this time: variable types; graphical & numerical descriptive methods

next time:

read: DD
(A) ch. 1, 2
(B) ch. 1-2

STAT 7
1 Oct 19

Quiz 1

due 1 week from today; tentatively

hwk 1 due Fri 11 Oct 19

(12.37)

at random without replacement = simple random sampling (SRS)

method	pro	con
IID	math easier	less informative
SRS	more informative	math harder

②

sample — disease? ✓ Y

$\begin{bmatrix} 1s \\ \& \\ 0s \end{bmatrix}$ $\begin{bmatrix} 1s \\ \& \\ 0s \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{113} \end{bmatrix}$ (Y) / "y sub 1" or "y1"

$n = 113$

$y_{113} = y_n$

"theta hat"

mean $\theta = ?$ mean $\bar{y} = \hat{\theta}$

mean $= \frac{0+0+\ldots+1+0+\ldots+0}{113} = \frac{1}{113} \doteq 0.9\%$

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

"y bar"

$= \frac{1}{n}(y_1 + \ldots + y_n)$

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

capital sigma

pop

$\begin{bmatrix} 3 \\ 4 \\ 11 \end{bmatrix}$ at random → $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$

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

Special cases

① $n = 1 \rightarrow$ IID = SRS

② $n = N \rightarrow$ census → SRS no uncertainty

→ IID still uncertainty

③ n (1,000) << N (200,000,000) → IID ≈ SRS

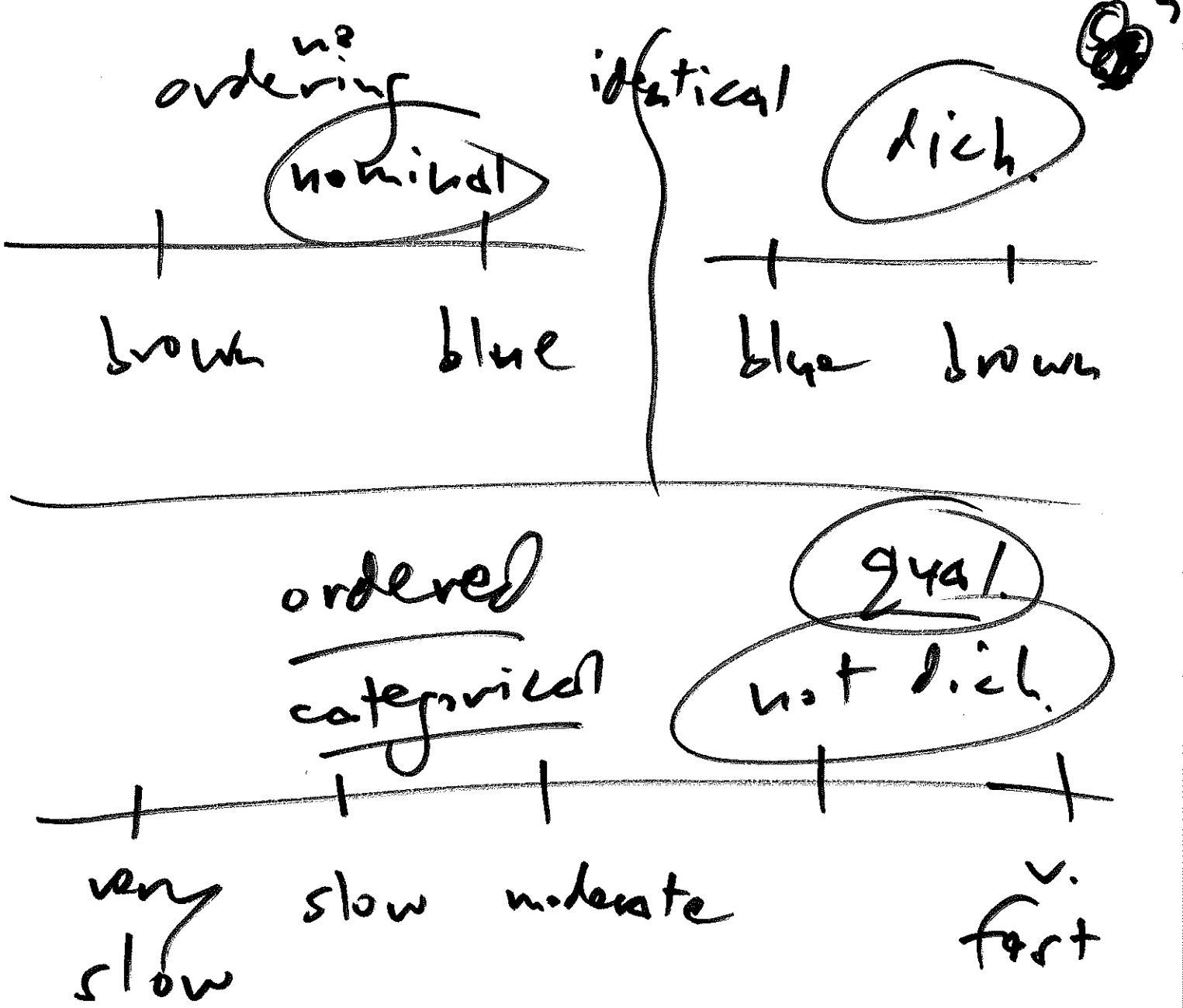
is a lot smaller than

variable	possible values	
eye color	qualitative: brown, blue	qual
	categorical	
plant size: height	14.1 cm ≈ 1 in (2.54 cm)	
	continuous	
	quantitative	quant
	numerical	
# leaves	19	quant
	discrete	

No unique place on # line for category 'brown' → equal

variable	possible values
maze running time	v. slow, slow, ..., v. fast

equal


no ordering
nominal
identical
rich
brown
blue
blue
brown

ordered categorical
qual.
not rich
very slow
slow
moderate
v. fast