

this time:

variable types; numerical / descriptive summaries

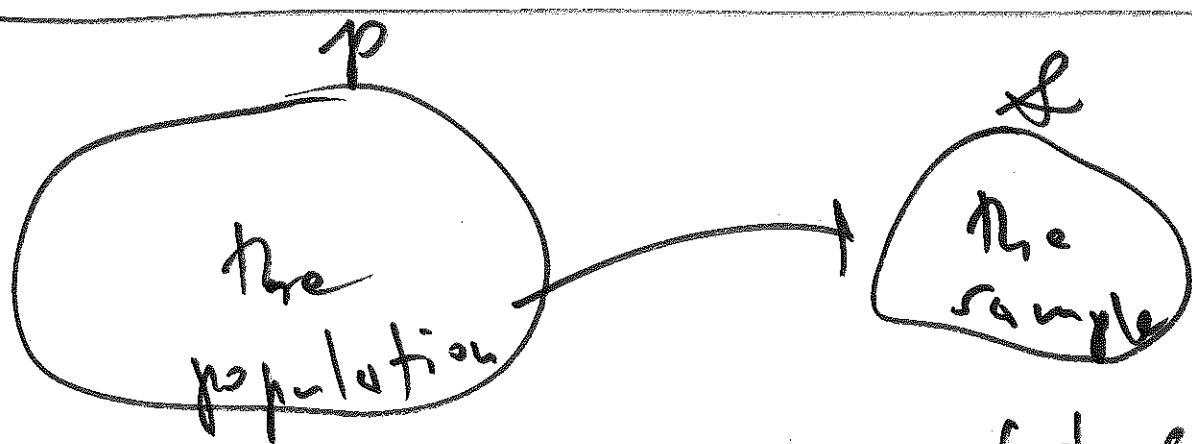
AM57  
22 Jun  
2016

next graphical time: descriptive summaries

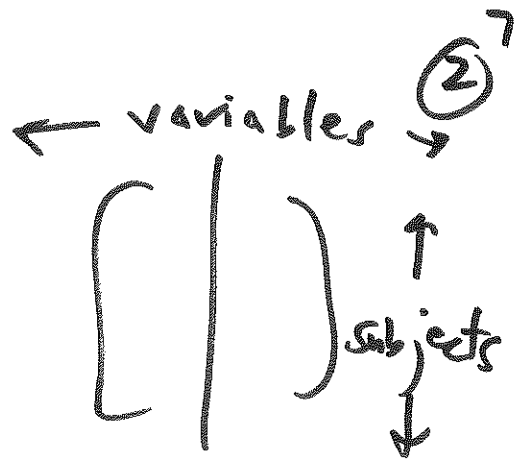
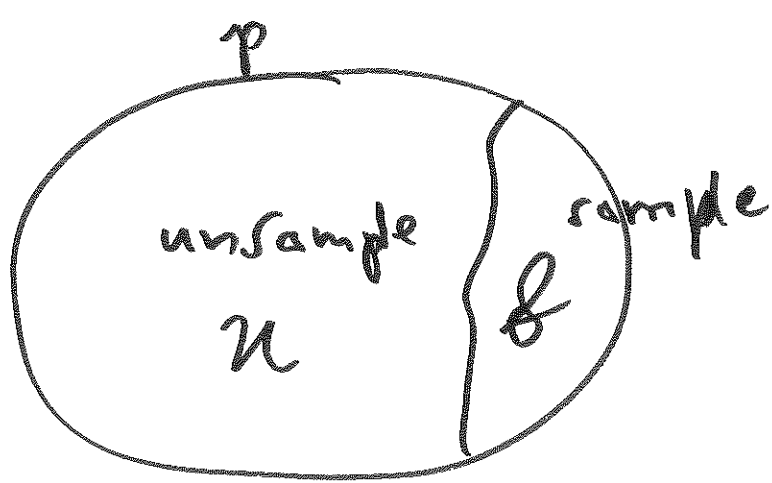
disc. sec. in 9AM TUE ①  
E2/194 starting tomorrow

temporary office hours:

DD	Th (to tomorrow)	10.30 - 11.30	BE 357C
GD	_____	12.30 - 1.30	BE <del>257C</del> 312CID
DD	F	11.30 - 12.30 2-3	BE 357C
GD	F	11.30 - 12.30	BE 312CID

