

Quiz 1 due next Tuesday
 Homework 1 due Friday Oct 21

Stat 7 - Lecture 2 - October 1, 2019

Statistics: study of uncertainty, how to measure it and what to do about it

uncertainty: state of incomplete/imperfect information of something of interest

example: percentage θ of deer at UCSC who are diseased

total number of deer: $N = 850$

mean = $\theta = ?$

$N = 850$

0
0
1
0
1

→ at random

sample disease
 sample size
 $n = 113$

1s
&
0s

mean = $\frac{1}{113}$

$$\frac{0 + 0 + \dots + 0}{850} = \text{mean of } 1s = p = \theta$$

proportion

sample is representative
 * want sample to be similar to unsampled deer in all pop. in all relevant ways

y

sample disease

unsampled } sampled

* estimate

1s
&
0s

=

y_1
y_2
y_3
\vdots
y_{113}

= y_n

\uparrow
 $n = 113$
 \downarrow

mean .9% = $\frac{1}{113} = \frac{0 + 0 + 1 + 0 + 0}{113}$

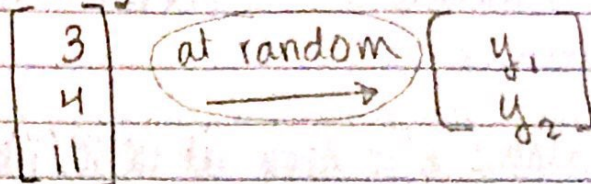
$$\bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n}$$

↳ $\frac{1}{n}(y_1 + \dots + y_n)$

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

$n \leftarrow$ largest value
 $i=1 \leftarrow$ smallest value

population



\leftarrow at random with replacement: independent, identically distributed (IID)

at random without replacement: simple random sampling (SRS)

method	Pro	Con
IID	math is easier (second draw doesn't depend on first)	less informative
SRS	more informative (no risk of getting the same sample more than once)	math is harder

Special Cases

1. $n=1 \rightarrow$ IID = SRS

2. $n=N \rightarrow$ SRS = no uncertainty

\rightarrow IID = still some uncertainty

3. $n \ll N \rightarrow$ IID = SRS

is a lot smaller than \nearrow

lex. $n \ll N$

1,000 \ll 200,000,000

no possibility of getting same subject twice)

Lecture notes page 12

example

variable	Possible values	
eye color	brown, blue	← qualitative (qual) ← categorical
(plant size): height	14 cm (~6 in)	} quantitative (quant) numerical
number of leaves	19	

- R-21: variable types

- no unique place on the number line for category 'brown' → qual

- 14 cm is continuous (possibility of height being 14.2)

- 19 leaves is discrete

example

variable	Possible values
maze running time	very slow, slow, ..., very fast

qualitative - ordered categorical - not dichotomous

v. slow slow moderate fast very fast

VS

nominal order - dichotomous

brown blue

blue brown