Chapter 3 Descriptive Statistics for Quantitative Data

We know how to form a histogram from Chpt 2.

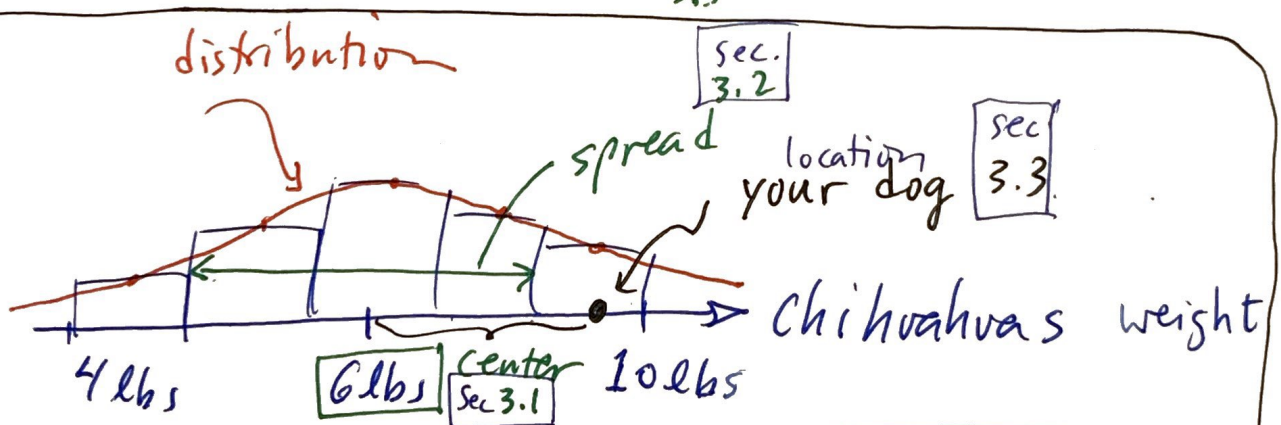
Here we learn methods that help us assign numbers to those histograms, dot plots etc.

The Shape of the Frequency Polygon connecting the bars of a histogram

Given a "distribution" of data

- we can determine the location of its center (3.1)
- we can determine the spread (3.2) of the distribution
- Then when we have an individual data point we can show its location/position (3.3) in the distribution

EX



Q: It is unusual that a chihuahua has a weight of 9lbs?

We answer these kinds of questions in this Chapter.

mean

sum of all data in the set

Def: $\bar{x} \equiv \frac{\sum x_i}{N}$ ← total number of data points

((Frequently the mean is refered to as the average))

Ex Mean quiz score

- Raw: 10, 9, 10, 3, 8, 7, 10, 9, 4 $N = 9$
- mean: $\frac{10+9+10+3+8+7+10+9+4}{9} = \frac{70}{9} = 7.78$

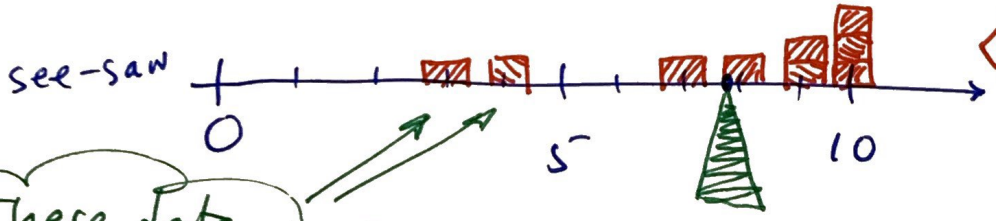
calculator: clear out [2nd] [CSR]

10 [Σ+]
9 [Σ+]
3 [Σ+]
⋮
9 [Σ+]
4 [Σ+]

[2nd] [$\bar{x}$] → 7.78

The mean is Not the best measure of center for skewed data

- Interpretation of the mean: let data = bricks



These data scores will decrease the mean alot!

Balancing Point of the Data Set

Median The middle data point is called the median. Half the data is to the left and half is to the right of the median

The Calculation of the median depends on whether we have an even or odd count in our data set

Ex Odd set : 92, 83, 68, 85, 78 $N = 5$

1. Order the data 68, 78, 83, 85, 92
2. Pick the middle number
3. That number called the median of the set

Ex Even Set : 20, 15, 12, 27, 13, 19, 13, 21

1. Order the data: 12, 13, 13, 15, 19, 20, 21, 27
2. We have no middle number so average the two middle numbers instead:

$$\frac{15+19}{2} = \frac{34}{2} = 17$$

3. The median is 17 even though 17 is not a member of our data we call it the median nevertheless!

- We saw that the mean is highly influenced by an extreme point, the median however is not as influenced. (4)

Ex Here are the annual incomes of 5 families in a five-plex row of apartments.

