Social Isolation Indices methodology

Southampton City Council

Statistical terms & Z-Scoring

Key terms and concepts – Mean and standard deviation

southampton
dataobservatory

- **Mean: All the values added up, divided by how many values you have giving a statistical average**

- **Standard deviation (SD)– a statistical measure of how the far data is distributed from the mean**



Employees of a Cereal Factory

Mean = 10.2 yr
SD = 4.1 yr
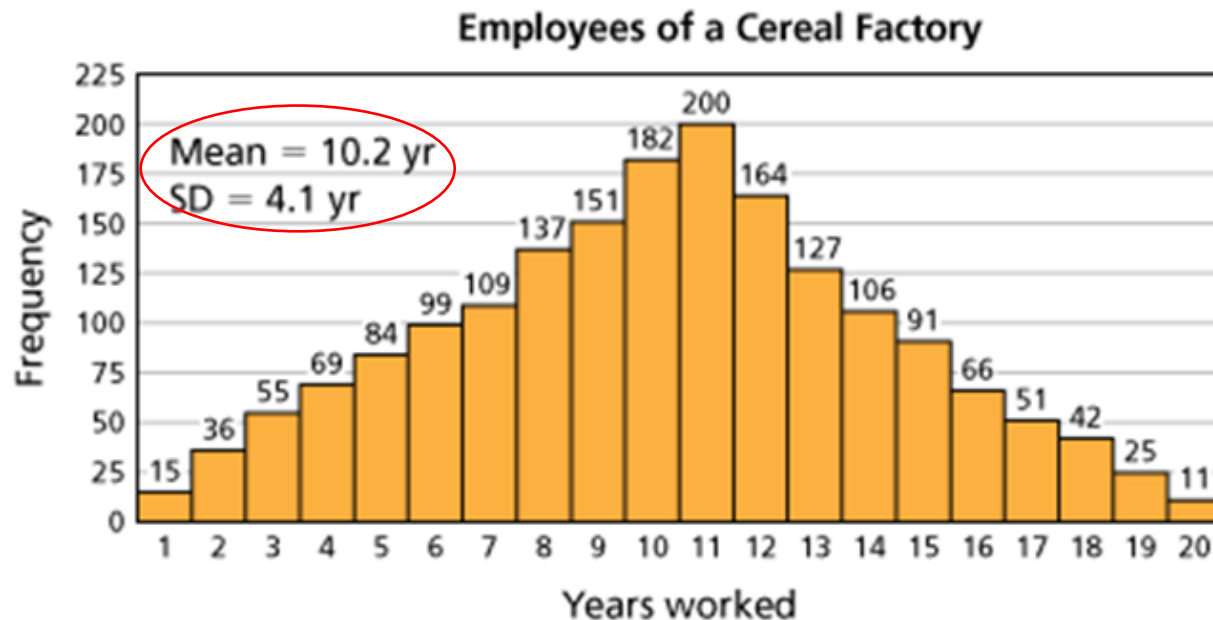
- 68.3% of the data is +/- a standard deviation distance from the mean
- One standard deviation back from the mean is 10.2 - 4.1 = 6.1
- One standard deviation forward from the mean is 10.2 + 4.1 = 14.3
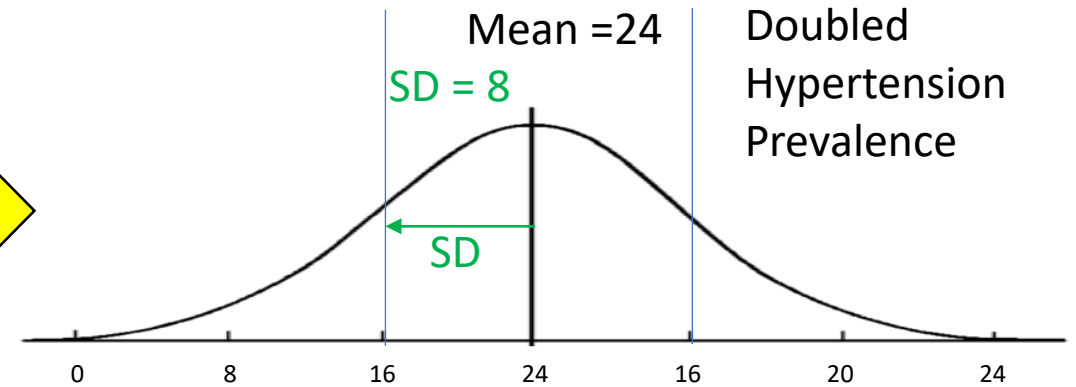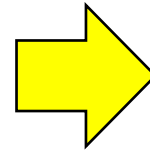- 68.3% of the employees have worked there for between 6.1 years and 14.3 years

Data can be transformed in different ways

It can be doubled…..

