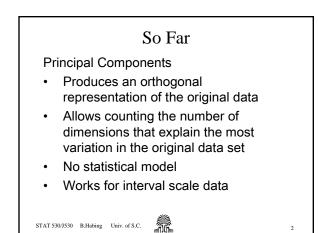


Department of Statistics LeConte 203 Telephone: 803-777-3578 E-mail: habing@stat.sc.edu

STAT 530/J530 B.Habing Univ. of S.C. M



So Far

2

Factor Analysis

- Fits a model to explain the observed ٠ variables in terms of underlying latent factors
- Lots of options ٠
- Is a linear model that requires interval ٠ data
- Some procedures require multivariate ٠ normality



So Far

Multidimensional Scaling

- Provides a graphical display of high dimensional data in fewer dimensions
- Can work with any kind of data if the appropriate distance measure is used
- Classical method is equivalent to principal components
- Non-metric methods allow for presenting more groups in fewer dimensions and better focus on small distances

STAT 530/J530 B.Habing Univ. of S.C.

So Far

Cluster Analysis

- Provides a graphical display of high dimensional data by producing a dendogram of various clusters.
- Can work with any kind of data if the appropriate distance measure and linkage are used
- No definitive way of choosing the right number of clusters.

STAT 530/J530 B.Habing Univ. of S.C.

So Far

5

MANOVA

- Provides a test of the hypothesis that several populations have the same mean vector
- Assumes multivariate normality, equal covariances, and independence
- This often isn't the actual question of interest

So Far

Discriminant Analysis

- Finds the linear combinations of variables that best distinguishes between the groups of interest
- Works with any interval data
- Is the optimal procedure if the data is multivariate normal with equal covariances
- Provides posterior probabilities assuming the equal covariance normality

M

STAT 530/J530 B.Habing Univ. of S.C.

So Far

Logistic Regression

- Predicts group membership from interval scale variables
- If the logistic curve is appropriate it provides a test of the hypothesis similar to MANOVA
- If the logistic curve is appropriate it provides a predicted group membership
- Does not require multivariate normality

STAT 530/J530 B.Habing Univ. of S.C.

Next

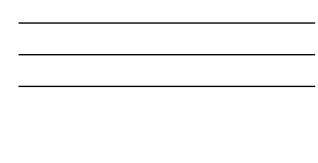
8

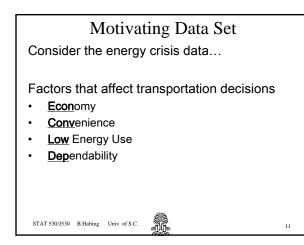
Today: Overview of multiple regression and introduction to canonical correlation analysis

Thursday 17th: Homework 8 is due, canonical correlation analysis continued

Tuesday 22nd: Brief homework 9 is due, final exam is posted, begin optional topic and course evals

Next
Thursday 24 th : Thanksgiving – No Class
Tuesday 29 th : Optional topic continued
Thursday 1 st : Homework 10 is due, ice cream field trip as penance for Homework 6 grade being late! With time for questions while we eat.
5:30pm Tuesday, December 6 th – Final Exam is Due
STAT 530/J530 B.Habing Univ. of S.C. 10





Motivating Data Set

Views on the Energy Crisis

- Q1 If the energy shortage gets any worse, the country will be in bad shape.
- Q4 Saving energy requires you to make major sacrifices.
- Q6 Utility companies should be allowed to burn cheaper fuel even though this would cause more pollution.
- Q9 Rationing of energy resources will be necessary for at least the next five years.
- Q10 Conserving electricity will save me money in the long run. STAT SSUSSO BHAME TUNN of S.C.

4

Motivating Data Set

- Q13 There is not much an average citizen can do to save electricity.
- Q17 We should forget about reducing pollution until our energy problems are solved.
- Q18 My personal conservation efforts have little impact on total consumption of energy.
- Q19 Because of the abundance of coal, industries should be encouraged to switch to coal as a fuel despite the air pollution it causes.

13