- ordered : 25k, 31k, 34k, 44k, 56k
- mean $\frac{25+31+34+44+56}{5} * 1000 = 38{,}000\$$ → mean

Now One family inherits 1.025 M dollars in the next yr. ((let that be the #25000 family))

- mean : $\frac{1{,}025k + 31k + 34k + 44k + 56k}{5} = 238{,}000$ → mean

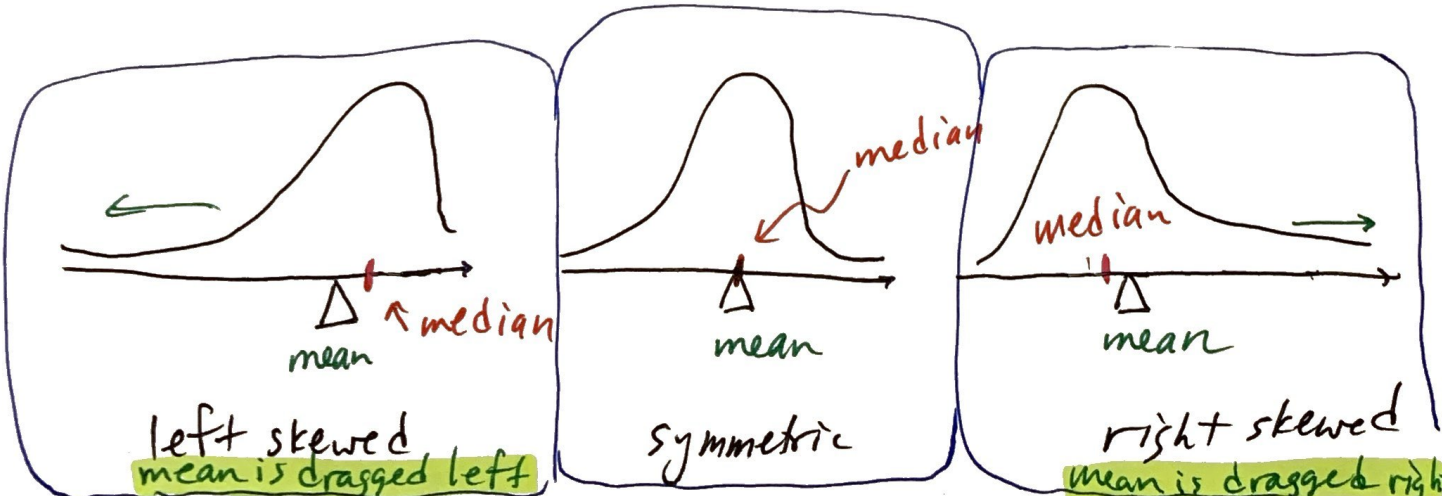
pre-inheritance

- mode : 25, 31, 34, 44, 56 → mode = 34,000

post-inheritance

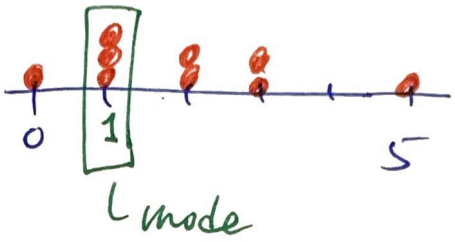
- mode 31, 34, 44, 56, 1025 → mode = 44,000

(*) Observations: The mean and median can be summarized in skewed distributions as:

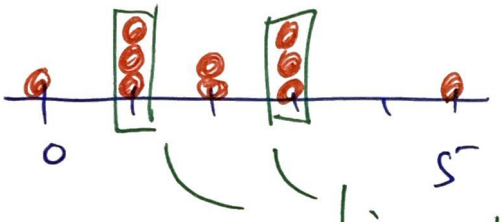


Mode The most frequent data point is called the mode. {"most popular"}

Ex: ordered data: 0, 1, 1, 1, 2, 2, 3, 3, 5 — 1 = mode



After picking a quiz up from the T.L.C. my ordered dat is (new) 0, 1, 1, 1, 2, 2, 3, 3, 3, 5



- mode can be applied to qualitative (categorical data)

Ex: makes of cars in a parking lot

Honda	Toyota	Toyota	Honda	Ford
Chevy	Nissan	Ford	Chevy	Chevy
Honda	Dodge	Ford	Ford	Toyota

Count:	Honda	Toyota	Ford	Chevy	Niss.	Dodg.
	III	III	IIII	III	I	I

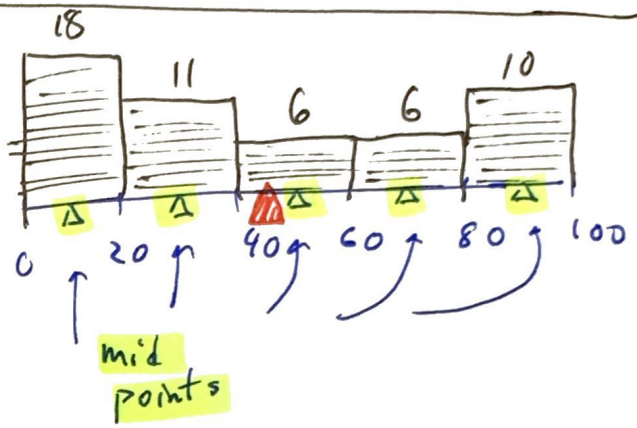
mode: Ford

We have completed the three measures of center.

⊛ Lets look at how to estimate the mean when we are given a frequency table -OR- a histogram:

Ex

class	Freq	mid point
0-19	18	10
20-39	11	30
40-59	6	50
60-79	6	70
80-99	10	90
	51	



To estimate the mean we select the middle number of each class

$$\bar{x} = \frac{\sum_{\text{class}} \text{midpoint} * \text{Frequency}}{\text{Total Data Points}}$$

Counts

$$= \frac{10*18 + 30*11 + 50*6 + 70*6 + 90*10}{51}$$

$$= \frac{2130}{51} = 41.76$$

$$\bar{x} \approx 41.8$$

approximate

* Summary:

Distribution	Center
• Unimodal symmetric	mean
• Unimodal skewed / outliers	median
• scattered distribution	mode