భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Biostatistics BT2023

Lecture 10

Himanshu Joshi
6 September 2022

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Start-up notes

Reading material folder

Standard deviation

Derivations ?

$$\sigma = \sqrt{\frac{\sum_n (x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum_n x_i^2 - n\bar{x}^2}{n}}$$

n used for Population

$$s = \sqrt{\frac{\sum_n {x_i}^2 - n\bar{x}^2}{n-1}}$$

n-1 used for Sample

Variance = σ^2

Standard deviation by transformation

y=ax $$\sigma_y = a\sigma_x$$

y=a+x $$\sigma_y = \sigma_x$$

y=a+cx $$\sigma_y = c\sigma_x$$

Standard error in Mean

$$\sigma_M = \frac{\sigma}{\sqrt{n}}$$

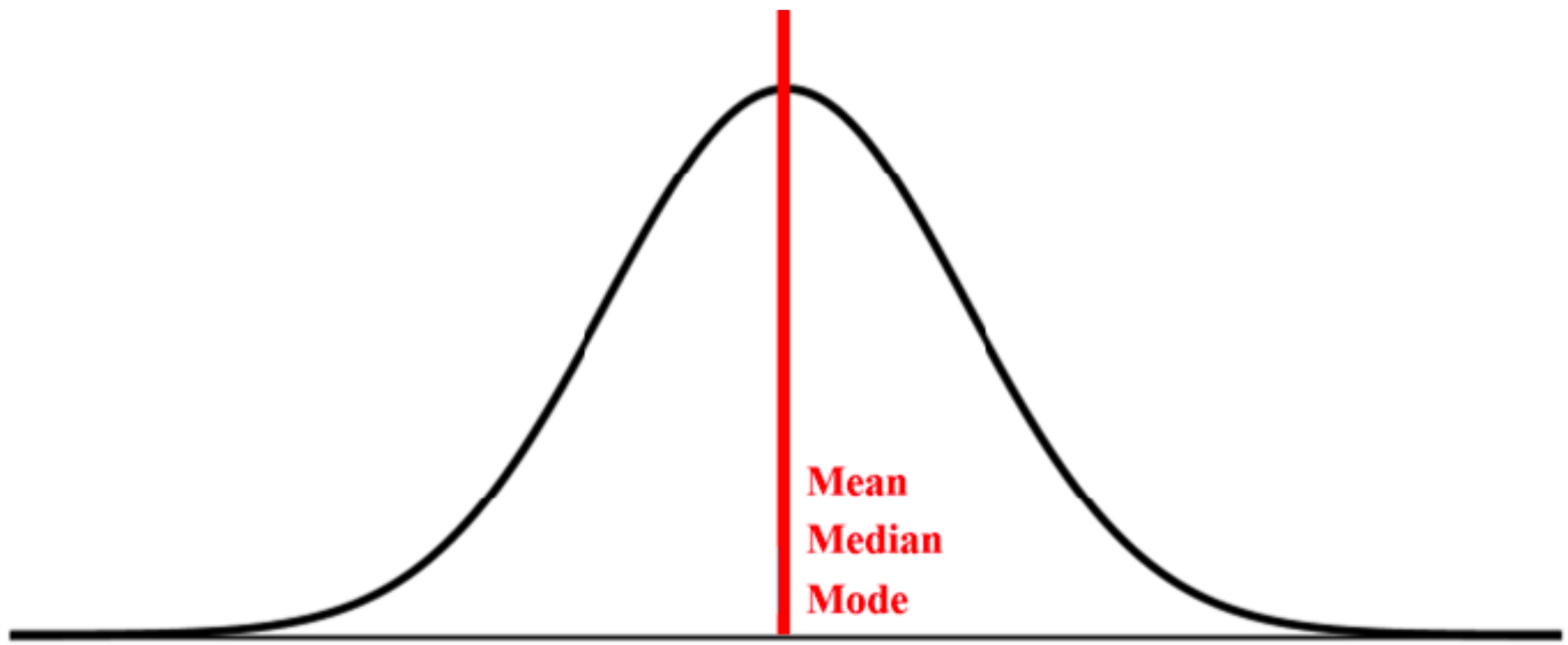
Z-score

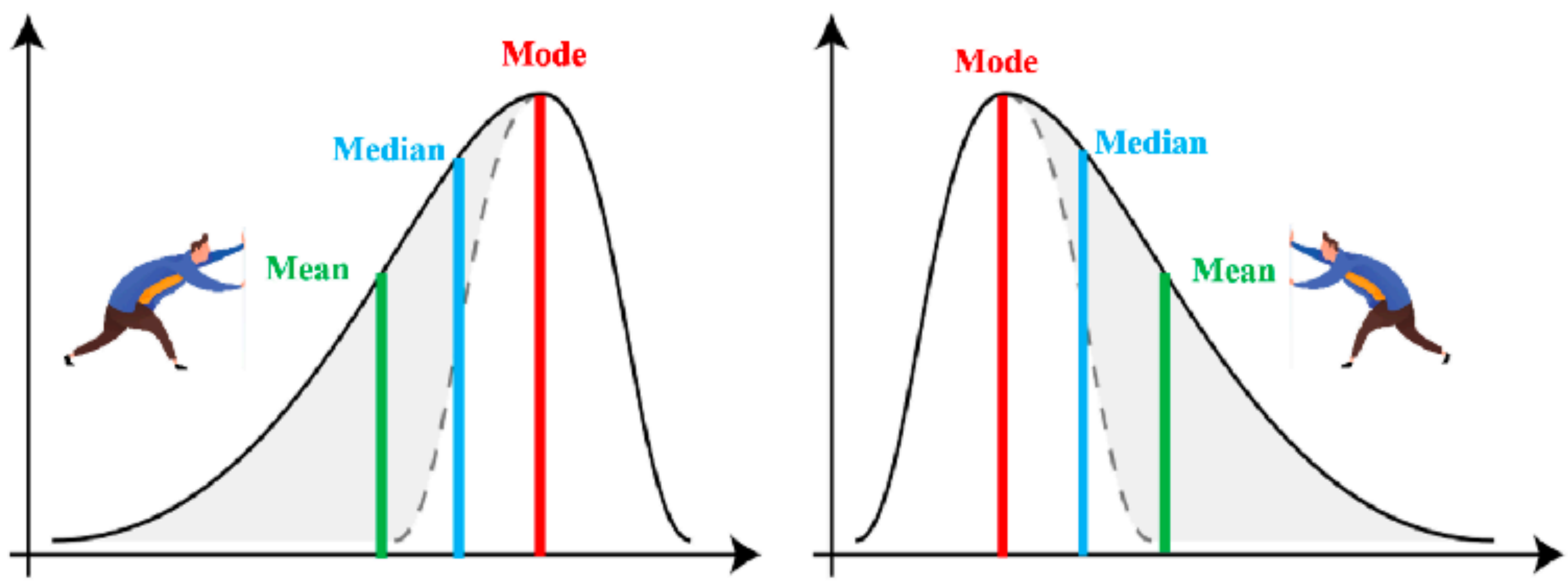
$$Z_{score} = \frac{x - \bar{x}}{\sigma}$$

To find if the data points is to the outliers

Z_score > 3 are called outlier

Skewness





Reference
Towards Data Science

Absolute skewness

Mean - Mode

Karl Pearson coefficient of skewness

$$S_k = \frac{Mean - Mode}{Stanadrd\ deviation} = \frac{3(Mean - Median)}{Standard\ deviation}$$

