this time: graphical & numerical

next time: descriptive methods

read: JJ ch. 1, 2, 3

key: visualize raw data

population (pop.)
all mule deer on 31 Dec 2006

1 column for each variable

1 row for each deer in pop

pop

① N = no
② Y = yes

pop. size

disease?	
N	0
N	0
Y	1
Y	1
Y	1
N	0

N = ?
(ex. 100-1000)

9 parameters

mean θ

pop

unsample $\mathcal{U}$ | sample $\mathcal{S}$

partition

pop: all WSFE deer on 31 (deer) — sample: the observed deer

1 = Y
0 = N

disease?

N = 800: [1s, 2, 0s] → (at random) → disease? [1s, 2, 0s] $n = 150$ ← sample size

mean θ ← unknown parameter

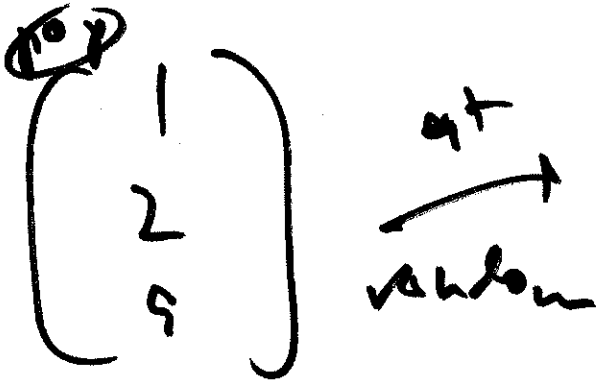
mean $\hat{\theta}$ = sample estimate of θ

if sample good, $\hat{\theta}$ is a good est. of θ

representative sampling

goal in sampling: try to make sample & unsample as similar as possible in all relevant ways

a simple way to achieve this: choose sample at random from pop.



---

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

---

at random without replacement
= simple random sampling (SRS)

① In practice people typically do SRS (or sampling equivalent) but math is easier for IID

② if n is a lot smaller than N ($n \ll N$), SRS ≈ IID

SRS : Neyman (1925)

qualitative or categorical

variable	values
eye color	brown, blue (dichotomous)
	nominal
success in maze running	1 (v. slow) 2 (slow)
	3 (mod.) 4 (fast)
	5 (v. fast)

qual

ordered categorical

=

ordinal

size of plant:

height	15.2 cm	quantitative, continuous, ratio-scale
# leaves	64	discrete, quant, ratio
growing temp. with most daily	19°C	quant, cont, interval

if 0 on scale stands for absence of thing measured, ratio;

if 0 arbitrary, interval

structural zero between possible values

wing length

4.4
3.6
|
|
3.9

$n = 24$

1 row for each butterfly

sort

3.3 cm
3.5
3.6
3.6
}
4.4
4.5

raw frequency

values	count
3.3	1
3.4	0
3.5	1
3.6	2
.	
.	
.	
4.4	1
4.5	1

$n = 24$

raw frequency distribution for wing length