types of data science questions

emilypole

21/06/2020

Types of data science questions/analyses

- Descriptive
- Exploratory
- Inferential
- Predictive
- Causal
- Mechanistic

##descriptive Summarising/describing a set of data. usually the first bite of a proj with a new dataset. examples incl measuring central tendency (mean, median, mode) or variability (range, SD, variance).

these are for summarising the data, not for generalising and making conclusions about what the data can tell you.

##exploratory For uncovering relationships in the data (correlation not causation). useful for formulating hypotheses and deciding what to do next

#inferential using a small sample of data to *infer* something about the population at large. extrapolation, statistical modelling etc. falls down whn the sample isnt representative of the population (obv)

> i think this is like if you know you have a normal distribution, then knowing the mean of the sample means you know the mean of the populaton

#predicitve using current data to make predictions about future data. like exploratory

> as well as, presumably inferential?

, depdends a lot on the variables you pick, and is correlative not causitive.

#causal manipulating one variable to see what happens to another. gives you a causal relationship. depends on the assumptions underlying the analysis and is v easy to get wrong. data is usually analysed in aggregate, an average causal effect (paracetamol helps headaches) might not be present in every single case in the analysis .

#mechanistic uncommon in data science. the goal is to understand the exact change in one variable that leads to an exact change in another. used in physics/engineering, usually too noisy for bio stuff.

> not sure what exact change means here precisely