

DS203 Assignment4

-Kalp Vyas

Q1. Python data types -> based on numerical data types like integer and float and non numerical data types as objects

Datatypes done in class -> numerical data types can be classified as integer, quantised and continuous data types and non numerical ones also can be nominal (randomly named/labeled) or ordinal (variables in a specific order) which makes these kind of data types more useful than python datatypes in statistics.

- Q2
- a) Descriptive
 - b) Predictive
 - c) Predictive
 - d) Exploratory
 - e) Descriptive
 - f) Descriptive
 - g) Exploratory

Q3 a) Dataset found:

https://www.researchgate.net/publication/275098614_Explaining_the_Rural-Urban_Gap_in_Infant_Mortality_in_India

Exploratory: Extracting information about the rural and urban morality rates in India and the factors like Gini coefficient and dispersion measure to understand the difference in the rural and urban areas.

Descriptive: The odds of neonatal death were lower for neonates born to mothers with secondary level education (O R = 0.60, p = 0.01) compared to those born to illiterate mothers. The odds of neon death were lower for infants born to unemployed mothers(O R =0.89 p=0.00) compared to those whose worked as agricultural worker/farmer/laborer.

Predictive: Predicting 5 factors which affect the IMR and change in neonatal mortality in urban and rural areas.

Prescriptive: Ensuring the consumption of an adequate quantity of Tetanus Toxoid (TT) injections by pregnant mothers, targeting vulnerable groups like young, first time and Scheduled Caste mothers, and improving overall household environment by increasing access to improved toilets, electricity, and pukka houses could also contribute to further reductions in neonatal mortality in rural India.

b) Dataset found:
<https://www.kaggle.com/atrexler/american-bellwethers-svm-analysis/data>

Exploratory: Getting datasets related to each sector, which can be used to predict best bellwethers.

Descriptive: Normalize the number of samples used to create each model (1000 datapoints), the PUMAs with more samples performed better, though the correlation is clearly lower. This suggests that better sampling leads to better

datasets, which of course makes plenty of sense.

Predictive: Use each SVM model to fit a subset of other PUMAs and predict occupations. Calculate mean scores to evaluate how effective the model was, then rank PUMAs by their mean scores to evaluate how well each PUMA predicts behaviour (occupations) in the other PUMAs.

Prescriptive: Average age and income were also slightly correlated with average PUMA predictor. I suspect the reason for this is that the average incomes and ages in the better predictor PUMAs cluster with the national averages of these values

c) Dataset found:

<https://data.gov.sa/Data/en/dataset/the-development-of-road-construction-2003>

Exploratory: Extracting data about the areas affected due to the various factors of construction of road.

Descriptive: Not applicable

Predictive: Predicting effects of construction of the roads in the places like health of people living near the area, socio-economic factors etc.

Prescriptive: If a systematic approach is planned it become easier to arrange approvals under Environment Acts and procedures. Now detail checklists and details are available on website of MOEF and other law enforcement authorities. These details and formats can be downloaded and application can be made either on line or on hard copies to competent authorities and approvals can be arranged in a fixed time frame.