this time: variable types; histograms

next time: mean, SD

read: both DD books ch. 1, 2

AMS 7 2 Oct 18

(1)

DD office BE 357C

P = population = all deer at UCSC as of 31 Aug 2018

Y (yes) (1)
N (no) (0)

disease?

N
N
Y
⋮
N

→ [0, 0, 1, ⋮, 0]

N = 800 ← population size

mean $\theta = p_P = ?$ (recode)

parameter

sample (S) = the observed deer

at random

Y disease?
y_1 0
y_2 1
y_3 0
y_4 0
⋮
$y_n = y_{241}$ 0

$\begin{pmatrix} Y \\ N \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

n = sample size = 241

mean $\bar{y}$ = ("ybar") $= \frac{2}{241} \doteq 0.8\%$

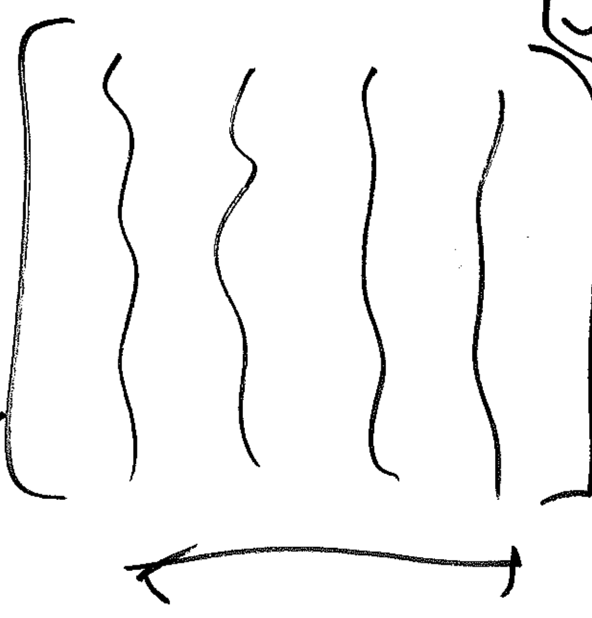
---

general dataset

(matrix)

1 column for each variable ← things we measure on subjects

1 row for each subjects (individuals) (elements)

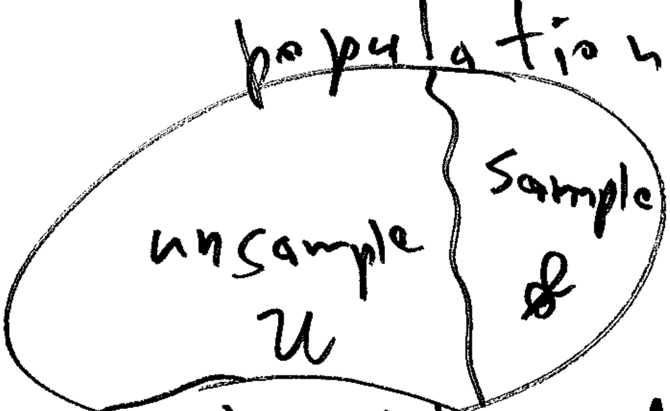


"any variable that takes on only 2 possible values is dichotomous (binary)

---

mean of a column of numbers

= average of that column = $\frac{\text{sum of all \#s}}{\text{how many \#s there are}}$



12.38

goal: try to sample so that sample & unsample are as similar as possible in all relevant ways

probability

1650: Pascal, Fermat

statistics

Bayes (1760)

simplest method: choose sample ③

at random

Neyman (1925)

(at random with replacement) = (independent identically distributed) IID

(at random without replacement) = (simple random sampling) (SRS)

SRS is more informative than IID, but the math is harder for SRS

(n = 1) SRS = IID

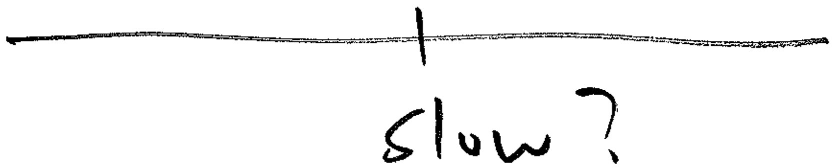
(n ≪ (is a lot smaller than) N) (SRS ≐ IID)

sample   analysis

~~Neyman (1925)/10~~

(4)

variable	possible values for variable (qual)
eye color	brown, blue (dichotomous) qualitative categorical (nominal)
success in running a maze	1 (very slow) 2 (slow) 3 (moderate) 4 (fast) 5 (very fast) — qual ordinal
size of a plant → conceptually → continuous	height (cm) (2.54 cm/in) 4.031... ex. 4.0 cm — quantitative (quant) (numerical)
	# leaves 12, 13, 14 — quant (discrete)
growing temperature (°C) (°F) ~~(K)~~ produces most buds	78°F


slow?