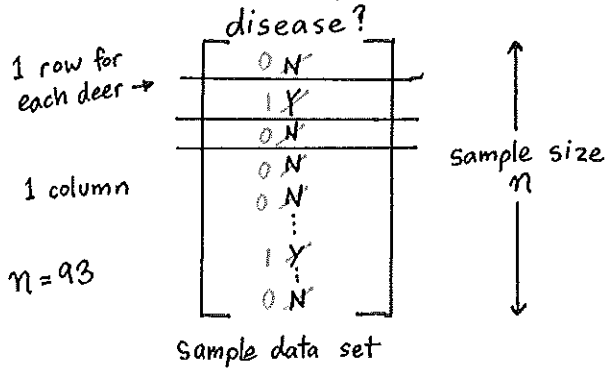
AMS 7

29 Sept. 15

This Time: Intro, types of variables; samples & populations

Next Time: numerical & graphical descriptive methods



Y = yes → 1

N = no → 0

Dichotomous

↳ Binary

↳ sum of 1s & 0s = # of dei with dis

Y = 4

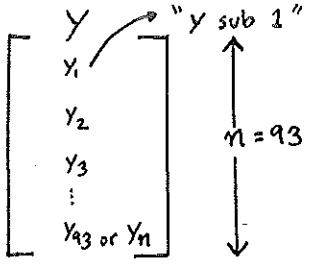
N = 89

N = 93 → 4/93 = 4.3%

1 row for each subject 1 column for each variable (thing we measure)

average = $\frac{\text{sum}}{n}$ = mean = % of the deer with disease

general data set: (with subscipt notation)



mean $\bar{Y}$ ← "Y bar"

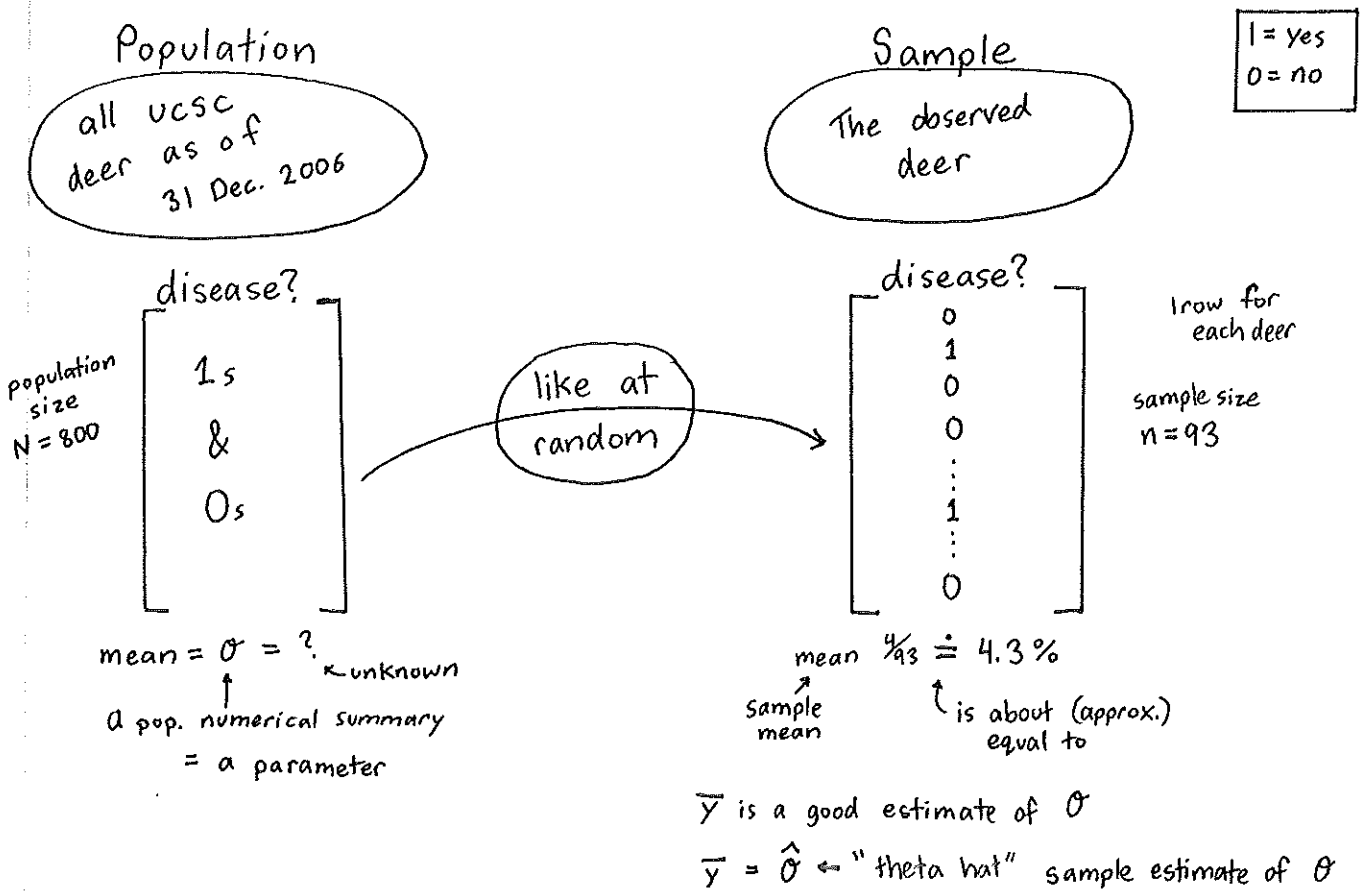
$$\text{mean} = \frac{Y_1 + Y_2 + \ldots Y_{93}}{93} = \frac{Y_1 + Y_2 + \ldots + Y_n}{n} = \frac{1}{n}\left(Y_1 + Y_2 + \ldots + Y_n\right)$$

