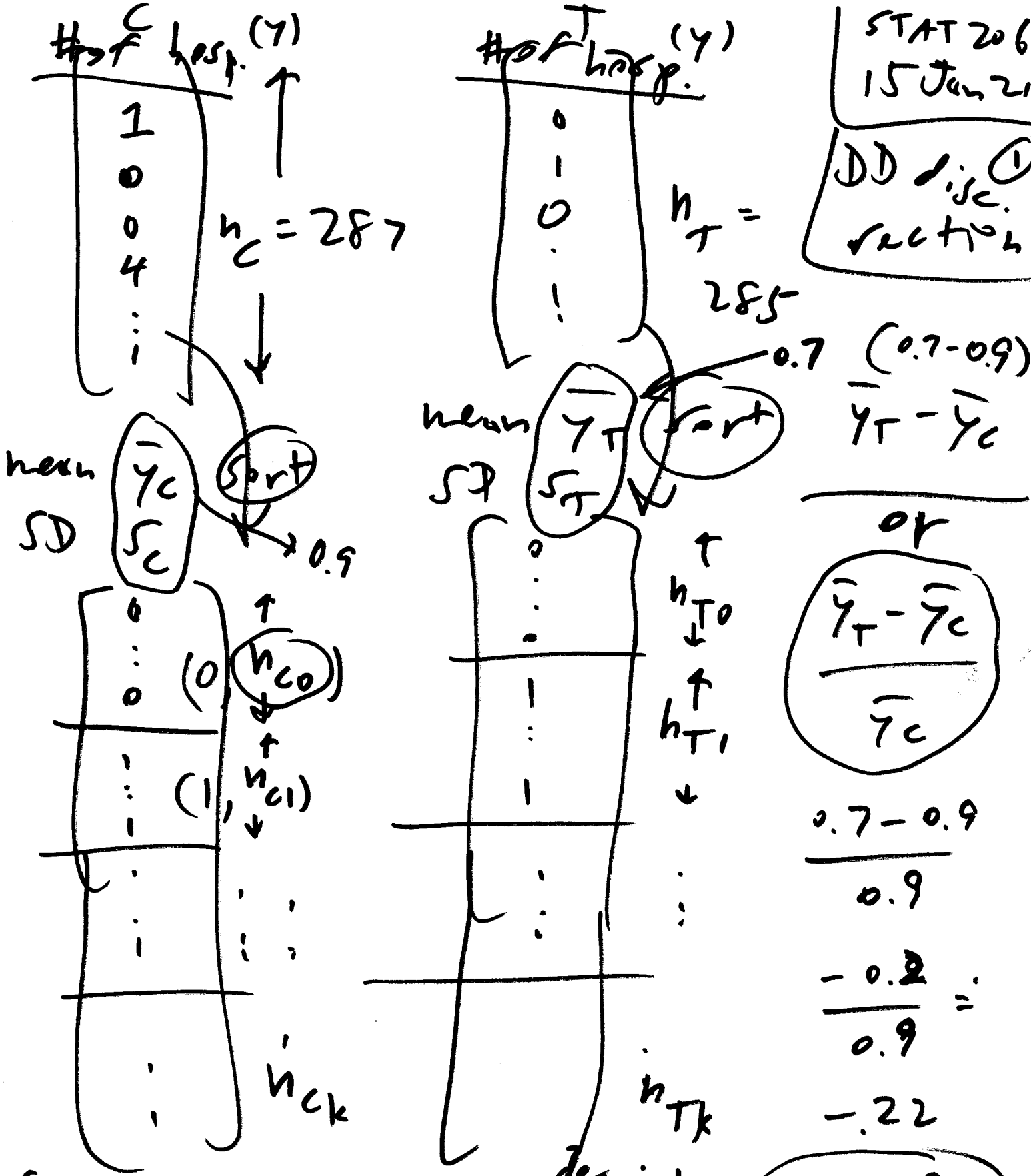
STAT 206
15 Jan 21

DD disc. ①
reaction



(0.7-0.9)

