

Data Analysis and Visualisation

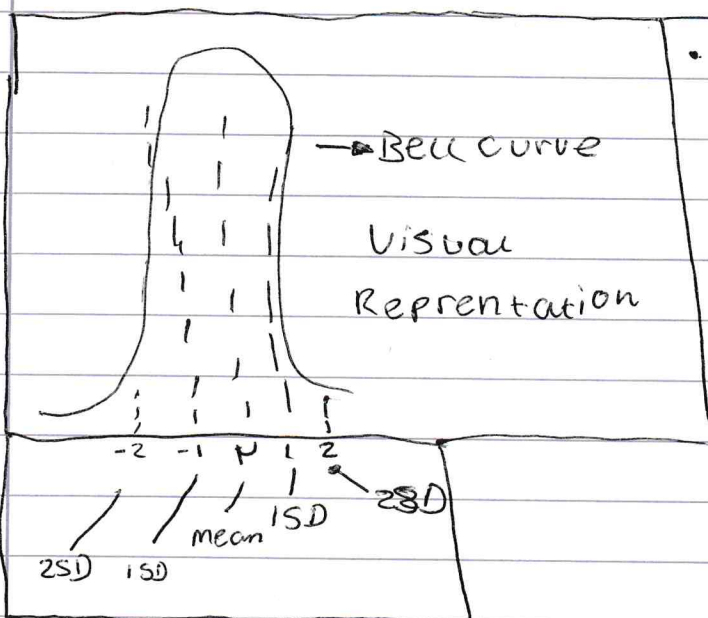
Bad Graph

- Non-labelled x/y-Axis
- Doesn't represent the same thing
- Lack in information
- Data can be misleading
- Doesn't tell you anything
- Too much information
- Too many keys
- Boring
- Lacks appearance
- Graphs / charts need purpose
- Not providing new information
- Lack of indication of most recent stats
- Non-clear ~~writes~~ writing / not accessible
- Small writing
- Translucent
- Fuzzy
- Some data is all the same. No purpose
- Useless information
-

Self-study

Standard deviation

- Measure of how data is spread out
- Shows how it is spread out around mean or average
- Bell curve (normal distribution, a tool to understand SD).
- Mean represented by μ .
- In the two dark ~~to~~ ~~sid~~ next to mean, = one ~~to~~ SD from mean
- Mean = 0, 2σ = 2 SD from mean



- Unfair to alter boundaries because of current stats compared to previous.

- Don't always want to see everything on the graph
- Visualisation not always for fine detail
- Fine detail, look at the numbers on the x & y-axis

Probability Distribution

What is it?

- Probability of an event occurring
- Show ~~for~~ simple events like coin toss
- Complex events
- Types of distribution, table, binomial, normal,
- The sum of all probability is 100% / 1
- ~~No~~ Area under curve = 100% of Data
- can be shown as, formula, table, graphs
-

Standard Distribution ^{class}

- Widely used
- ~~B~~ known as bell curve
- Natural phenomenon fit bell curve (height)
- μ = Greek word / ~~mean~~ used for mean / average
- Percentage of scores find for an SD from mean ~~are~~ ^{score}
- σ = Sigma
- ~~B~~ Bell curve is symmetrical
- Half of data on left of curve
- Half of data on right of curve

SD =
Standard
deviation

- Controls spread of distribution
- Smaller SD says that data is tightly clustered ^{around} mean,
- Normal distribution = tower

- Mean, mode, median = central
- Symmetrical at center
- Half value - left and right
- Measure of dispersment in statistics
- How much your data is spread around mean
- Mean represented by μ
- Each segment represents one SD
- 2 σ means 2SD
- Beyond 2 σ , data is outlier, blips in data
- $$S = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

- Σ = Capital Sigma / Add up ~~every~~ all values
- $\bar{x}$ = Average
- N = amount of numbers in data set
- $\sqrt{\quad}$ = Square root
- x = not yet known Value

• 12, 15, 17, 20, 30, 31, 43, 44, 54

• $\bar{x} = 9$

• $266 \div 9 = 29.555^2$

$$\begin{aligned}
 (12 - 29.556)^2 &= 308.213 + \\
 (15 - 29.556)^2 &= 211.877 + \\
 (17 - 29.556)^2 &= 157.683 + \\
 (20 - 29.556)^2 &= 91.317 + \\
 (30 - 29.556)^2 &= 0.443 + \\
 (31 - 29.556)^2 &= 1.80.7412.085 + \\
 (43 - 29.556)^2 &= 180.741 \\
 (44 - 29.556)^2 &= 208.629 + \\
 (51 - 29.556)^2 &= 597.531 = 1758.222
 \end{aligned}$$

$$1758.222 \div n(n) - 1 = 8 = \sqrt{219.778} = 14.825$$

4 6 7 11

$$A = \bar{x} = 4$$

$$n = 28 \div 4 = 14$$

$$(14 - 4) = 10 +$$

$$(14 - 6) = 8 +$$

$$(14 - 7) = 7 +$$

$$(14 - 11) = 3 +$$

$$= 28 \div 3 = \sqrt{9.333333333} = 3.05505$$

$$0.6324933$$

$$n - 1 \quad n = 4 = 3$$

HOMEWORK

1. PRACTISE STANDARD DEVIATION
2. RESEARCH BINOMIAL ~~AE~~ DISTRIBUTION