



Quantitative Data Properties

Loco por los Datos

Quantitative Data Properties

 Trend.

 Dispersion.

 Shape.

crisys
Loco por los Datos

Quantitative Data Properties

□ Trend

■ Arithmetic mean

No	Employee ID	First Name	Last Name	Age	Worked years	Salary	Status	Grade
1	1000001	John	Denver	23	1	\$500	Single	Elementary
2	1000002	Peter	Hank	30	3	\$900	Married	High School
3	1000003	Jack	Sullivan	27	2	\$900	Married	High School
4	1000004	Marco	Aurelio	40	8	\$1,500	Married	Master Degree
5	1000005	Claudia	Perez	35	5	\$1,300	Single	Master Degree

$$\bar{x} = \frac{\sum_{i=1}^n X_i}{n}$$

Average salary = $\frac{500 + 900 + 900 + 1,500 + 1,300}{5} = 1,020$

Quantitative Data Properties

□ Trend

■ Median (case 1)

Employee ID	First Name	Last Name	Age	Worked years	Salary	Status	Grade
1000001	John	Denver	23	1	\$500	Single	Elementary
1000002	Peter	Hank	30	3	\$900	Married	High School
1000003	Jack	Sullivan	27	2	\$900	Married	High School
1000004	Marco	Aurelio	40	8	\$1,500	Married	Master Degree
1000005	Claudia	Perez	35	5	\$1,300	Single	Master Degree

\$500	\$900	\$900	\$1,500	\$1,300
1	2	3	4	5
\$500	\$900	\$900	\$1,300	\$1,500



Quantitative Data Properties

□ Trend

■ Median (case 2)

CustomerID	Type	Payments	Purchases	Sales	Refunds	Country	Continent
10000	Person	Cash	120,000	150,000	240	Canada	America
10001	Company	Cash	521,400	651,750	1,043	Japón	Asia
10002	Company	Credit Card	451,000	563,750	902	Mexico	America
10003	Company	Transference	565,000	706,250	1,130	España	Europe
10004	Person	Transference	512,300	640,375	1,024	Argentina	America
10005	Person	Transference	415,500	519,375	0	Canada	America
10006	Company	Credit Card	696,300	870,375	1,392	EEUU	America
10007	Person	Cash	741,000	926,250	1,482	Chile	America

150000	651750	563750	1206250	640375	519375	1120375	926250
1	2	3	4	5	6	7	8
150000	519375	563750	640375	651750	926250	1120375	1206250

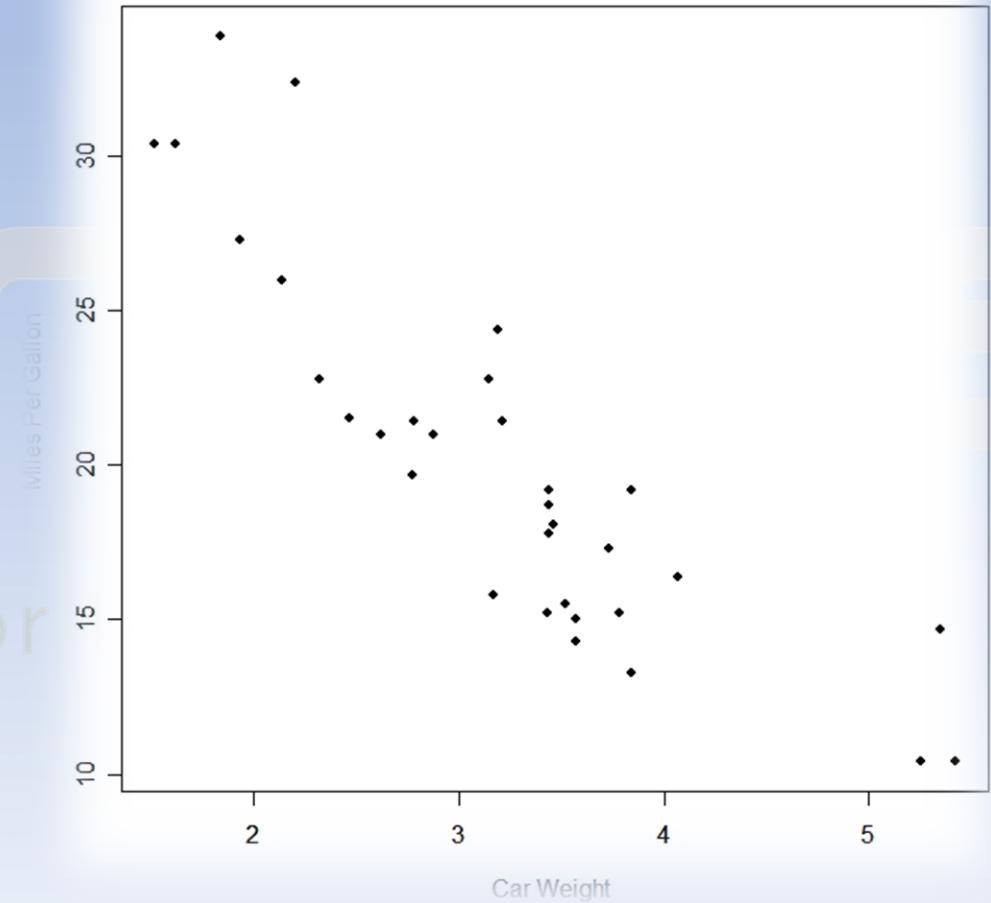
$$\text{Median} = \frac{640,375 + 651,750}{2} = 646,063$$