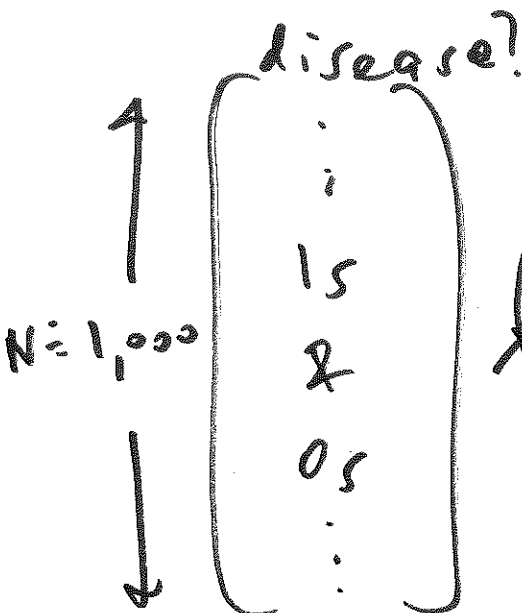


get data here

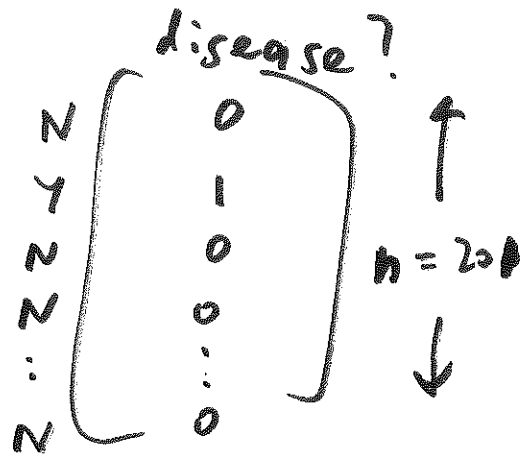


population  
 all users deer  
 at 31 Dec to 15

sample  
 the observed  
 deer



at  
 random



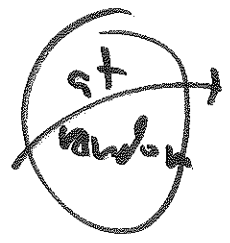
$$\text{mean } \hat{p} = \frac{3}{201} = 1.5\%$$

random sampling: Jerzy Neyman (1927)

gdinolov@soe.ucsc.edu

$N=3$   $\mu = \mu$

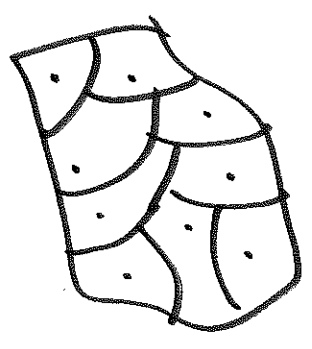
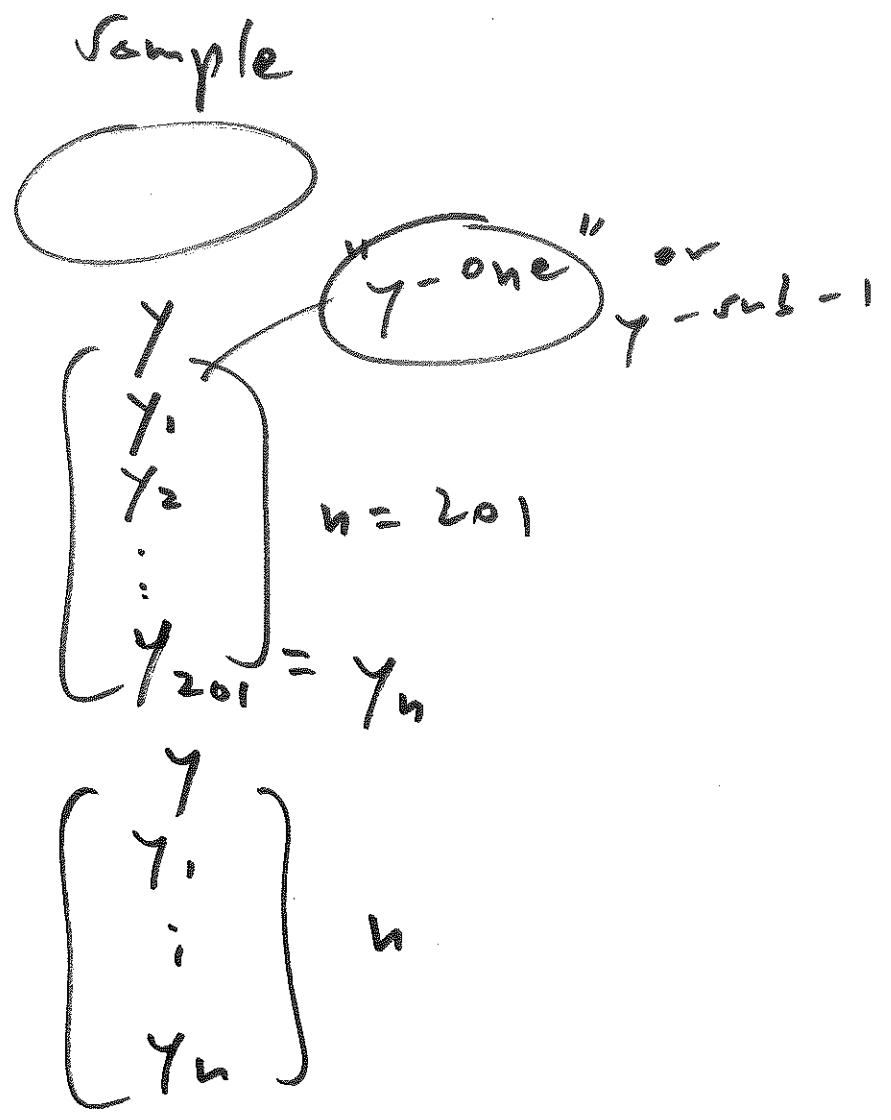
$$\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix}$$