Bowley’s coefficient of skewness

$$S_k = \frac{Q_3 - Q_1 - 2Median}{Q_3 - Q_1}$$

$$-1 \geq S_k(Bowley's) = \leq +1$$

Skewness

Kelly’s measure of skewness

$$S_k = \frac{P_{10} + P_{90} - 2Median}{P_{90} - P_{10}}$$

Since Bowley’s coefficient ignore 50 % of the values in the extreme, hence this method was suggested by Kelly.

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Skewness coefficient based on moments

Moments in mechanics is refers to the rotating effect of a force, in statistics it is used to describe the peculiarities in the frequency distribution

$First\ moment\ : \mu_1$ $\quad \dfrac{\sum_N (X-\bar{X})}{N} \quad \dfrac{\sum_N f(X-\bar{X})}{N}$

Always 0

$Second\ moment\ : \mu_2$ $\quad \dfrac{\sum_N (X-\bar{X})^2}{N} \quad \dfrac{\sum_N f(X-\bar{X})^2}{N}$

Measure the variance

$Third\ moment\ : \mu_3$ $\quad \dfrac{\sum_N (X-\bar{X})^3}{N} \quad \dfrac{\sum_N f(X-\bar{X})^3}{N}$

Measure skewness

$Forth\ moment\ : \mu_4$ $\quad \dfrac{\sum_N (X-\bar{X})^4}{N} \quad \dfrac{\sum_N f(X-\bar{X})^4}{N}$

