

AMS 7

29 Sept. 15

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

Next Time: numerical & graphical descriptive methods

1 row for each deer →

1 column

$n = 93$

disease?

0	N
1	Y
0	N
0	N
⋮	⋮
1	Y
0	N

sample data set

sample size n

$y = \text{yes} \rightarrow 1$

$N = \text{no} \rightarrow 0$

Dichotomous
↳ Binary
↳ sum of 1s & 0s = # of deer with dis

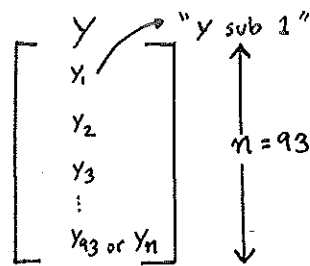
y = 4
N = 89
 $N = 93 \rightarrow \frac{4}{93} = 4.3\%$

1 row for each subject

1 column for each variable (thing we measure)

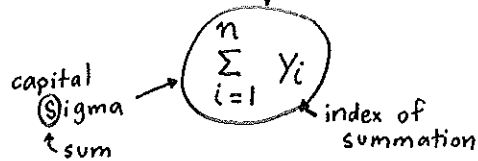
average = $\frac{\text{sum}}{n}$ = mean = % of the deer with disease

general data set:
(with subscript notation)



mean \bar{y} ← "y bar"

$$\text{mean} = \frac{y_1 + y_2 + \dots + y_{93}}{93} = \frac{y_1 + y_2 + \dots + y_n}{n} = \frac{1}{n} (y_1 + y_2 + \dots + y_n)$$



$$\begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix}$$

$$(y_1 - \bar{y}) + (y_2 - \bar{y}) + \dots + (y_n - \bar{y})$$

$$= \sum_{i=1}^n (y_i - \bar{y})$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$\text{mean } \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \rightarrow \frac{1}{n} (y_1 + y_2 + \dots + y_n)$$

Population
all UCSC deer as of 31 Dec. 2006

Sample
The observed deer

1 = yes
0 = no

disease?
[1s
&
0s]

Population size
N = 800

like at random

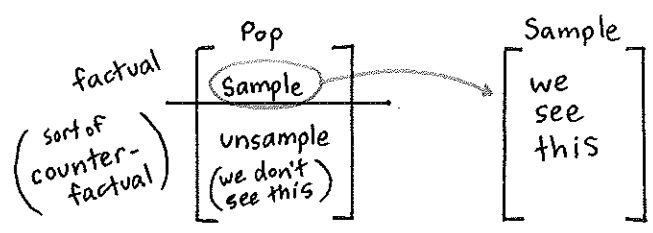
disease?
[0
1
0
0
⋮
1
⋮
0]

1 row for each deer
sample size
n = 93

mean = θ = ? unknown
↑
a pop. numerical summary
= a parameter

mean $\frac{4}{93} \approx 4.3\%$
↑
sample mean
↑ is about (approx.) equal to

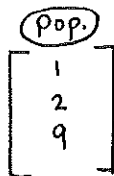
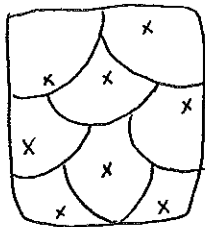
\bar{y} is a good estimate of θ
 $\bar{y} = \hat{\theta}$ ← "theta hat" sample estimate of θ



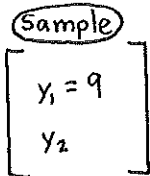
basic sampling principle = try to make sampled & unsampled subjects in pop. as similar as possible in all relevant ways.

How do this? Choose sample at random from pop.

campus :
geographic



at
random



after drawing q , do we put it back in?

At random with replacement (put first draw back in):
independent identically distributed (IID) sampling
has easier math

At random without replacement:
simple random sampling (SRS)
more informative than IID

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