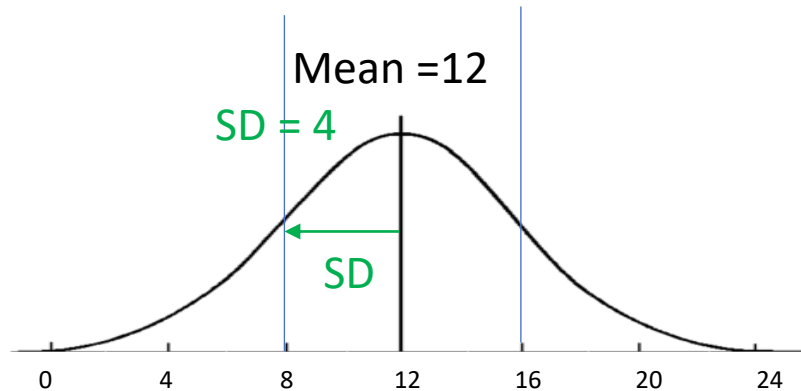Hypertension prevalence

Double each value in the data x2

Mean =12

SD = 4

SD
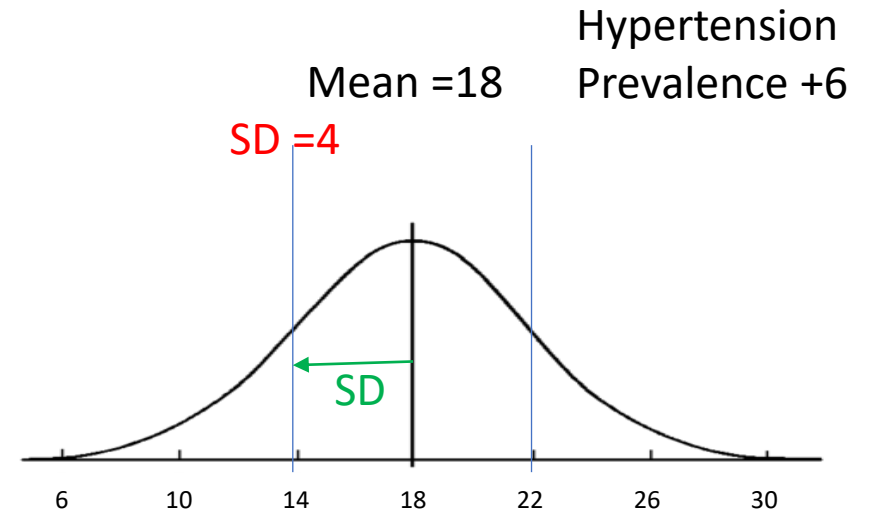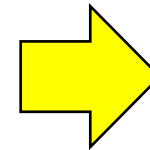
0   4   8   12   16   20   24

x2

Mean =24

SD = 8

SD

0   8   16   24   16   20   24

Doubled Hypertension Prevalence
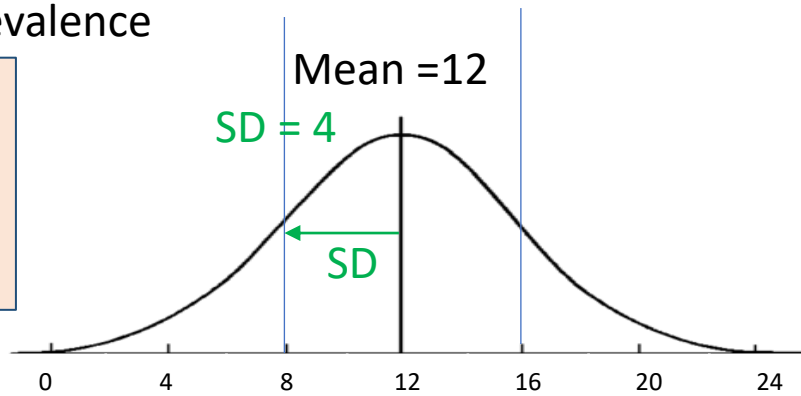
Or each value can have a factor (number) added/subtracted from it…