sample

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \quad n=2$$

ST is  
more  
 informative  
 than IID,  
 but the math  
 is easier  
 with IID)



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

"y-bar" =  $\frac{1}{n} (y_1 + \dots + y_n)$

$$\bar{y} = \frac{1}{n} (y_1 + \dots + y_n)$$

$$= \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} (y_1 + y_2 + \dots + y_n)$$

Greek capital sigma

index of summation (ex. i or j)

the sum as i runs from 1 to n

2 possibilities for random sampling:

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

Variable

values the variable can take on

eye color in an animal

qualitative (categorical)  
brown, blue

dichotomous (circled)      nominal (circled)

success at running a maze

1 (very slow), 2 (slow), 3 (moderate), 4 (fast), 5 (very fast)

ordered categorical (circled) = ordinal (circled)

size of plant:

(conceptually) continuous (circled)

height

63.2, 61.8 (in cm)

ratio scale (circled)

quantitative (numerical)

# leaves

discrete (circled)

47, 62, ...

quant (circled)

plant: growing temperature that produces the most buds

quant. (circled)

continuous (circled)

81.4°F

79.7°F, ...

(in °F)

interval-scale (circled)

10.52 (circled)

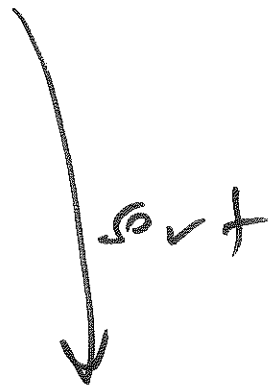
sample  
the observed  
butterflies

(y) wing length (cm)

