

Exam Variant 1

1. Problem

Peter tries to find something to watch on a TV. He has already switched 70 channel and noticed that 8 of them were showing reality shows, 9 were showing old films and the rest were showing politics.

Calculate the likelihood ratio statistic to test the hypothesis that politics are twice as likely to appear on TV as reality shows and three times more likely than films.

2. Problem

Find two clusters of retail companies using complete linkage hierarchical clustering and the following distance matrix:

$$D_{i,j} = \begin{bmatrix} 0 & & & & \\ 6.3 & 0 & & & \\ 6 & 7.2 & 0 & & \\ 2.8 & 4 & 4.5 & 0 & \\ 23 & 28.3 & 28.2 & 25.8 & 0 \end{bmatrix}$$

where $i, j = 1, \dots, 5$.

Which observations will be in a smaller cluster?

3. Problem

Wonka's chocolate factory made questionnaire for its customers to determine their tastes. They collected data for three features (color, size and sweetness preferences) with mean vector is $\begin{pmatrix} 2.8 \\ 5.2 \\ 7.9 \end{pmatrix}$. All standard deviations are equal to 1. The eigenvector of the correlation matrix that corresponds to a principal component is $\begin{pmatrix} -0.316 \\ 0.707 \\ 0.632 \end{pmatrix}$. Use PCA algorithm to transform the feature vector $\begin{pmatrix} 2.7 \\ 6.6 \\ 8.8 \end{pmatrix}$ for Johnny into the given principal component.

4. Problem

We want to determine how *age* and *income* influence the probability to buy White Monster energy drink. The estimated logit regression is:

$$\hat{P}(y_i = 1) = \Lambda(-2 - 0.013 \cdot \text{age}_i - 0.069 \cdot \text{income}_i).$$

where y – binary variable equals to 1 if a person has bought White Monster, 0 otherwise. Find the marginal effect of an increase in age for 19 year old person with income of and $b = 3$ thousand dollars.

5. Problem

There are three groups of customers based on their share of spendings o luxury goods. The dataset consists of 50 observations on 3 different customer characteristics. The eigenvalues for discriminant functions are $\gamma_1 = 0.124$ and $\gamma_2 = 0.1415$. Calculate appropriate statistics to test if customers groups are significantly different from each other.

6. Problem

We want to check if gender is associated with how people make orders in “Tasty & Dot”.

	Male	Female
Orders delivery	22	91
Eats indoors	56	64

Using the data above calculate contingency coefficient.