Add six to each value

Hypertension prevalence

Mean =12

SD = 4

SD

0   4   8   12   16   20   24

+6

Hypertension Prevalence +6

Mean =18

SD =4

SD

6   10   14   18   22   26   30
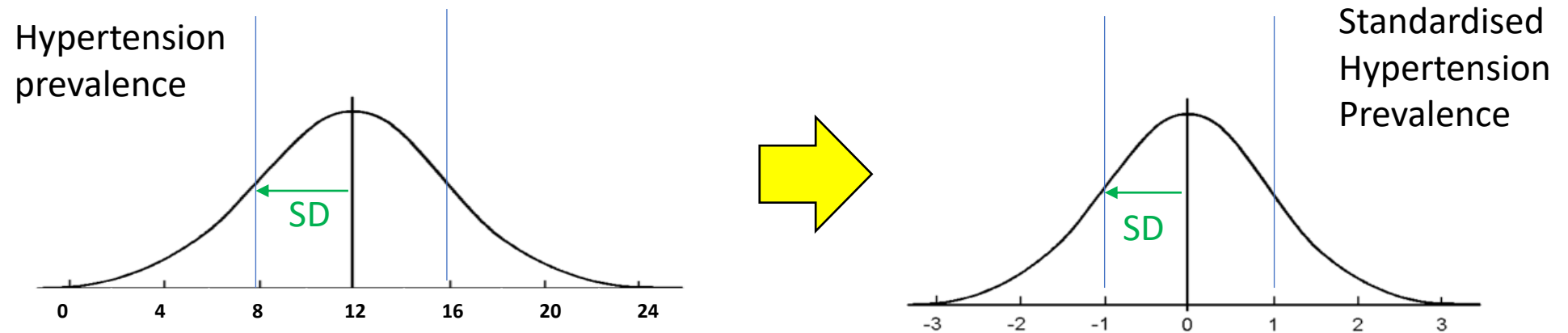
- **We can transform data using Z scoring, that transforms any distribution of data to have a mean of zero and a SD of 1**

- **Z scoring…… (value – mean)/SD…..**
- **Take the mean of 12…. 12-12 /0 =0…..**
- **Or the value of 8, one standard deviation before the mean… (8-12)/4…. -4/4….. = -1**
- **Or the value of 16, one standard deviation after the mean… (16-12)/4…. 4/4…. = -1**
- **The height of each point on the chart is the frequency, here how many LSOAs have that value, the height remains the same when transformed**
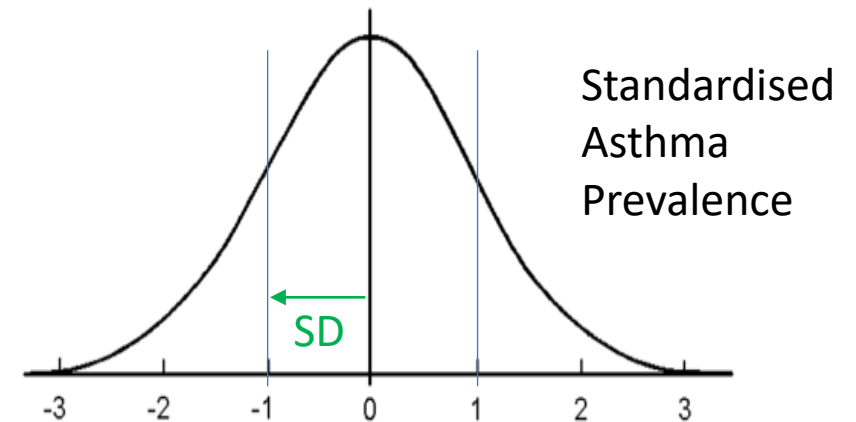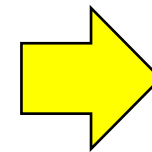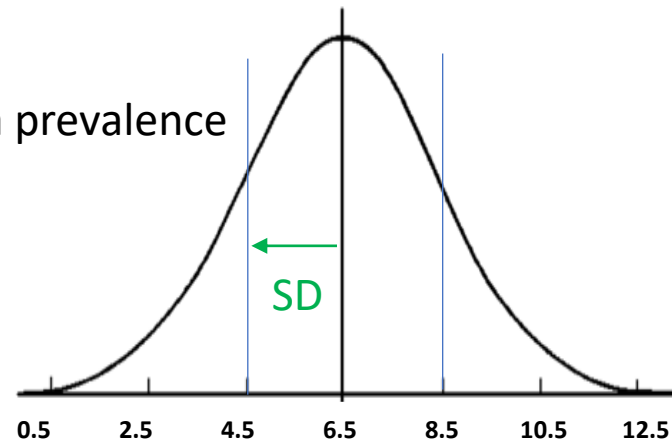
| Value | minus mean | Divide by SD | Z-score |

