Lecture 1: What to learn ?

A Conceptual framework for Statistical Analysis

- develop Language for Statistical reasoning and probabilistic argument

- **Day 1 (today) : variable, mean, median, standard deviation**
- **Can get more involved and confusing, inconsistency,**
- **If well-trained, understandable from the context;**
- **math (precise), humanity(flexible); errors (tolerable);**
- **Useful for biological, studies : events of interests are either abundant and rare.**
- **most of time it works; not always,**
- **Challenges: to know which method works, when, why?**
- **Examples first, then Generalization;**
- **how to ask questions (problem formulation);**

Mean and Median

- Why not use median?
- Hard to manipulate mathematically?
- Median price of this week (gas) is $1.80
- Last week : $2.0
- What is the median price for last 14 days?
- Hard! How about if last week's median is $1.80
- Still hard.
- The answer : anything is possible! Give Examples.
- Note(require Math): Minimize average of absolute distances.

Measure of dispersion

- Maximum - minimum=range
- Average distance from average
- Average distance from median
- Interquartile range= third quartile - first quartile
- Standard deviation = square root of average squared distance from mean
- The most popular one is standard deviation (SD)

Step by Step illustration for finding median through Stem-leaf plot

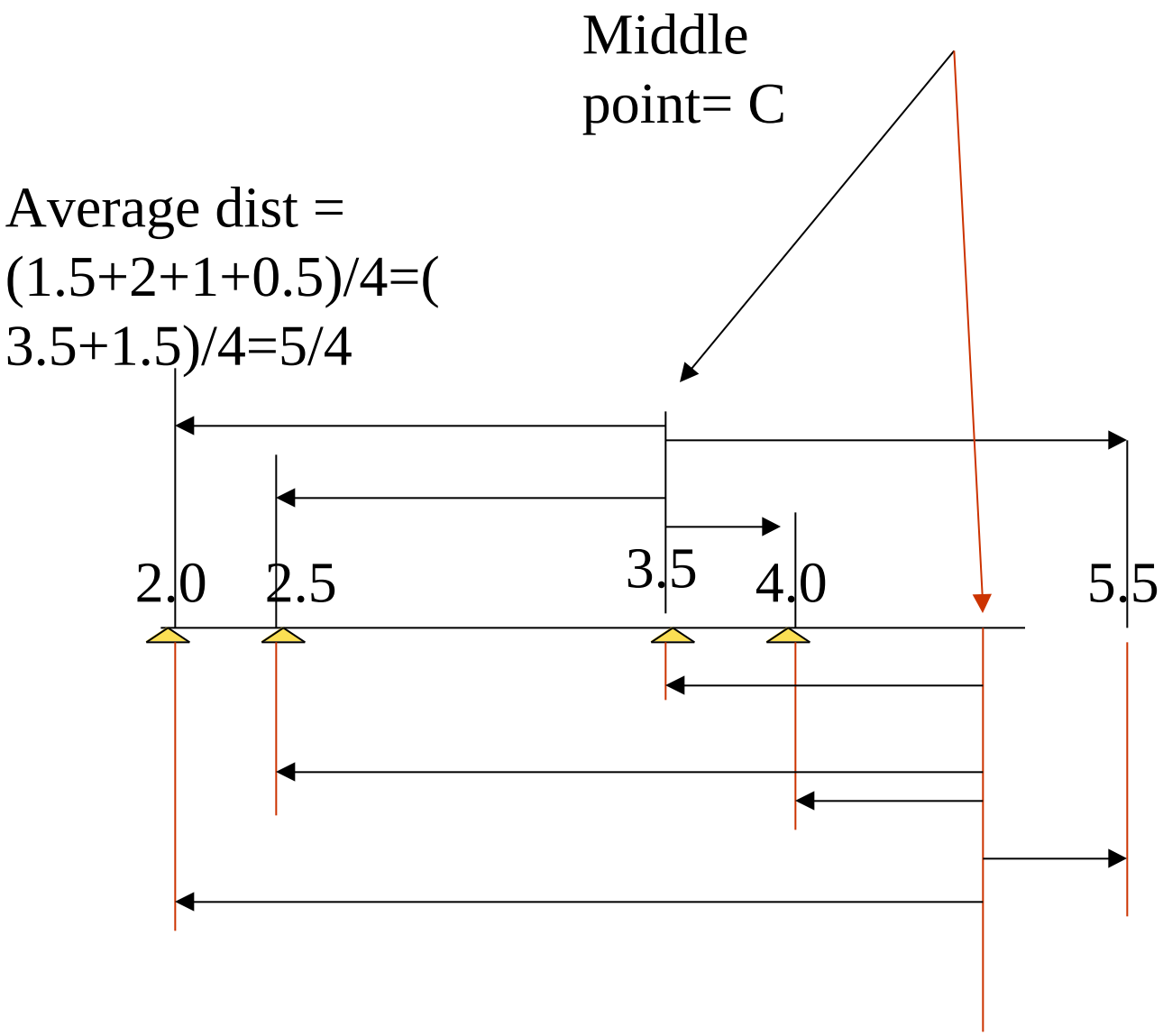
STAT 13 -Lecture 1

From stem-leaf to histogram

- Using drug response data

- NOT all bar charts are histograms!!!
- NCBI's COMPARE

- Histograms have to do with "frequencies"


Middle
point= C
Average dist =
(1.5+2+1+0.5)/4=(
3.5+1.5)/4=5/4
2.0
2.5
3.5
4.0
5.5

Homework 1 (due Tuesday 2nd week)

STAT 13 -Lecture 1