$\bar{y}_T - \bar{y}_C$

or

$$\frac{\bar{y}_T - \bar{y}_C}{\bar{y}_C}$$

$$\frac{0.7 - 0.9}{0.9}$$

$$\frac{-0.2}{0.9} =$$

$-.22$

= 22% decline

description

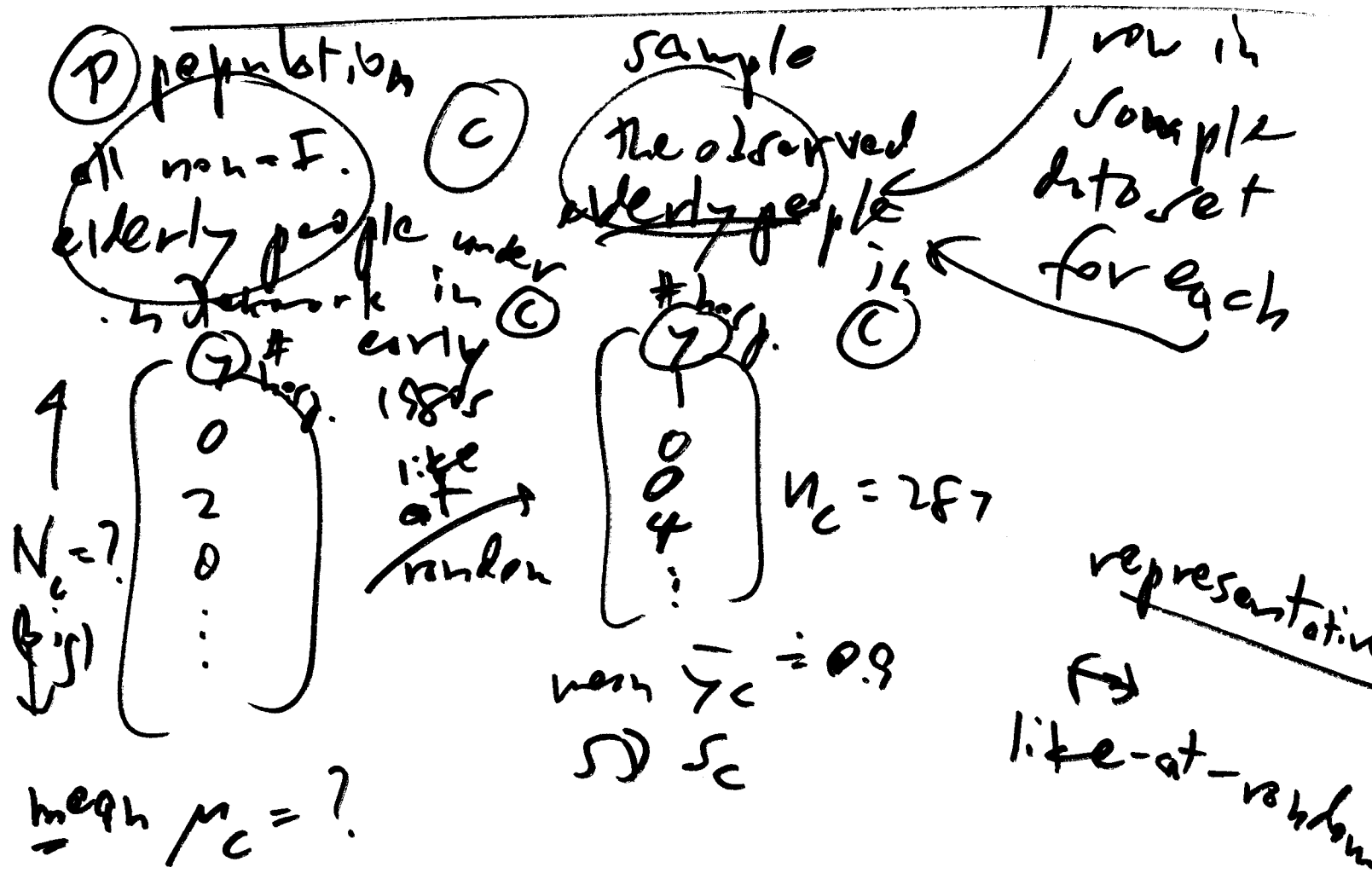
practically significant

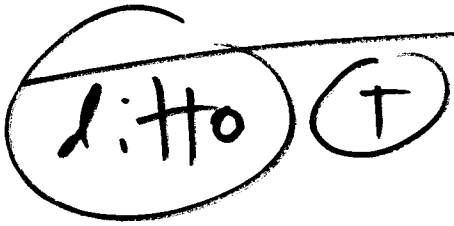
large in practical terms! yes

(k positive integer)

