

STA723 Case Study 2 - Group 5 Report

Youngsoo Baek, Phuc Nguyen, Irene Ji

Executive Summary

1. Introduction

2. Materials and Methods

3. Results

3.1 Exploratory Data Analysis and Preprocessing

...It is unclear what quantity `availability_365` variable is measuring, or how precise a measure it can serve as for whatever quantity inherent to a listing.

3.2 Main Results

...Assuming log price and log number of monthly reviews are both isotropic and intrinsically stationary processes, we can define its semivariogram as

$$\gamma(\|\mathbf{h}\|) \equiv \frac{1}{2} \mathbb{E}[Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})].$$

The customary, simple nonparametric estimator for semivariogram is (BCG, 2011)

$$\hat{\gamma}(d) = \frac{1}{2|N(d)|} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(d)} [Y(\mathbf{s}_i) - Y(\mathbf{s}_j)]^2,$$

where $N(d)$ consists of all pairs of locations that have a pairwise Euclidean distance d .

...It is clear that for monthly review rates, we observe a negative spatial autocorrelation: closer things have more different response rather than similar. This must raise alarming sign to possible efforts to apply traditional spatial models for this response, as naive conditional autoregressive models assumes positive spatial autocorrelation.

4. Discussion

References

Sarthak, N., Post: “Availability_365=0?”, Discussion thread: New York City Airbnb Open Data, Kaggle. <https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data/discussion/111835>

Banrjee, S., Carlin, B. P., and Gelfand, A. E. (2011). *Hierarchical Modeling and Analysis for Spatial Data*.

Appendix: Figures and Tables