STAT 7: Statistical Methods for the Biological, Environmental, and Health Sciences

David Draper ⟶ Office Hours: Tue 2-3:30pm
Thurs 2-3:30pm   Jacks Lounge

Office: E2 539B (hardly ever there)
Tele: 459-1295
email: draper@ucsc.edu ⟩ best way to reach
  ↳ subject line: STAT 7 student, Fall 2019

Lab ⟶ mandetory for only the first week
entirely online
added a final project
  ↳ show that you know how to use jump

rgutie17@ucsc.edu
Go to Rene Gutierrez for grade fix on assignment/test
                        ↑
              any questions w/ grades

Take Home midterm and Final   There is MSI
⏜
during this
time, he will have
a office hour everyday,
including weekend

Discussion sections (mandetory)
- content in discussion sections identical all week long
- can go to any section if not able to go to the one
      your enrolled in   9 section
- 55 min: cover problems   9 quizzes
- 10 min: easy quiz that you can take w/you
      upload as pdf on canvas

Quizzes due Tues   No discussion sections for entire
                          Thanksgiving week

there is a survey at the end of course
    80% or more students do survey = plus bump
                                        for all

HW (25%.) → 4 times HW
    drops lowest HW score

Midterm (25%) → end of fifth week

Discussion Sections (20%.) → each quiz ~ 2% of grade
    drops lowest quiz
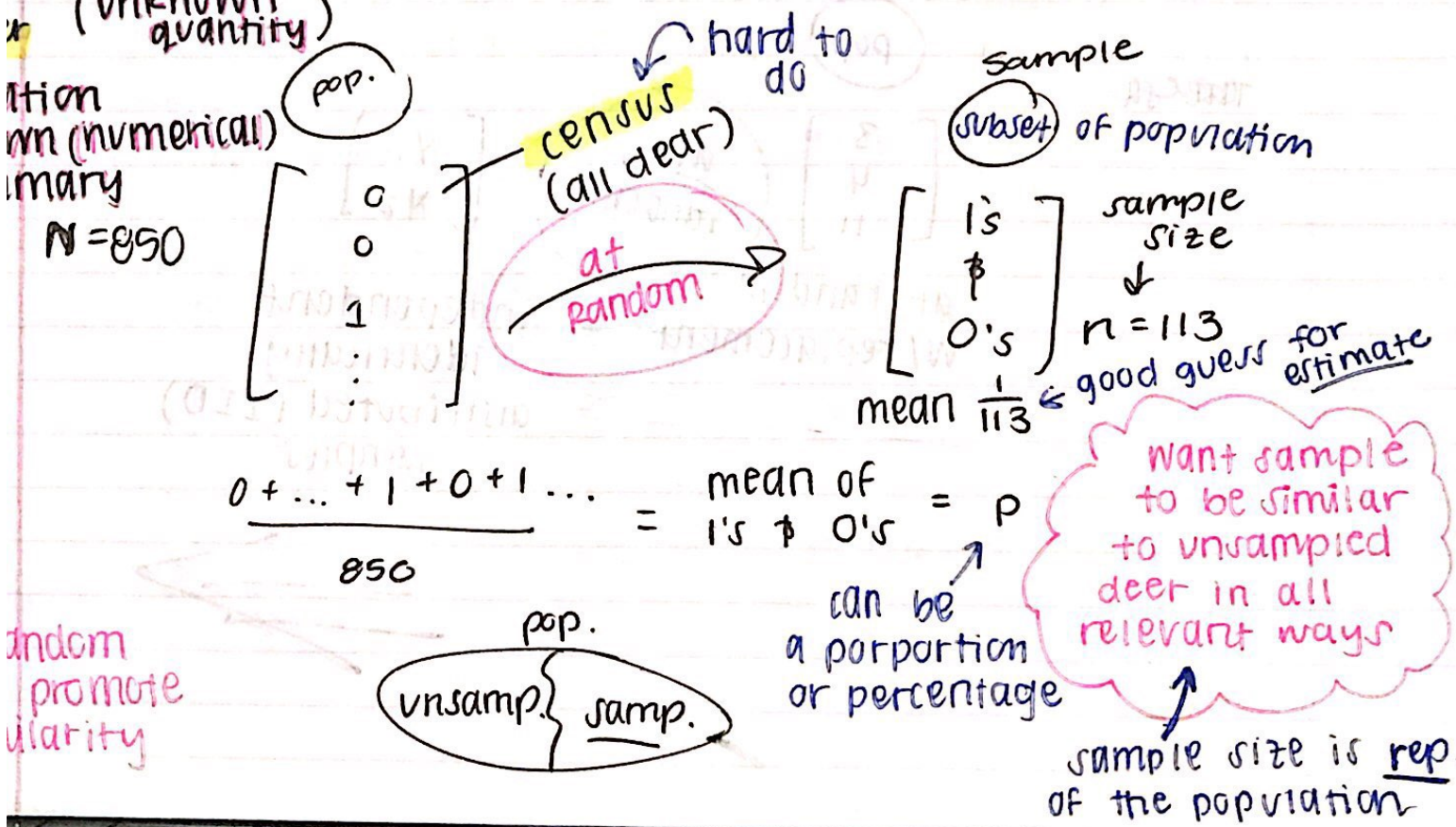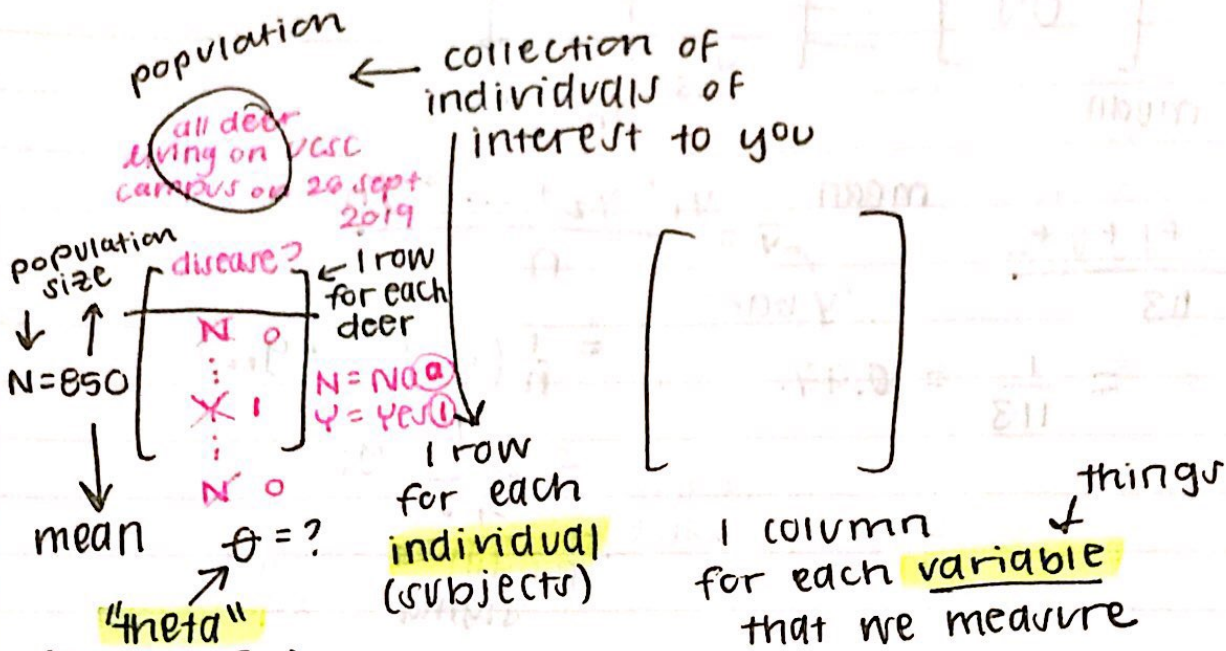
Final Exam
    (30%)