Quantitative Data Properties

□ Dispersion

- Standard Deviation.
- Low value indicate closeness.
- High value indicate farness.

Scatterplot Example



Quantitative Data Properties

Why we should be concerned about data dispersion?

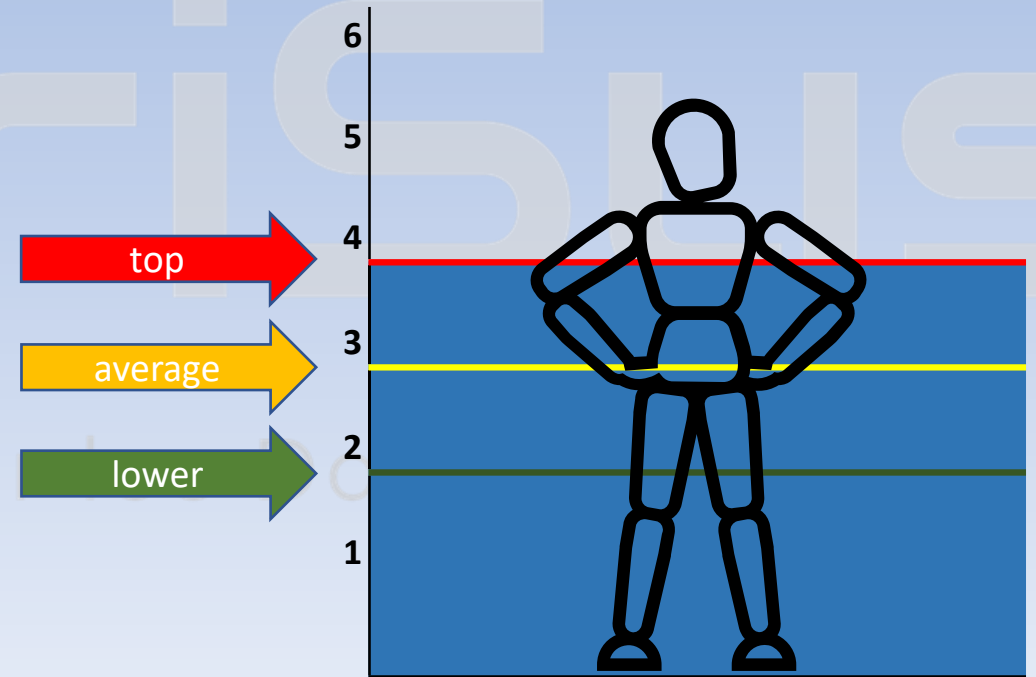
Depth average = 3 feet

Variation average = 1 feet

$3 + 1$

$3 - 1$

Between 4 and 2 feet.



Quantitative Data Properties

Why we should be concerned about data dispersion?

