Abstract: Today, almost every action we take generates data: mobile communications, participation in social networks, web searches, business transactions, physical activity or sleep monitoring, clinical tests... the list is long and growing rapidly.

What information can be extracted from this data? Can we use it to predict behaviors or other variables? The key to integrating so many sources of information is often to find smaller underlying structures to be analyzed, without losing relevant information.

Mathematical statistics tries to give a rigorous answer by appealing to the concept of sufficiency. In this talk we will give an introduction to the topic, we will briefly review the recent contributions and we will discuss some of the problems that still remain to be solved.