capital Sigma → sum

$\sum_{i=1}^{n} Y_i$ ← index of summation

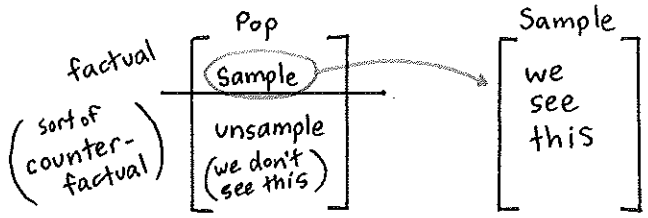
$$\begin{bmatrix} Y_1 - \bar{Y} \\ Y_2 - \bar{Y} \\ \vdots \\ Y_n - \bar{Y} \end{bmatrix} \qquad (Y_1 - \bar{Y}) + (Y_2 - \bar{Y}) + \ldots + (Y_n - \bar{Y}) = \sum_{i=1}^{N} (Y_i - \bar{Y})$$

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \qquad \text{mean } \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\left(Y_1 + Y_2 + \ldots + Y_n\right)$$



$\bar{y}$ is a good estimate of θ

$\bar{y} = \hat{\theta}$ ← "theta hat" sample estimate of θ

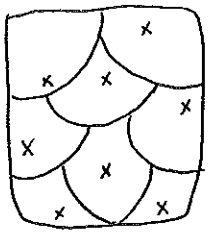


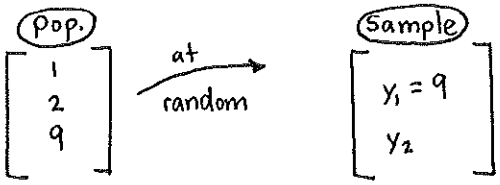
basic sampling principle = try to make sampled & unsampled subjects in pop. as <u>similar</u> as possible in all <u>relevant</u> ways.

How do this? Choose sample at random from pop.

campus:

geographic





after drawing 9, do we put it back in?

At random <u>with</u> replacement (put first draw back in):

<u>i</u>ndependent <u>i</u>dentically <u>d</u>istributed (IID) sampling

has easier math

At random <u>without</u> replacement:

<u>s</u>imple <u>r</u>andom <u>s</u>ampling (SRS)

more informative than IID

Variable		Possible Values	
eye color		blue (1), brown (0)	← dichotomous (binary) ← qualitative/ catagorical
hair color		black, brown, red, white	
plant	height (cm)	7.42, 8.20, ...	← quantitative/numerical (can be found on a number line)
	leaves		