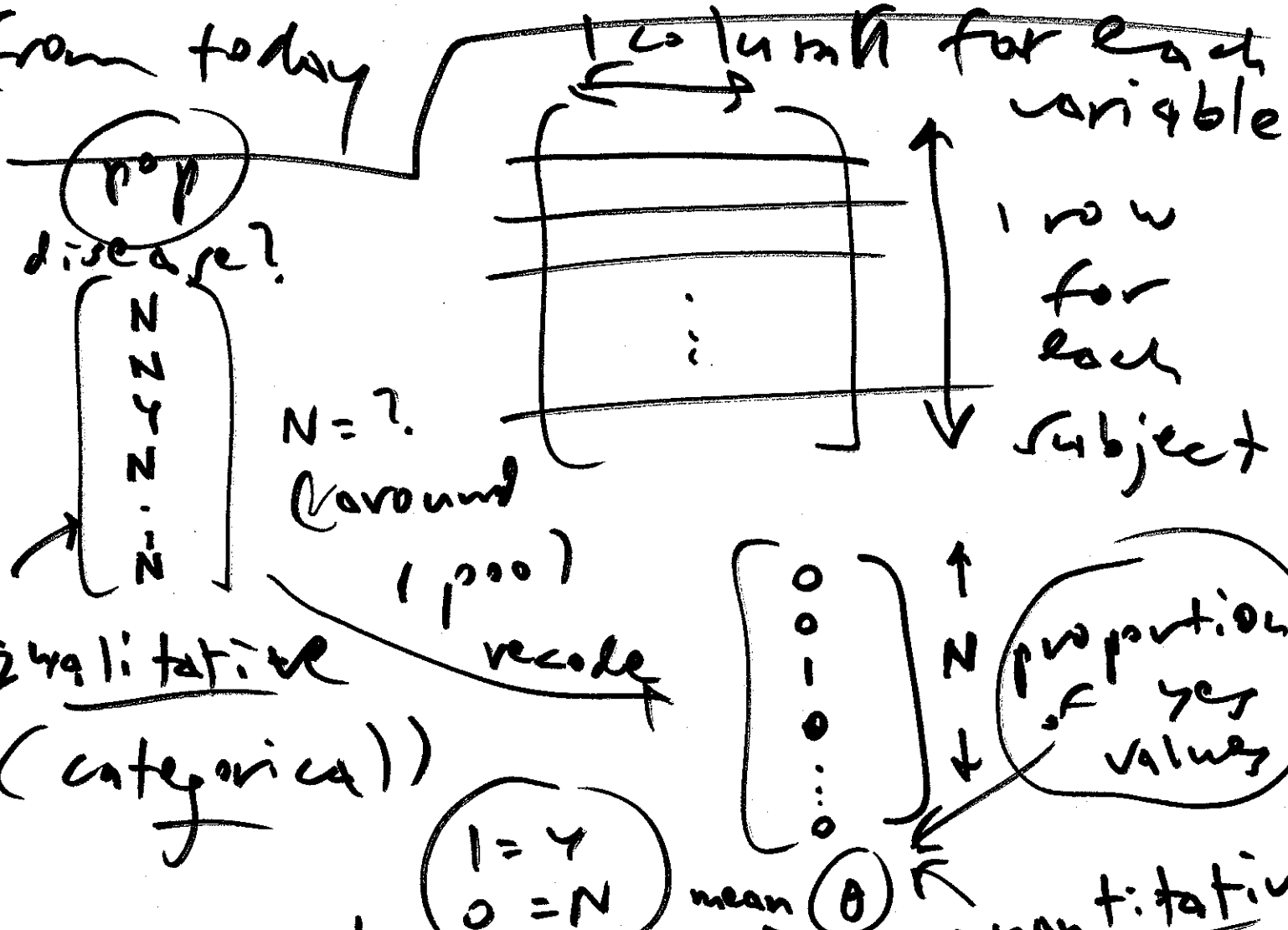
this: variable types;
time: histograms

next time: measures of center & spread

AMS 30 dep of F

read: ① TXT ch 1, 2

reader available from Joan Taylor at back of room: $20 cash; no more in class handouts 1 week from today



pop: all deer living on UCSC campus in July 2008

1 = Y
0 = N

sample: the observed deer

disease?

$N = 1{,}000$: $[\,1_s,\ 1,\ 0_s\,]$ — disease?