    Everything uploaded w/ canvas
            pdf files

General cotent

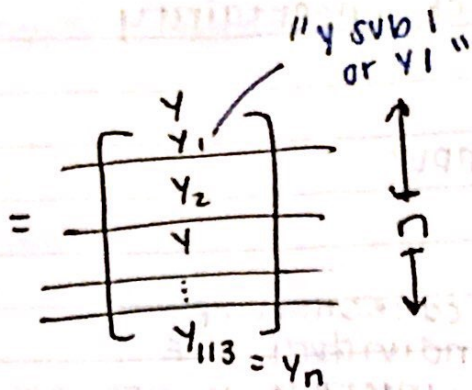Statistics : study of uncertainty

850 deers on campus   — terms

popvlation ← collection of
individvals of
(all deer living on UCSC campus on 26 sept 2019)   interest to you

population size   disease?   ← 1 row for each deer
↓   ↑
N=850
$$\begin{bmatrix} N & 0 \\ \vdots \\ X & 1 \\ \vdots \\ N & 0 \end{bmatrix}$$
N = No $\textcircled{0}$
Y = Yes $\textcircled{1}$
1 row for each individual (subjects)

$$\begin{bmatrix} \\ \\ \\ \end{bmatrix}$$
1 column for each variable that we measure   ← things

mean $\theta = ?$
"theta" (vnknown quantity)

pop. (numerical) summary
N =850
pop.
$$\begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \end{bmatrix}$$
census (all dear) ← hard to do
at random →

sample
(subset) of population

$$\begin{bmatrix} 1's \\ \$ \\ 0's \end{bmatrix}$$
sample size ↓
n = 113 ← good guess for estimate
mean $\frac{1}{113}$

$$\frac{0 + \ldots + 1 + 0 + 1 \ldots}{850} = \text{mean of 1's \& 0's} = P$$

can be a porportion or percentage

want sample to be similar to vnsampled deer in all relevant ways

pop.
(vnsamp. samp.)

sample size is **rep.** of the population

random promote ularity

sample

disease

$$\begin{bmatrix} 1's \\ \& \\ 0's \end{bmatrix} = \begin{bmatrix} Y \\ Y_1 \\ Y_2 \\ Y \\ \vdots \\ Y_{113} = Y_n \end{bmatrix} \Bigg\updownarrow n$$

"y sub 1 or Y1"

mean

$$\frac{0+0+...+1+0+..}{113}$$

$$= \frac{1}{113} = 0.9\%$$

mean

$$\bar{y} = \frac{Y_1 + Y_2 + ... + Y_n}{n}$$

"y bar"

$$= \frac{1}{n}(y_1 + ... + y_n)$$

$$= \frac{1}{n}\sum_{i=1}^{n} y_i$$

largest value

capital sigma

smallest value

pop

TEREESA

$$\begin{bmatrix} 3 \\ 4 \\ 11 \end{bmatrix} \xrightarrow[\text{random}]{at} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$$

at random
w/ replacement = independent
identically
distributed (IID)
samples

STATS 10/1
This variable types

- HW due friday Oct 11
- do survey
- quiz 1 due 1 week from today

1.1 introduction

statistics → study of uncertainty
uncertainty → state of incomplete or imperfect info
            about something of interest to you

method	pro	con
I I D	math is easier	less informative
SRS	more informative	math is harder

special cases
① $n=1 \longrightarrow$ IID = SRS
② $n=N$ $\nearrow$ ~~census~~ SRS no uncertainty left
          $\searrow$ IID still uncertainty

③ $n \underset{\text{is a lot smaller than}}{\boxed{<<}} \underset{200,000,000}{N} \longrightarrow \boxed{\text{IID}} = SRS$

population: a collection of subjects or elements or individuals of interest to me

dichotomous ~ yes, no

Binary coding ~ yes = 1
                no = 0

Parameter: numerical summary of a population

$\bar{\theta}$ "theta Hat"

An estimate $\bar{\theta}$ of a
    population parameter θ
is also sometimes called a statistic

variable	possible values
eye color	(brown, blue) **qualitative variable** ~categorical~
(plant size) height	=6in 14 cm (2.54 cm/inch) **quantitative variable** ~numerical~
# leaves	(19)

~vous~ (continuous)

no unique place on # line for category 'brown' → qual

variable	possible values
maze running time	v. slow, slow, ..., v. fast qualitative

brown blue | blue brown
↑ ↗ no diff

very slow   slow   moderate   fast

ordered categorical
not dichotomas
qual.