

Subject: Statistics STD, Z-Score

Date: _____

z-Score

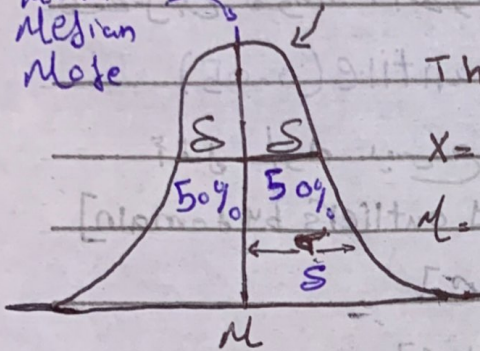
$$\mu = 0, \sigma^2 = 1$$

• normally distributed

من خلال هذا الشكل إذا كانت البيانات متوزعة بشكل طبيعي

Mean
Median
Mode

Bell curve



The Normal distribution:

$X = \text{continuous random variable}$

$\mu = \text{mean}$ $\sigma = \text{std}$ $\sigma^2 = \text{variance}$

area under curve = 1

منافذ تحال حول μ

z-Score

(Mean = 0, $\sigma = 1$) Standardize

standard deviation \Rightarrow المسافة بين البيانات

وال μ

• برسم المنحنى ده كلنا ن detect outliers

• بعد ما نوزن $z\text{-score}$ نرى

$$z\text{-score} = \frac{X - \bar{X}}{\text{std}}$$

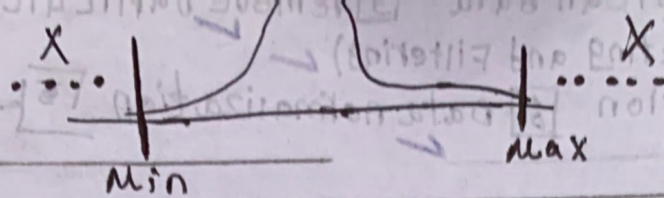
normalization \Leftarrow بتعمل

new values $\rightarrow -3 \rightarrow 3$

الاي اكبر، اقل من
خارج ال range

Alaziz

Subject: _____



Date: _____

Std method The Best thing using 3 or 4 std to remove outliers

```
# upper_limit = data['height'].mean() + 3 *  
data['height'].std()
```

```
# Lower Limit =
```

outliers detection

```
data[(data['height'] > upper_limit) | (data['height'] < lower)]
```

show all the outlier detection

Remove outliers

data_no_outliers =

```
data[(data['height'] < upper_limit) & (data['height'] > lower)]
```

DONE ✓ outliers Removed

Z-score method is The same std

$$Z = \frac{x - \bar{x}}{s}$$

range 3, -3

new col

```
data['Z-Score'] = (data['height'].mean() -  
data['height'].mean()) / data['height'].std()
```

```
data['Z-Score'] < 3 & data['Z-Score'] > -3
```

remove out

أفكس الشروط
detection

Alaziz