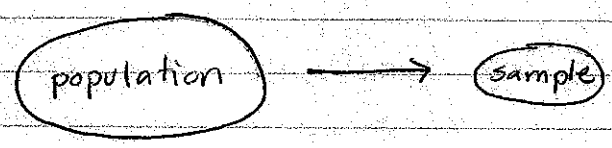


Jan. 8, 2009
1.1 Introduction

Deer w/ chronic wasting disease



data set:
1 column for each variable
1 row for each subject
n = sample size = 100

disease?
no
yes
no
no
1 row for each deer
n = 100

using 1's & 0's

disease?
0
1
0
0
1 = yes
0 = no

sum(s) = # of diseased deer
mean(\bar{y}) = s/n = proportion of diseased deer in population sample
* this is a good estimate of θ , but still has a measure of uncertainty

Population (all deer at UCSC in Dec. 31, 2006)

pop. size = $N = ?$
(maybe around 100 - 1,000)

disease?
1's
or
0's

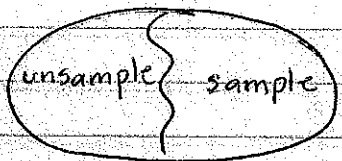
mean = $\theta = ?$
= p parameter

Sample (observed deer)

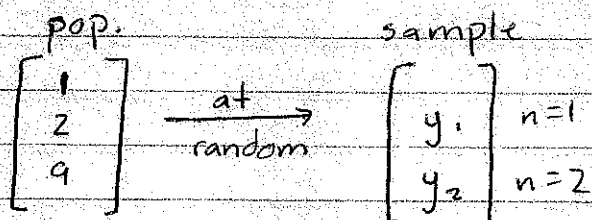
disease?
1's
or
0's

n = 100
mean $\bar{y} = 1/100 = \hat{p} = 1\%$
estimate

Jan. 8, 2009 (pg. 2)

population

- goal: make sample and unsample as similar as possible in all relevant ways
- simple method to achieve goal: choose sample elements at random from the population (all subjects have an equal chance of being chosen)



- 2 types of random sampling
 1. at random without replacement = simple random sampling (SRS)
 2. at random with replacement = independent identically distributed (IID)

- SRS is more informative than IID (so this is what people do), but the math is easier with IID

- when n is a lot smaller than N , SRS and IID are about the same

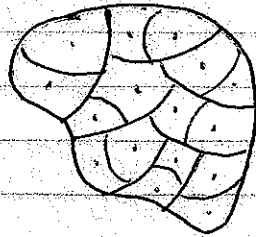
* people do SRS, but when $n \ll N$, they use math from IID

θ = parameter

$\hat{\theta}$ = estimate of θ from sample data

Jan. 8, 2009 (pg. 3)

- You could also partition the campus into n subsets and have n people get data on 1 deer in each subset, so it's like SRS



1.2 Data Types

variable	Values	
eye color in animal of study	blue, brown (dichotomous)	qualitative, nominal
success in maze-running	very slow, slow, moderate, fast, very fast	qualitative, ordinal (have ordering to them)
Size of a plant: 1. # of leaves	... 44, 45 ...	quantitative, discreet
2. Height	111.47 cm, etc.	quantitative, continuous, ratio (true zero has direct physical meaning)
growing temp. at which most buds were produced	25, 240° C	quantitative, discreet continuous interval (does not have true zero)

