monophthongs of Philippine Hybrid Hokkien in Manila. *Journal of Pidgin and Creole Languages*. 35(2).

Hay, Jennifer, Paul Warren, & Katie Drager. 2006. Factors Influencing Speech Perception in the Context of a Merger-in-Progress. *Journal of Phonetics* 34(4).

Johnson, Daniel Ezra. 2015. Quantifying Vowel Overlap with Bhattacharyya's Affinity. NWAV44. Toronto.

Kelley, Matthew C. & Benjamin V. Tucker. 2020. A Comparison of Four Vowel Overlap Measures. *JASA* 147(1).

Kettig, Thomas T. 2021. Ha'ina 'Ia Mai Ana Ka Puana: The Vowels of 'Ōlelo Hawai'i. Ph.D Dissertation. University of Hawai'i at Mānoa.

Nycz, Jennifer & Lauren Hall-Lew. 2013. Best Practices in Measuring Vowel Merger. *Proceedings of Meetings on Acoustics* 20(1): 060008.

Pillai, K. C. S. 1955.Some New Test Criteria in Multivariate Analysis. *The Annals of Mathematical Statistics* 26(1).

Nadeu, Marianna & Margaret E. L. Renwick. 2016. Variation in the Lexical Distribution and Implementation of Phonetically Similar Phonemes in Catalan. *Journal of Phonetics* 58.

Tse, Holman. 2015. Beyond the Monolingual Core and out into the Wild: A Variationist Study of Early Bilingualism and Sound Change in Toronto Heritage Cantonese. PhD Dissertation. University of Pittsburgh.

joeystanley.com
@joey_stan
joey_stanley@byu.edu

betsysneller.github.io
@betsysneller
sneller7@msu.edu