

Question 1

We take a random sample of individuals in a population and identify whether they smoke and if they have cancer. We observe that there is a strong relationship between whether a person in the sample smoked or not and whether they have lung cancer. We claim that the smoking is related to lung cancer in the larger population. We explain we think that the reason for this relationship is because cigarette smoke contains known carcinogens such as arsenic and benzene, which make cells in the lungs become cancerous.

- This is an example of a mechanistic data analysis.
- This is an example of a predictive data analysis.
- This is an example of a causal data analysis.
- This is an example of an inferential data analysis.

Question 2

What is the most important thing in Data Science?

- The question you are trying to answer.
- Hacking skills.
- Machine learning and prediction.
- The data.

Question 3

If the goal of a study was to relate Martha Stewart Living Subscribers to Our Site's Users based on the number of people that lived in each region of the US, what would be the potential problem?

datascitoolbox%2Fheatmap.png

Link to image Source: <http://xkcd.com/1138/>

- There would be confounding because the number of people that live in an area is related to both Martha Stewart Living Subscribers and Our Site's Users.
- We would be data dredging because we collected data on thousands of places in the United States.
- We would be performing inference on the relationship between Martha Stewart Living Subscribers and Our Site's Users.
- We wouldn't know the sensitivity of our predictions.

Question 4

What is an experimental design tool that can be used to address variables that may be confounders at the design phase of an experiment?

- Randomization
- Taking a convenience sample.
- Data cleaning.
- Only using non-confounding variables.

Question 5

What is the reason behind the explosion of interest in big data?

- We have better experimental design now than previously.
- We recently discovered ways to use data to answer scientific and business questions.
- We recently discovered ways to use data to make predictions.
- The price and difficulty of collecting and storing data has dramatically dropped.