

## STATISTICAL ANALYSIS AND DATA MINING

Original Article

## Principal axes analysis of symbolic histogram variables

Sun Makosso - Kallyth 

First published: 15 September 2015

<https://doi.org/10.1002/sam.11270>

About



Access

als



are

## Abstract

We present a new method to perform a principal axes analysis of symbolic histogram variables. In the symbolic data analysis framework, several Histogram Principal component Analysis (Histogram PCA) have been proposed. Some approaches focus on the relationships between some specific features of histograms such as the means or the quantiles. Others use the association for distributional variables based on the squared Wasserstein distance. In this paper, we propose two new approaches. The first one uses new correlation measures based on Fisher's z scores between corresponding bins of histogram variables. We also suggest the use of the estimator proposed by Olkin and Pratt. In the first approach, histogram variables must have the same number of bins. The second proposed approach, by contrast, extends the previous proposed correlations by considering the corresponding quantiles. This second approach can be used when histograms do not have the same number of bins.

About Wiley Online Library Help & Support Opportunities Connect with Wiley 