Measure Kurtosis

Skewness

Skewness coffcient $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$

adjusted Fisher-Pearson coefficient $$\frac{\sqrt{N(N-1)}}{N} \frac{\mu_3^2}{\mu_2^3}$$

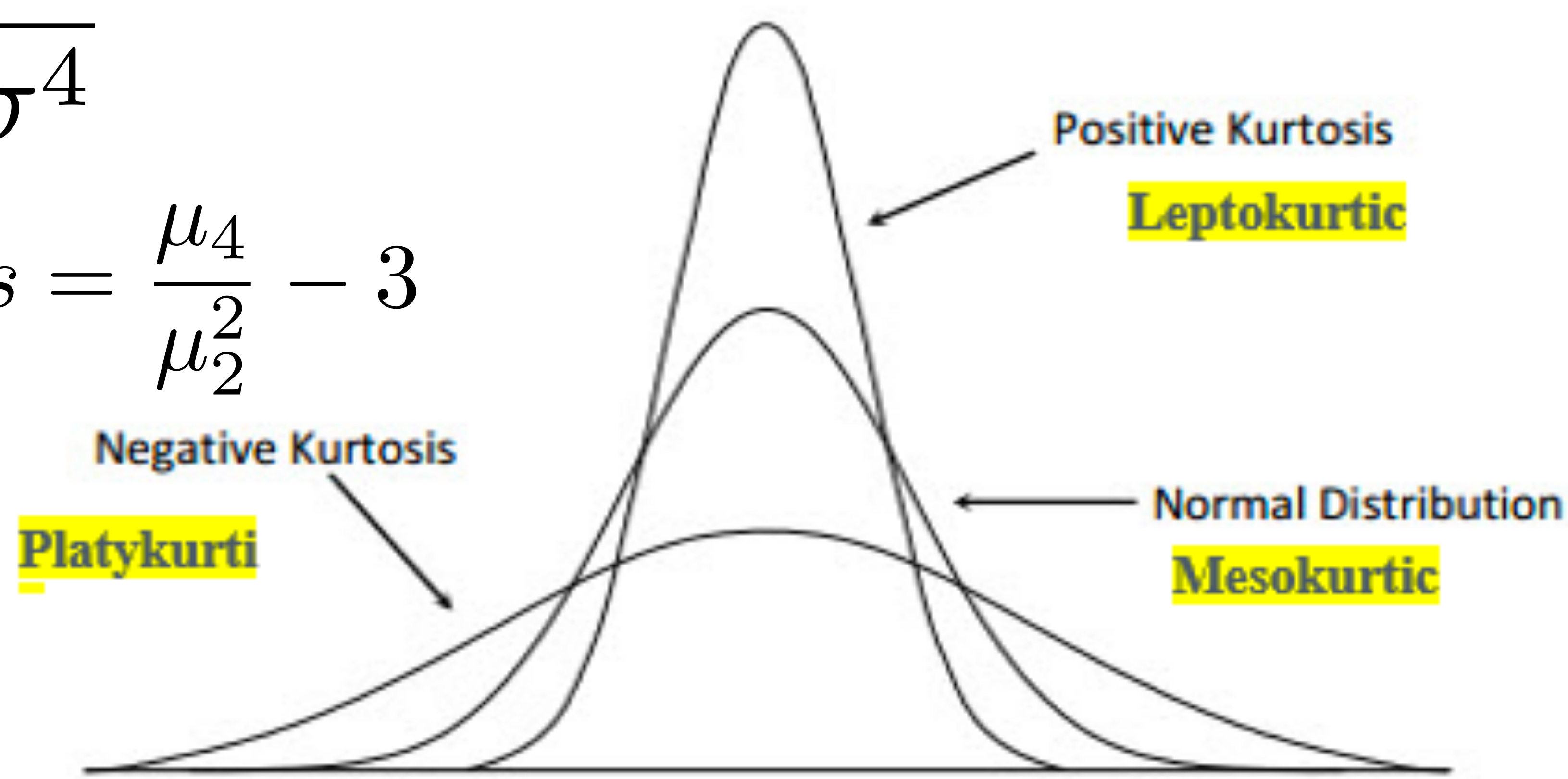
The pre factor approaches to 1 as N tends to large values

Kurtosis

Measure of “tailedness” or peakedness

$$K|X| = \frac{\mu_4}{\sigma^4}$$

$$Excess\ Kurtosis = \frac{\mu_4}{\mu_2^2} - 3$$



Positive kurtosis indicates a "heavy-tailed" distribution >> Leptokutic
Negative kurtosis indicates a "light tailed" distribution >> Platykurtic

I want you to think about these coefficients, look the formula closely, see if you can relate to a mathematical formula you have seen before

Violin plot

A Violin Plot is used to visualize the distribution of the data and its probability density.