- 4.4 =  $y_1$
- 3.6 =  $y_2$
- ⋮
- 3.9 =  $y_n$

$n = 24$

mean  $\bar{y} =$



- 3.3
- 3.5
- 3.6
- 3.6
- ⋮
- 4.5

$n = 24$

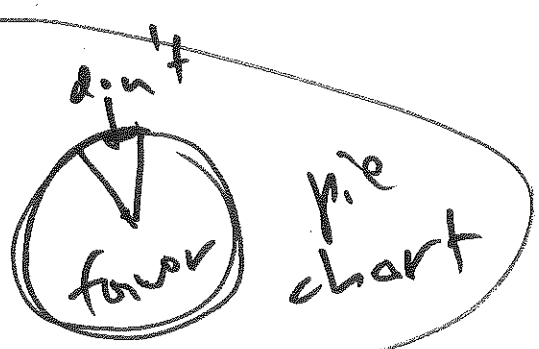
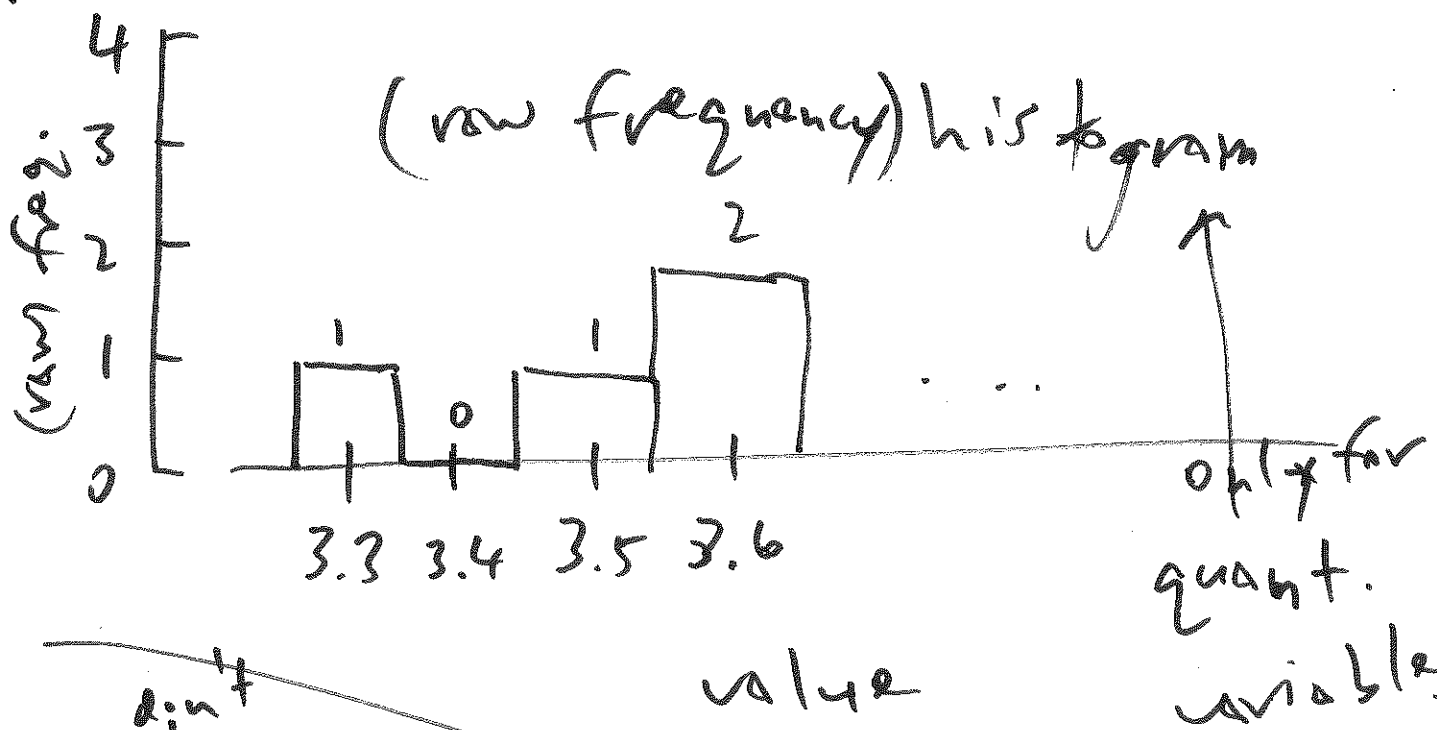
mean  $\bar{y}$

value	count (frequency)
-------	-------------------

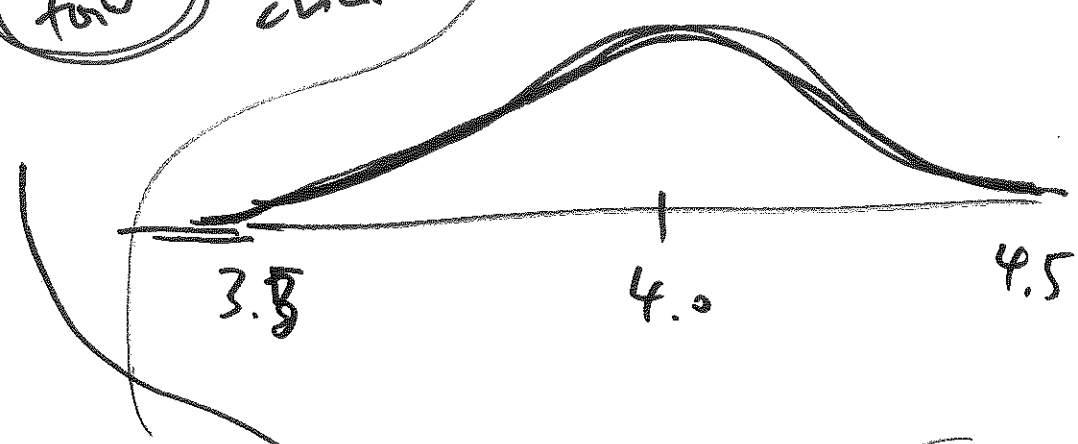
- 3.3
- 3.4
- 3.5
- 3.6
- ⋮
- 4.5

- 1
- 0
- 1
- 2
- ⋮
- 1

sum  
 $n = 24$



(disc. or cont., int. or ratio)



favor legitimation?

- Y
- N
- Y
- Y
- ...
- Y