- Z scoring…… means all indicators have the same mean of 0 and the same standard deviation of 1, which each value kept in the same order when ranked highest to lowest as the original data set
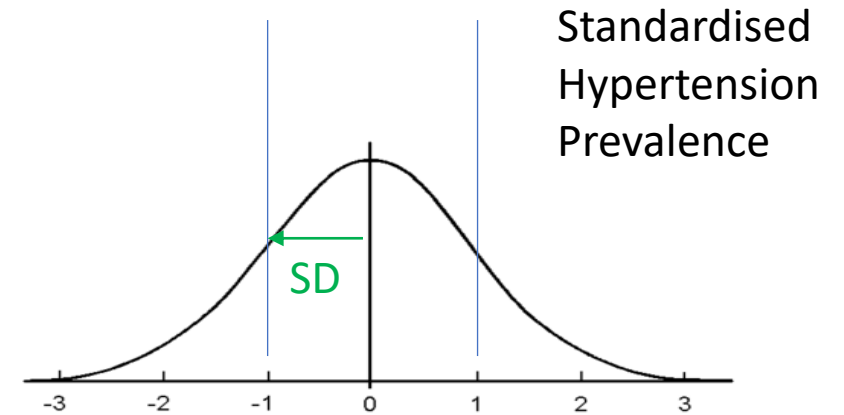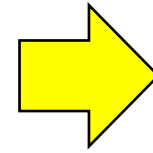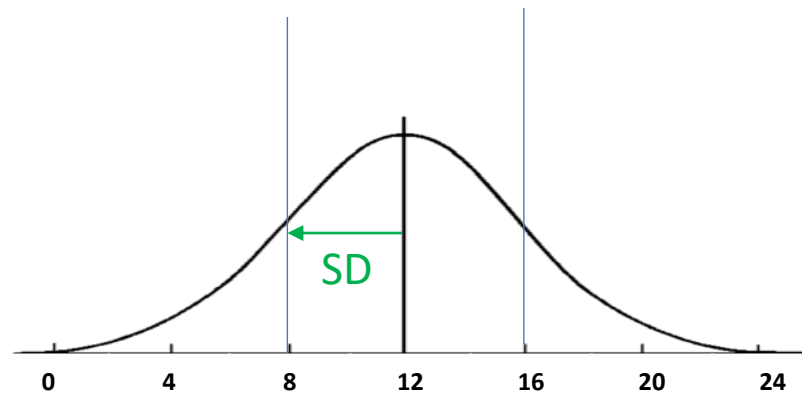
- Once on the same scale, the values of each LSOA can be added together, to make a combined overall score. You can do this for all clinical indicators to get an overall clinical domain score where each indicator is equally weighted

Pros
- It allows us to compare raw values between distributions and create relative combined domains, each indicator has an equal weighting when combined

- Allows comparison and combination of different measures, e.g. SMR (standardised mortality ratio), prevalence (%), DSR (directly age standardised rate), etc.

- Relatively simple to do and can be applied to create separate domains with multiple differing indicators

Cons
- Normal distribution is assumed (depends on Central Limit Theorem for those variables that are skewed, n>30)

- Interpretability is reduced as the z-score and combined domain scores no longer has the original unit, the unit of z-score is in SD