

1-30-19

MODULES 7, 8, 9

* DATA: WHAT IS DATA?

↳ CONSIST OF INDIVIDUALS & VARIABLES THAT DESCRIBES THE MEASUREMENT OR GROUPINGS

- WE WANT DATA TO BE MEANINGFUL

* TWO TYPES OF VARIABLES:

- 1) QUANTITATIVE (NUMERICAL)
- 2) CATEGORICAL (QUALITATIVE)

* QUANTITATIVE VARIABLES

↳ ALL NUMERICAL VAR'S HAVE UNITS (\$, lbs, MILES)

⇒ NUMBERS CAN BE CATEGORICAL / QUALITATIVE

↳ ZIP CODES, PHONE #'S, ADDRESS

* IDENTIFIERS: NUMBERS THAT ARE ASSOCIATED

DIRECTLY TO AN INDIVIDUAL

(SS #, STUDENT ID #, PHONE #)

LOOKING @ SETS OF NUMERICAL DATA

HISTOGRAM: PLOTS THE BIN COUNT AS

HEIGHTS OF BARS

* (PLZ REFERENCE NUMERICAL DATA

SLIDES FROM CLAS BLOG LINK) SLIDE 1

1-30-19

* OUTLIERS: DATA POINTS THAT LIE IN THE EXTREMES OF A DISTRIBUTION

* RELATIVE FREQ. HISTOGRAM: SLIDE #2 DISPLAYS % OF CASES IN EACH BIN INSTEAD OF A DIRECT COUNT
↳ RELATIVE FREQUENCY HISTOGRAMS ARE FAITHFUL TO THE "AREA PRINCIPLE," (DEFINED LATER...)

* STEM & LEAF: ADD'L NUMERICAL DISPLAY WHICH PROVIDES FURTHER DETAIL INTO HISTOGRAM DATA (SEE # WOMEN VS PULSE RATE. SLIDE 3)
→ CAN REPLACE THE HISTOGRAM & RETAIN THE SAME SHAPE TRENDS





* DOTPLOTS: SIMPLE DISPLAY THAT PLACES SLIDE 4-4 DOTS ALONG AN AXIS (SEE DOTPLOT SLIDE)

SHAPE / CENTER / SPREAD (SLIDE 5)

- WHEN DESCRIBING DISTRIBUTION

1-30-19

- * SHAPE:
- 1) # OF HUMPS
 - 2) SYMMETRIC OR SKEWED (OUTLIERS / SKEW DATA)
 - 3) UNUSUAL FEATURE

- SLIDE 6
- 1) HUMPS: UNIMODAL ; BIMODAL  ETC
 - 2) SKEWED (L or R)  
 - 3) FEATURES: MAGNITUDE OR PEAKS


* UNIFORM DATA: NO HUMPS; A RECTANGULAR SHAPE
 ↳ NO MODE (OR MODALITY) & ALL THE BARS ARE APPROX. SAME HEIGHT

SLIDE 8

* SYMMETRY: FOLDS THE HISTOGRAM ON A CENTRAL LINE & EDGES MATCH

SLIDE 9

→ TAILS: THINNER ENDS OF A DISTRIBUTION



SLIDE 10

→ A HISTOGRAM IS SKEWED TO THE SIDE OF A LONGER TAIL

* CENTER:

- 1) MEDIAN - VALUE \geq 1/2 DATA ABOVE/BELOW
- 2) MEAN (AVERAGE) [PT WHERE HISTOGRAM BALANCES]

SLIDE 13

SLIDE 12

1, 2, 3, 4, 5 \Rightarrow AVG 3

1, 2, 3, 4, 100 \Rightarrow AVG 22, MEDIAN 3

NOTE \Rightarrow TO FIND MEDIAN OF AN EVEN #'S DATA SET, FIND THE MEAN (AVG) OF THE TWO MIDDLE POINTS:

1-30-19

- NOTE: \downarrow USE MEDIAN WHEN DATA IS SKEWED!
USE MEAN WHEN DATA IS SYMMETRICAL

- IF THE DATA IS SYMMETRIC!
 \hookrightarrow MEAN \approx MEDIAN

- IF DATA IS SKEWED TO THE RIGHT (R)
THEN MEAN WILL BE LARGER THAN
MEDIAN

\Rightarrow IF SKEWED TO (L), THE MEDIAN
WILL BE LARGER

SPREAD: REPORT A MEASURE OF SPREAD
ALONG τ MEASURE OF CENTER WHEN
NUMERICALLY DESCRIBING A DISTRIBUTION

- RANGE: Δ BETWEEN MAX & MIN OF VALUES
RANGE = MAX - MIN [SLIDE 15]

- INTERQUARTILE RANGE (IQR) \Rightarrow ALWAYS
TO IGNORE EXTREME VALUES & LOOK
@ MIDDLE VALUES
 \hookrightarrow IF USING MEDIAN & DATA SKEWED,
USE IQR.

1-30-19

• QUANTILES: DIVIDE DATA INTO FOUR = SECTIONS

↳ Q_1 - 25% DATA LIES BELOW LOWER

QUANTILE

Q_2 - 25% DATA LIES ABOVE UPPER

QUANTILE

$$\Rightarrow IQR = Q_3 - Q_1$$

SLIDES 17/18