Statistic (統計量)

- The problem of dimensionality: We need indexes

- Mean $\bar{x} = \frac{\sum x}{n}$ Variance $s^2 = \frac{\sum (x - \bar{x})^2}{(n - 1)}$

Standard Deviation

Formula 1 $s = \sqrt{\frac{n \sum x^2 - (\sum x)^2}{n(n-1)}}$ Formula 2 $s = \sqrt{\frac{\sum (x - \bar{x})^2}{(n - 1)}}$

Index	New York City	Los Angeles
1	$ 1.00	$ 1.00
2	$ 2.00	$ 2.00
3	$ 3.00	$ 3.00
4	$ 3.00	$ 4.00
5	$ 5.00	$ 5.00
6	$ 6.00	$ 6.00
7	$ 7.00	$ 7.00
8	$ 8.00	$ 8.00
9	$ 9.00	$ 9.00
10	$ 11.00	$ 10.00
11	$ 66.00	

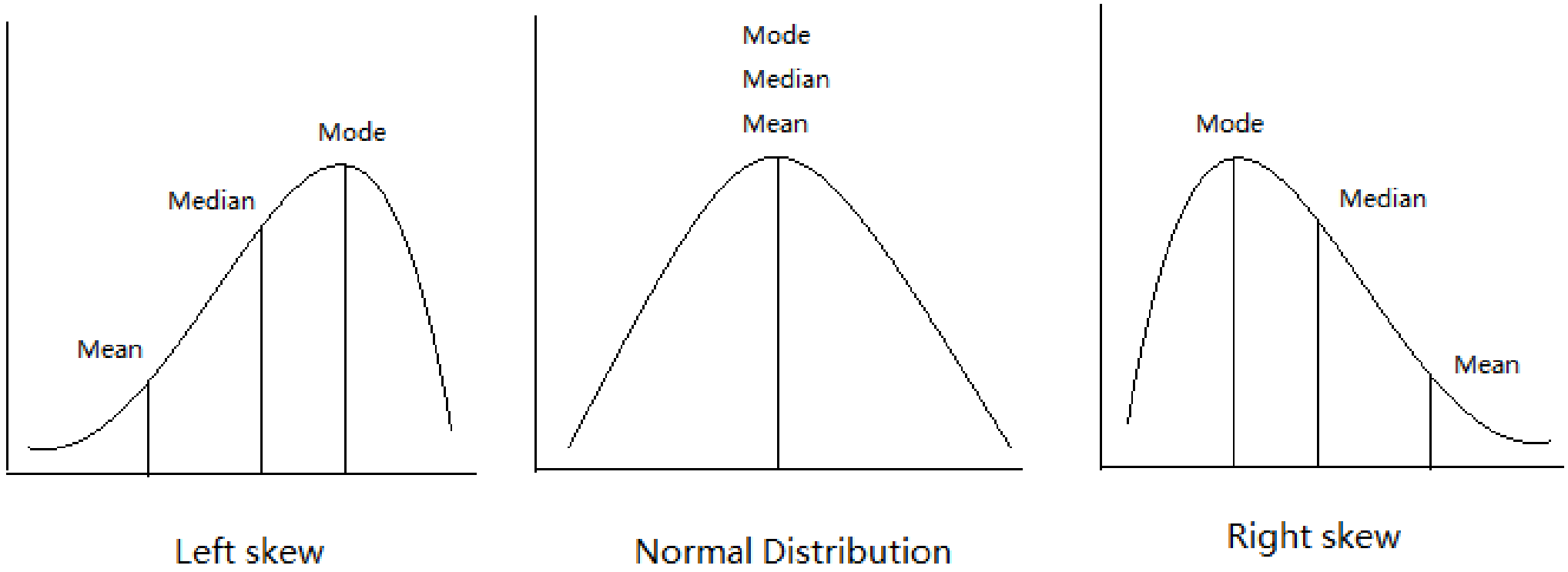
Median

Index	New York City	Los Angeles
1	$ 1.00	$ 1.00
2	$ 2.00	$ 2.00
3	$ 3.00	$ 3.00
4	$ 3.00	$ 4.00
5	$ 5.00	$ 5.00
6	$ 6.00	$ 6.00
7	$ 7.00	$ 7.00
8	$ 8.00	$ 8.00
9	$ 9.00	$ 9.00
10	$ 11.00	$ 10.00
11	$ 66.00	

Mode

Measures of Center

- Mean, Median vs Mode



Simpson's Paradox

授課語言	統計量	2000入學生	2011入學生
中文授課	學生人數	2000	2000
	被當人數	4	2
	比率	0.002	0.001
英語授課	學生人數	600	2500
	被當人數	9	25
	比率	0.015	0.01
全校	學生人數	2600	4500
	被當人數	13	27
	比率	0.05	0.06