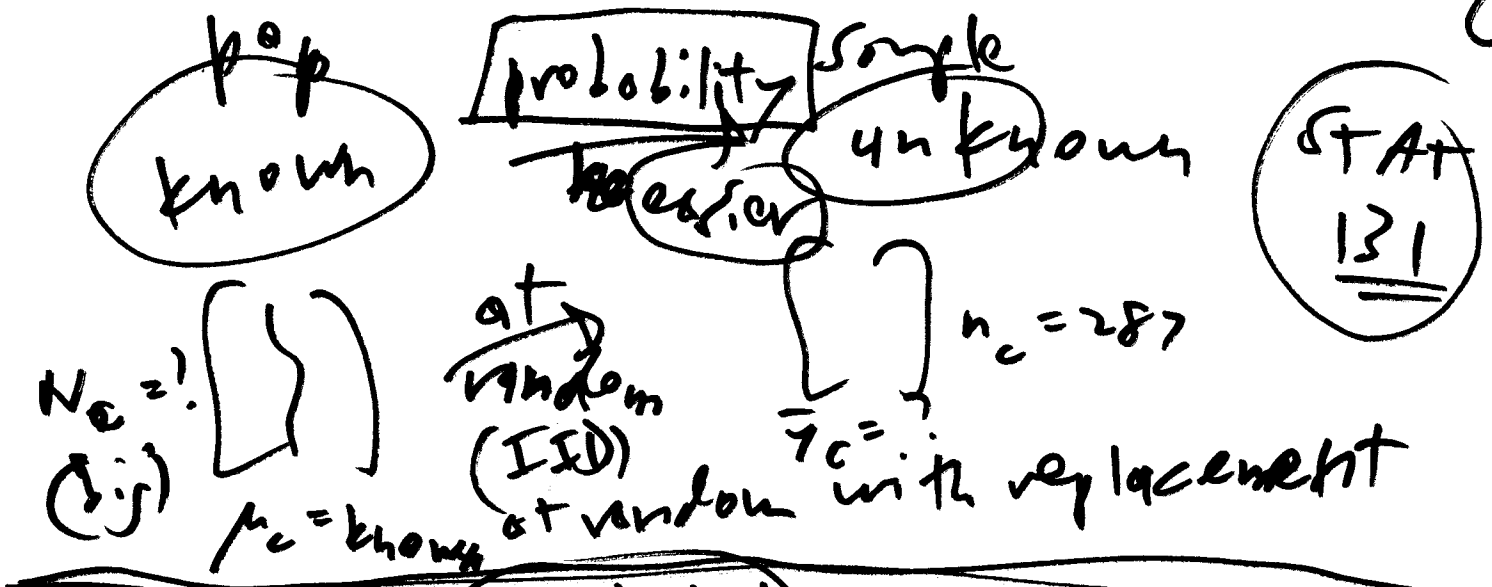
benefit

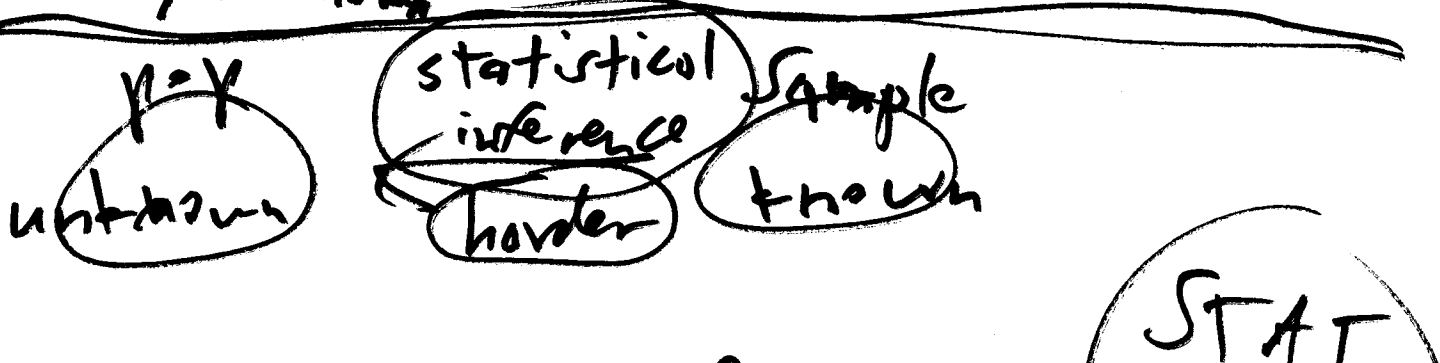
- statistical inference (generalizing outward from our data to a larger & more interesting world)





$\mu_T = ?$ $\quad \bar{y}_T = 0.7$





activities

(7) statistical data science

(1) probability calculations

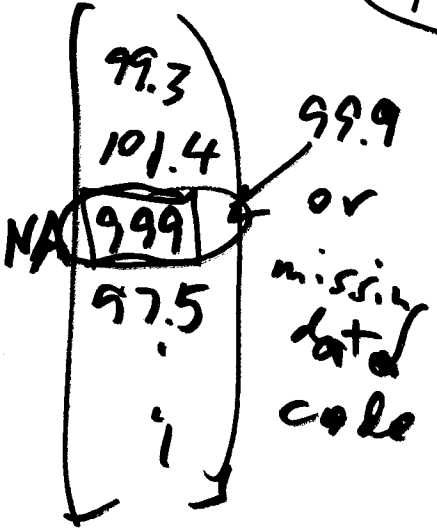
(2) design of data-gathering activities

(A) sampling from pop. (no interventions) (B) designed experiments (interventions)

(3) data curation (of existing data sets) ④

"data cleaning"

- coping with missing data ① can be 75% of total data science time
- sanity checks on all the variables (e.g. x_3 = temp. in °F at hospital admission



~~fight~~ garbage in ⇒ garbage out

here we're enriching the problem context

matches the data

~~min max mean~~ ③

(4) graphical & numerical descriptions of existing data sets ①

histograms, scatterplots, time series plots

(4) (5) statistical inference (generalizing out-of-sample accurately from existing data sets: unknown θ (parameters) Bayes

(6) statistical prediction of new data using D^* = unknown data Bayes

(7) decision-making (Bayesian decision theory or Bayesian game theory) ← 1 decision-maker

> 1 decision-makers