

this time: Samples & Populations ; histograms

next time: Measures of Center & Spread

* sections start next Mon

- statistics: the study of uncertainty
- uncertainty: state of incomplete information

$\theta = p$

↳ percentage that we don't know exactly

principle: to decrease your uncertainty (or not completely known) about something unknown to you, gather new good data

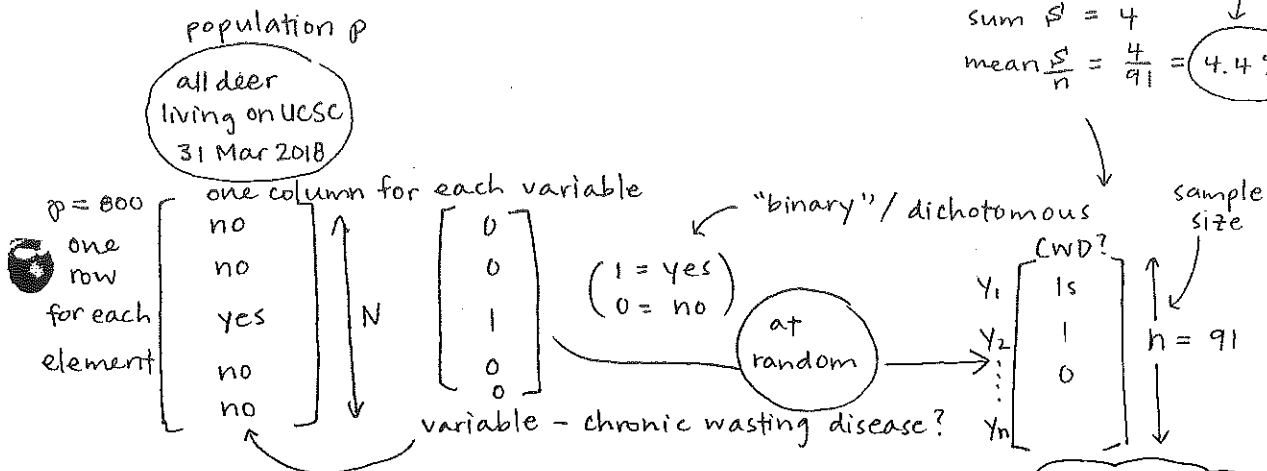
The set

p = the deer who live on UCSC campus as of 31 Mar 2018

↳ population: collection of subjects / elements

$\bar{y}, \hat{\theta}, \hat{p}$
estimate

sum $\beta = 4$
mean $\frac{\beta}{n} = \frac{4}{91} = 4.4\%$



β (sum) = # of deer w/ chronic wasting disease

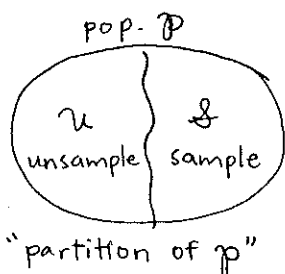
$\frac{\beta}{n}$ (mean) = ? = $\theta = p = ?$
parameter

- graphical & numerical summaries of data sets: descriptive statistics
- parameter: numerical summary of a population
- * not enough time & money for complete census

↳ subset \mathcal{S} of p

* goal of sampling: try to make \mathcal{S} and \mathcal{U} as similar as possible in all relevant ways.

- to achieve this goal, choose \mathcal{S} at random



2 simple random sampling methods:

① at random with replacement
(independent identically distributed (IID))

* ② at random without replacement
(simple random sample (SRS))

→ SRS is more informative than IID, but IID has easier math

when $n \ll N$ a lot smaller than

(sample size) (population size)

$SRS = IID$

- randomization can't guarantee perfect similarity in all relevant ways between \mathcal{S} & \mathcal{U} every time

① the bigger n gets, the more likely that \mathcal{S} & \mathcal{U} are relevantly similar

② we will learn methods to estimate how often randomization yields bad samples

← unrepresentative \mathcal{S} of \mathcal{U}

- since sampling was at random, $\hat{\theta} = \hat{p} = \bar{y} = \frac{\sum}{n} = 4.4\%$ is a good estimate of $\theta = p = \frac{\sum}{N}$ — how good?

- we think θ is around 4.4%, give or take ? %

variables & the values they take on

| variable | possible values |
|-------------------------|--|
| eye color | brown, blue 0 1 ← dichotomous |
| hair color | brown, black, red, white |
| success in maze-running | very slow slow moderate fast very fast |

circled text: categorical non-numerical qualitative

↳ its values do not have numerical value