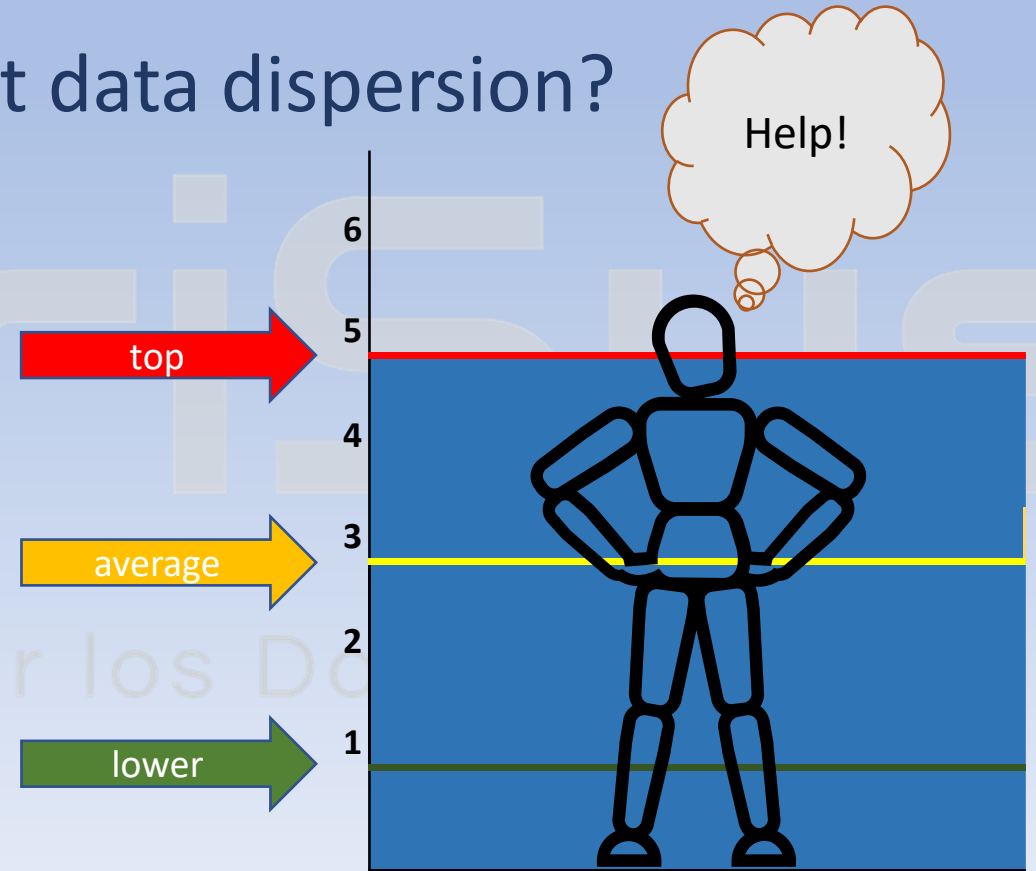
Depth average = 3 feet

Variation average = 2 feet

$3 + 2$

$3 - 2$

Between 1 and 5 feet.



Quantitative Data Properties

Variance:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n - 1}$$

x_i	\bar{x}	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
Salary	Average	Difference	Square Difference
500	1,020	-520	270,400
900	1,020	-120	14,400
900	1,020	-120	14,400
1,500	1,020	480	230,400
1,300	1,020	280	78,400
		0	608,000

$$s^2 = \frac{608,000}{5 - 1} = 152,000$$

Quantitative Data Properties

Standard Deviation:

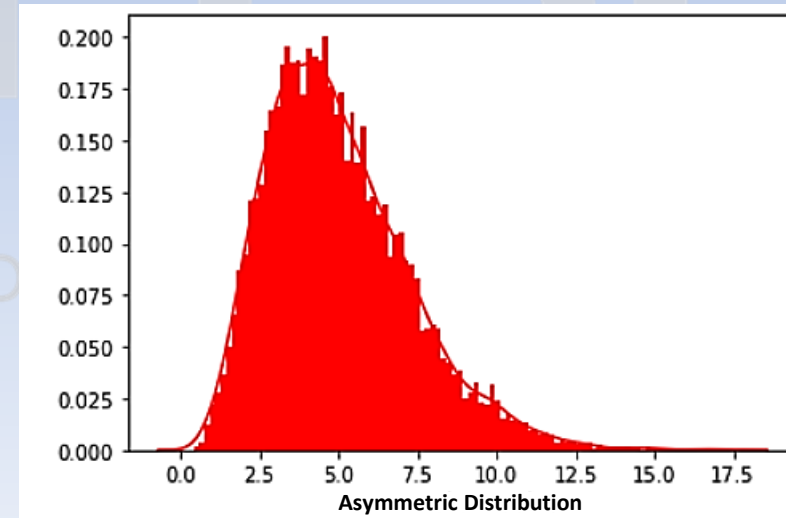
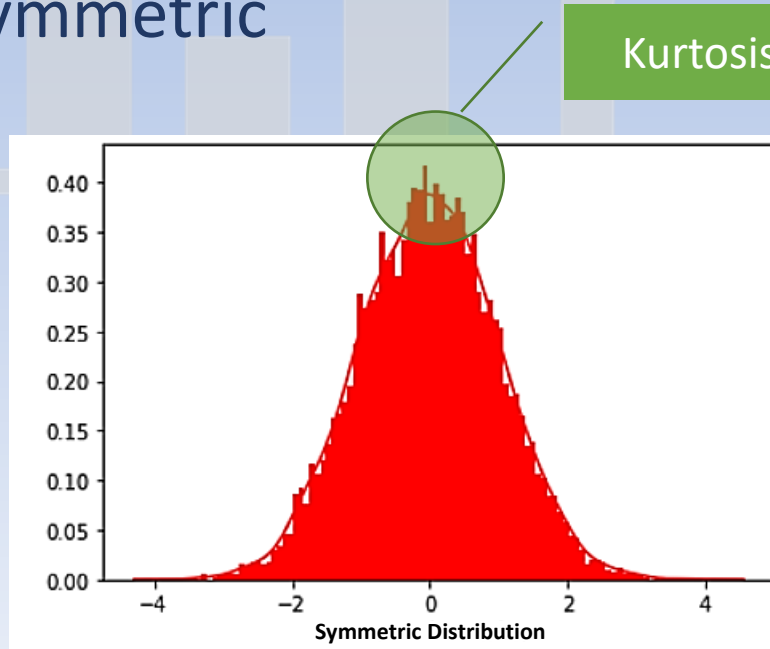
$$S = \sqrt{S^2} \quad S = \sqrt{152,000} \approx 390$$

The salaries move between ± 390 respect to the mean.

Quantitative Data Properties

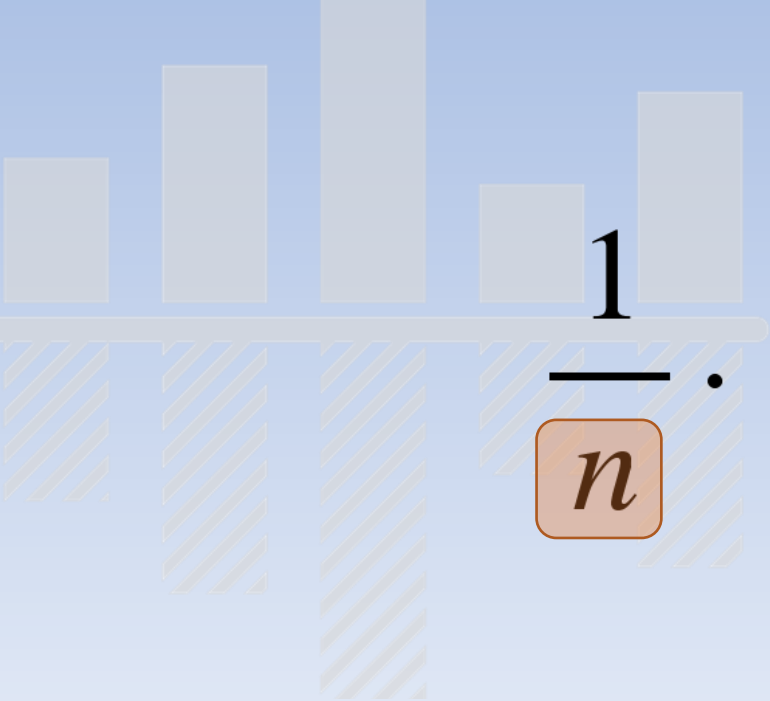
□ Shape

- Symmetric
- Asymmetric



Quantitative Data Properties

□ Skewness



$$\frac{1}{n} \cdot \frac{\sum_{i=1}^n (x_i - \bar{X})^3}{s^3}$$

The formula is annotated with colored boxes and arrows: an orange box around n , a yellow box around x_i , a green box around \bar{X} , and a blue box around s . Green arrows point from the right towards the 3 exponents on $(x_i - \bar{X})$ and s .

<https://en.wikipedia.org/wiki/Skewness>

Quantitative Data Properties

□ Kurtosis


$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^4$$

s^4

Two green arrows point to the terms $(x_i - \bar{X})^4$ and s^4 in the equation.

<https://en.wikipedia.org/wiki/Kurtosis>