

Statistics 579

Applied Multivariate Methods

Exam 1

The HATCO data set gives data from a survey of customers of the HATCO company. The data set consists of 100 observations on 14 separate variables, and is described in a separate document. For this exam, use only variables X1-X7, corresponding to perceptions of HATCO.

1. Determine if there are any major outliers in these data (but do not eliminate any observations in any of the subsequent analyses).
2. Determine if the 7 variables (corresponding to perceptions of HATCO) in this data set can be regarded as arising from a multivariate normal distribution.
3. Produce both a correlation matrix and a Scatterplot matrix of the 7 variables. Interpret.
4. Using Fisher's method with a bias adjustment, find a 95% confidence interval for the true population correlation between variable X2 (Price Level) and variable X5 (Overall Service).
5. Test whether the 7 variables X1-X7 can be regarded as statistically independent.
6. Perform a Principal Components Analysis using all 7 variables, and determine the actual dimensionality of the space in which the data fall (i.e., determine the number of principal components that would adequately describe these data).
7. Perform a Factor Analysis using all 7 variables. After determining the number of factors needed, compare the results produced from using both the iterated Principal Factor Method and the Maximum Likelihood method. Use a Varimax Rotation with both methods. Specifically, which variables load on which factors, and do the 2 methods produce the same results? Also, provide an interpretation of each factor that you identify. If necessary, retain any trivial factors.