~~how chosen?~~ → at random

$[\,1_s,\ 1,\ 0_s\,]$ — $n = ?$ (ex. 100) ← sample size

mean $\theta = ?$

mean $\bar{y} = \frac{2}{100} = 0.02 \doteq 2\%$

$\bar{y}$ is a good estimate of θ

$\hat{\theta}$ ("θ hat")

goal: want θ

representative sampling method:

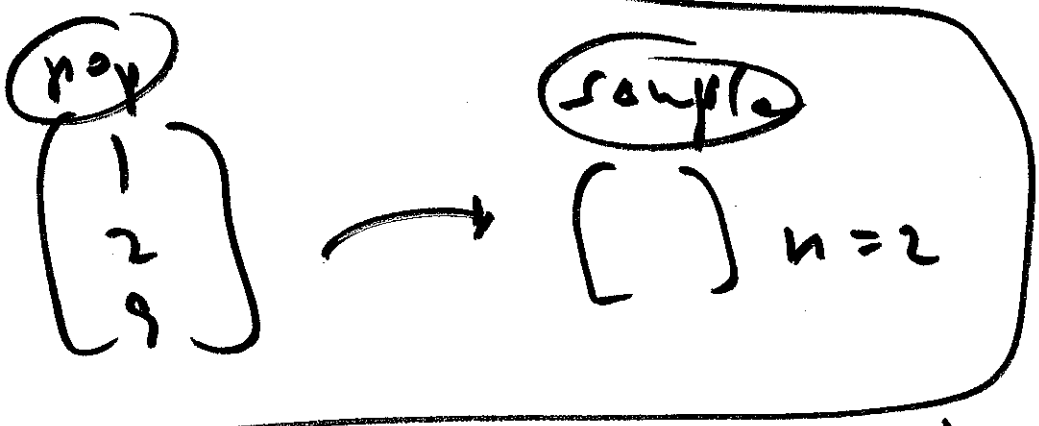
sample is similar to unsampled in all relevant ways

pop: unsampled | sample

how achieve this goal?

(Neyman, Fisher (1920s)) choose sample at random

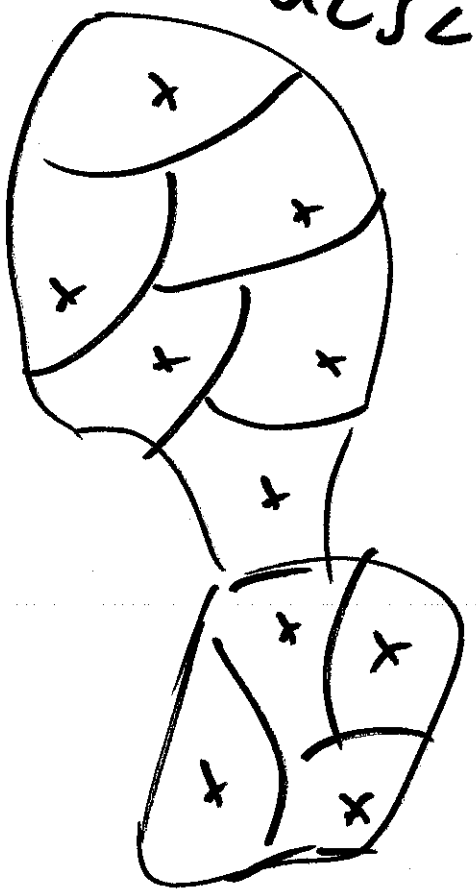
at random with replacement:



: independent
: identically distributed (IID)
sampling

at random without replacement: simple random sampling (SRS)

SRS is more informative than IID, so SRS is what is actually used often, but the math is easier for IID; if N is a lot bigger than n (N >> n) SRS is approximately IID and about the same SRS ≈ IID



UCSC

variable

eye color in animals

values

qualitative

nominal

brown, blue

dichotomous

quant

(-3)1 v. slow

(-1)2 slow

(0)3 mod

(+1)4 fast

(+3)5 v. fast

success in running a maze

quant

0, 1, 2, ...

# of leaves

between pos. val. (discrete)

size of a plant

ht in cm

quantitative

(e.g. 77 cm)

true zero

ratio

6.024 cm

no. of continuous

⊛ maze success

has order to it: ordinal

(considered categorical)

running temp at which most bugs produced

quant

interval

73° F

no obvs true zero

or

23° C

cont

wing length (cm)

4.4
3.6
⋮
3.9

$n = 24$

sort →

3.3
3.5
3.6
3.6
⋮
4.5

$n = 24$

value	(raw) frequency (count)
3.3	1
3.4	0
3.5	1
3.6	2
⋮	⋮
4.5	1
	$n = 24$

raw frequency distribution of wing length