

Type of data

Quantitative الكمية  
 Categorical التصنيفية

Categorical

ordinal  
 Nominal

Quantitative

Discrete  
 Continuous

نحو تحليل البيانات ناقش أربع جوانب رئيسية

① Centre

1. Mean  
 Median  
 Mode

② Spread

Range  
 Interquartile range IQR  
 Standard Deviation  
 Variance

Use Five number summary


Max  $Q_2$  (Median)  $Q_3$   $Q_1$  Min

Range = Max - Min


IQR =  $Q_3 - Q_1$ 

$$\text{Variance} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$
Standard =  $\sqrt{\text{Variance}}$ 

③ shape

Left Skewed 

Mean &lt; Median

Right Skewed 

Mean &gt; Median

Symmetric

Mean = Median = Mode

④ outliers نقاط تقع بعيداً عن باقي النقاط



## Simpson's Paradox

هي ظاهرة الاحتمالات والاحصاءات يظهر فيها اتجاه في عدة مجموعات مختلفة من البيانات ولكنه يختفي عند دمج هذه المجموعات

## Statistics and probability

$$P(A) = 1 - P(\bar{A})$$

نقطة

$$P(A \cdot A) = P(A) \cdot P(A)$$

عدد مرات الظهور عند تكرار الحدث  $n \cdot r$   
نقطة

## Binomial distributions

$$\frac{n!}{k!(n-k)!} \cdot p^k \cdot (1-p)^{n-k}$$

نقطة  $n \cdot k$