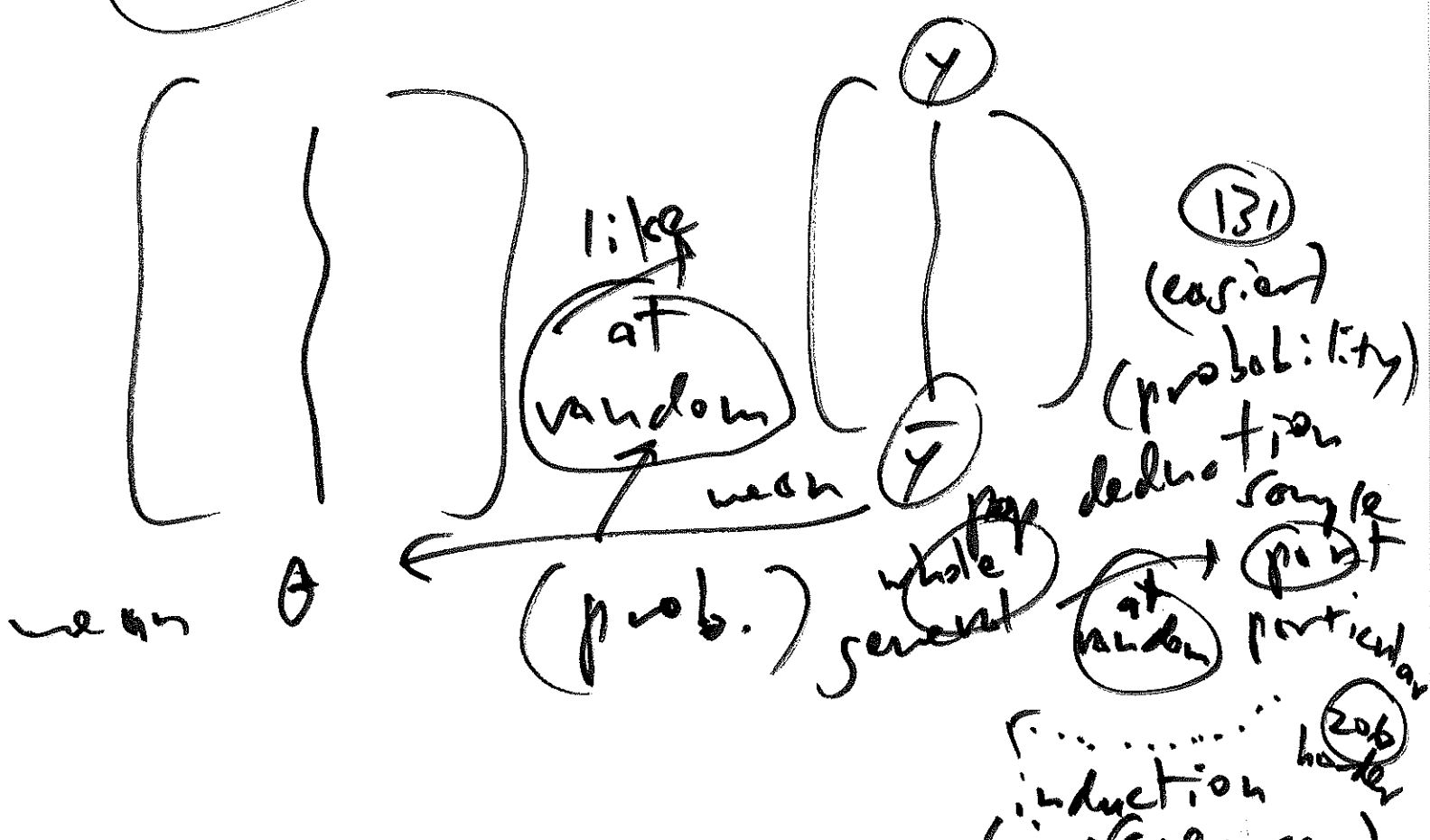
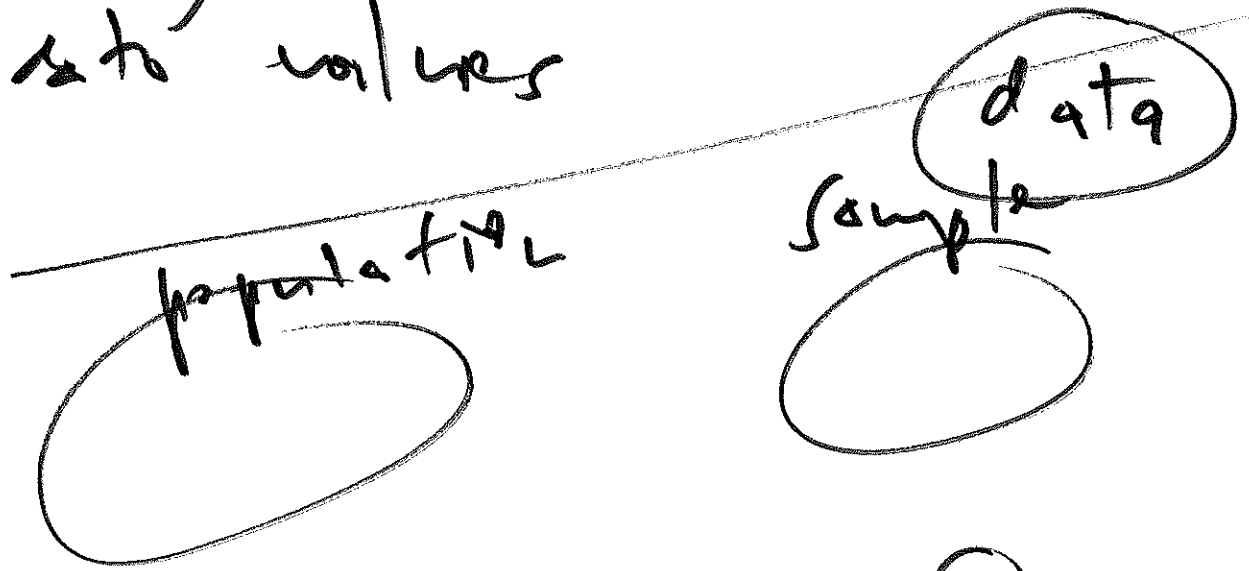




(missing values) | data  
(normalize) | analyzing

getting to know your  
data values



# 5<sup>th</sup> statistical data science <sup>(3)</sup>

activity:

optimal data -  
gathering  
(decisions)

experimental  
design

sampling

(A/B tests)

~ 1925 (Fisher)

Blood test truth  
 $P(\oplus \text{ and HIV}+) =$

$$P(\text{HIV}+) \cdot P(\oplus | \text{HIV}+)$$