Nikhila Kadiyala

2 October 2018 - Tuesday

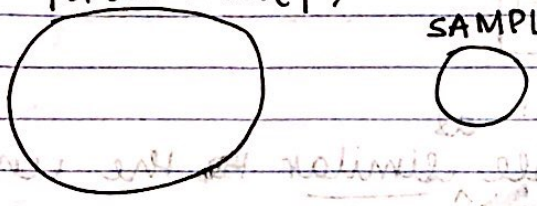
HW- next FRIDAY 10/12

DD office BF 3570

Variables

POPULATION (P)

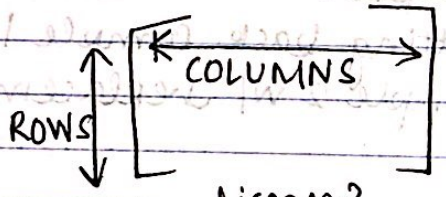
SAMPLE (S)



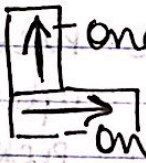
eg: POPULATION: All deer @ UCSC

DATA SET

Rows Columns (Matrix)



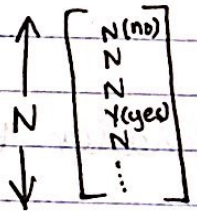
- General Data Set Structure



one row for each individual (SUBJECTS)

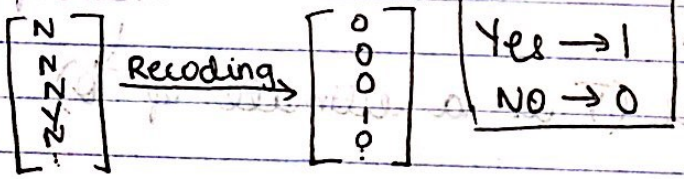
one column of for each variable VARIABLE: Measured

disease?



N ← population size ~800 deer

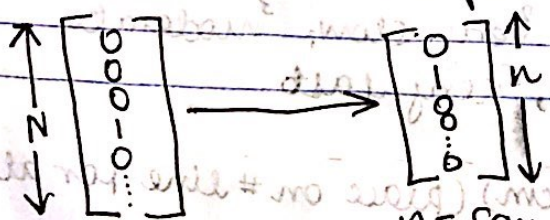
* Any variable that can take only 2 forms of possible values - DICHOTOMOUS BINARY



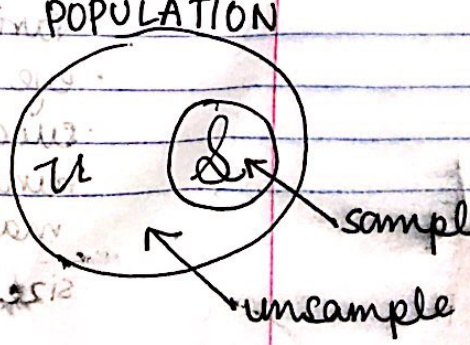
→ Why is this a good scale? Adding all #0s and 1s of the data set will give us the # of deer w/ the disease. (easy to keep track)

- MEAN = $\frac{\Sigma \text{ of all \# in data set}}{\text{How many \# there are}}$

SAMPLE: a subset of the population:



n = sample size



Mean of sample is a GOOD guess of how many deer have the disease in the population (GIVEN THE SAMPLE is GOOD)

$Y_1 \to 0$
$Y_2 \to 1$
$Y_3 \to 0$
$\vdots$
Y_{n-1}
Y_n

MEAN = $\bar{Y}$ ("Y bar")

GOAL IN SAMPLING: as
- Make the sample similar to the unsample in all RELEVANT ways.

as possib

SIMPLEST method of achieving the goal: CHOOSE sample at RANDOM.

(1) At Random W/ replacement: Pick sample 2 after putting back sample 1.
(IID - Independent)

(2) At random W/O replacement: pick sample 2 W/ replacement
(SRS - simple random sampling)

SRS is more informative than IID

IF $n=1$ ---> SRS = IID → IF n is a lot smaller than N ($n \ll N$) ---> SRS ≈ IID

Sampling method	SRS
Analysis "	IID

ESTIMATE: An educated guess ($\bar{Y}$ is a estimate of θ)

mean of population

POPULATION —(HOW CHOSEN?)→ SAMPLE

→ DATA TYPES

variable	Possible values
· eye colour	BLUE / BROWN (dichotomous)
· success in running a maze	1 slow, 2 very slow, 3 moderate 4 fast, 5 very fast
size of plant	Height (cm) (place on # line for all values) #leaves
growing temp.	78°F