The white dot in the middle shows the median value and the thick black bar in the centre represents the interquartile range

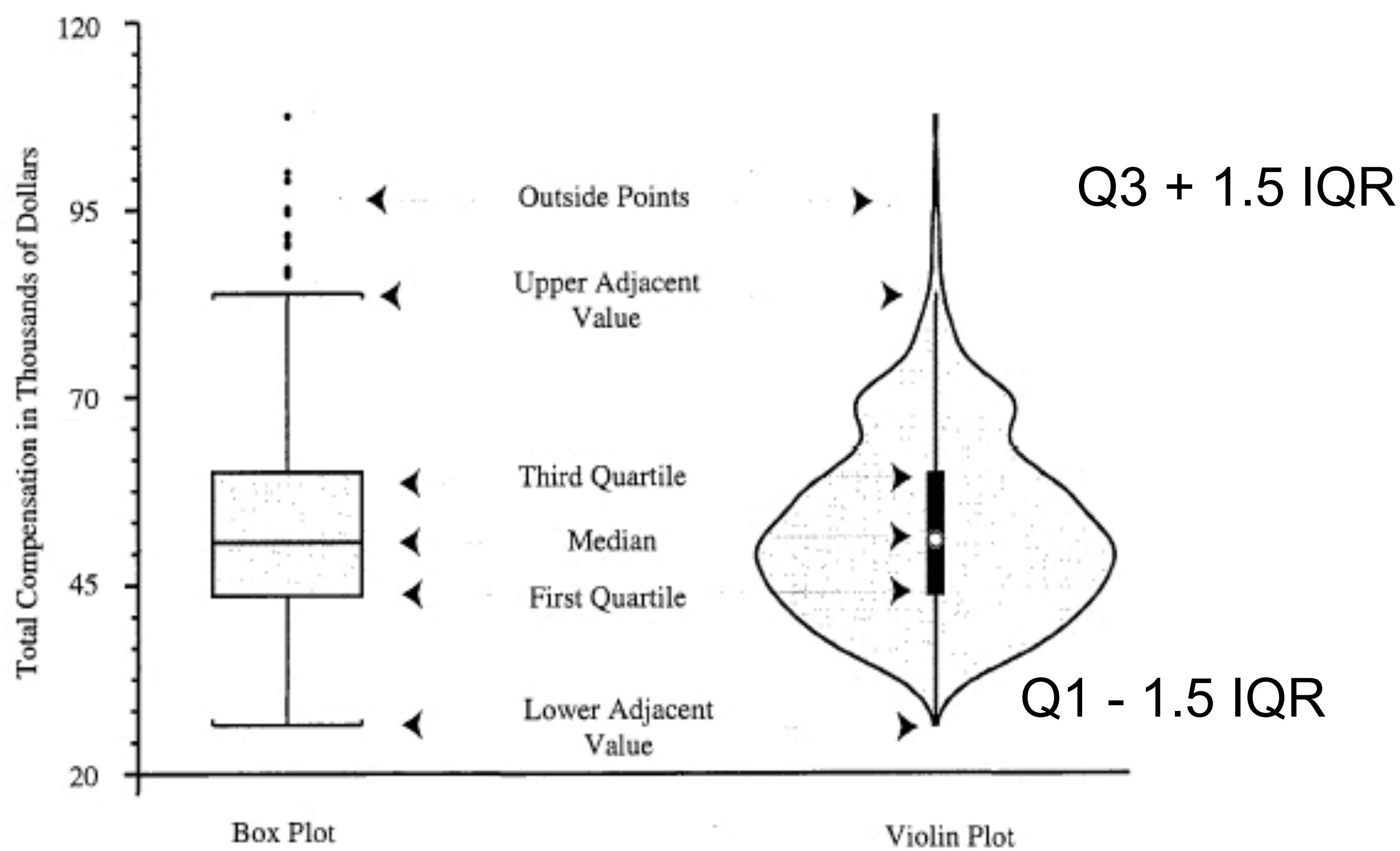


Figure 1. Common Components of Box Plot and Violin Plot. Total compensation for all academic ranks.

Reference
Towards Data Science

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Violin plot