Unit 20 : Applications of standard deviation

Learning Objectives

The Students should be able to:

- determine the standard deviation.
- calculate standard score from a given set of data.
- determine, in the case of normal distribution, the percentages of data lying within a certain number of standard deviations from the mean.

Activities

Web Reference:

http://www.mste.uiuc.edu/hill/dstate/dstate.html

http://frey.newcastle.edu.au/Stat/stat101.html

Reference

Suen, S.N. "Mathematics for Hong Kong 5A"; rev. ed.; Chapter 5; Canotta

1 Standard deviation

For ungrouped data $x_1, x_2,\ldots,x_n$, with a mean $\bar{x}$, the standard deviation (σ) is

$$\sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+\ldots+(x_n-\bar{x})^2}{n}}$$

For grouped data with class marks $x_1, x_2,\ldots,x_n$; corresponding frequencies $f_1,f_2,\ldots,f_n$, and a mean $\bar{x}$, the standard deviation (σ) is

$$\sqrt{\frac{(x_1-\bar{x})^2 f_1+(x_2-\bar{x})^2 f_2+(x_3-\bar{x})^2 f_3+\ldots+(x_n-\bar{x})^2 f_n}{f_1+f_2+f_3+\ldots+f_n}}$$

<u>Example 1</u>

The height of Basil team members at the 2002 FIFA World Cup is listed as following:

Marcos	1.93
Cafu	1.76
Lucio	1.88
Roque Junior	1.86
Edmilson	1.85
Carlos	1.68
Richardino	1.76
Silva	1.85
Ronaldo	1.83
Rivaldo	1.86
Ronaldinho	1.80

Calculate the average height, and the standard deviation:

<u>Solution</u>

mean $\bar{x}=\sum_i \frac{x_i}{n} = 1.82$ standard deviation $s = \sqrt{\frac{\sum_i (x_i-\bar{x})^2}{n}} = 0.07$

Example 2

Find the mean and standard deviation for the data given below:

Age group x	Population ('000)
5	676
15	885
25	1000
35	1267
45	1208
55	677
65	503
75	499

Solution

mean $\bar{x} = \sum_i \frac{f_i x_i}{f_i} =$____ $\quad s = \sqrt{\frac{\sum_i f_i (x_i - \bar{x})^2}{f_i}}$ = ____ (from calculator)

2. Applications of standard deviation

2.1. Standard score

Standard scores are used to compare students' performances in different tests.

For a distribution of marks with mean $\bar{x}$ and standard deviation σ,
The standard score z is

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

Example 3

In the final examination, John obtained 70 in Mathematics and 60 in English.
His results are compared with those for the whole class.

Subject	class average	class s.d.	John's mark
Mathematics	75	5	70
English	56	2	60

a) Find John's standard scores in Mathematics and English.
b) State, for each subject, whether his performance is above or below average.

Solution

a) standard score in Mathematics = $(70 - 75) \div 5 = -1$

standard score in English = $(60 - \quad) \div$ ___ =

b) John is below average in Mathematics. (70<75)
John is above average in English. (>56)

Although John's raw mark in Mathematics is higher than that in English (70>60), he is better in English than Mathematics when compared to other students.
This is reflected by the standard scores $(2 > -1)$.

Remark:
1. Negative standard score means below average.
2. Zero standard score means at the average.
3. Positive standard score means above average.

Example 4

The statistical data of an examination are tabulated below. A student obtained 43 marks and 48 marks in Information Technology and Office Software respectively. Calculate the standard scores that the student obtained in the examination.

Subject	mean	Standard deviation
Information Technology	51.5	13.1
Office Software	63.3	14.5

Solution

Standard score $Z = (x - \bar{x}) / \sigma$

Information Technology: $Z = (43 -$ ___$)/13.1 = -\ 0.$

Office Software: $Z = (48 -$ ____$)/14.5 = -\ 1.$

Example 5

A student obtained 65 marks in Engineering Science. The mark was equal to a standard score of 3.2. Calculate the standard deviation of the marks if the mean mark was 45.

Solution

$$\text{Standard deviation } \sigma = (x - \bar{x}) / Z$$
$$= (65 - 45)/3.2 =$$

Example 6

The mean mark and the standard deviation of the marks in an examination are 58.9 and 19.4 respectively. The standard score of a student is 1.5. What is the actual mark that the student obtained? .

Solution

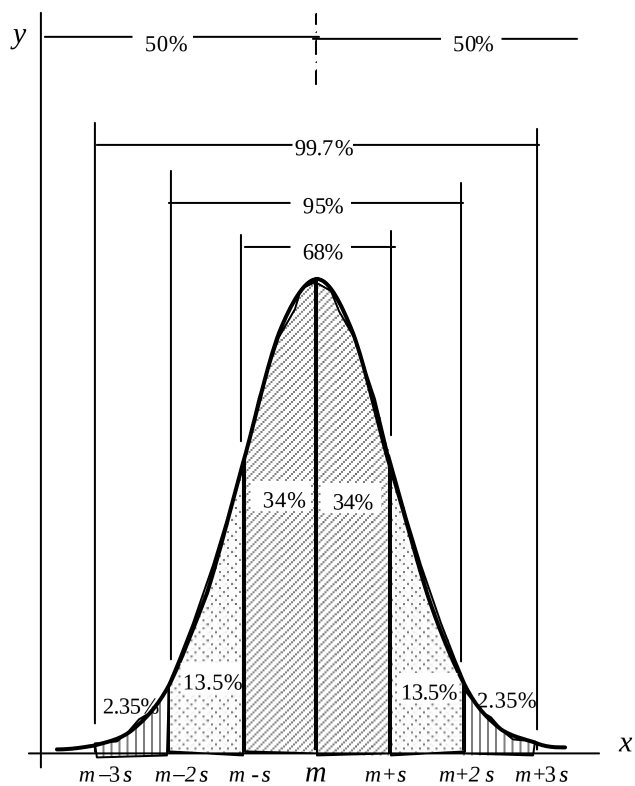
$$Z = (x - \bar{x}) / \sigma$$
$$= (x - 58.9)/19.4$$

$$x = 1.5 \times 19.4 + \_\_\_ =$$

2.2 Percentages of Normal data lying within a certain number of standard deviations from the mean

A distribution curve for a set of data is basically a frequency or relative frequency curve of the data. It is found that the distribution curves for a lot of commonly occurring data sets follow a certain pattern that came to be known as normal distributions.

A normal distribution has a bell-shaped curve as shown.



A normal curve has the following characteristics:

1. It is symmetrical about the mean.

2. Mean = mode = median. They all lie at the centre of the curve.

3. There are fewer data for values further away from the mean.

    a) about 68% of the data lie within 1 standard deviation from the mean.
    b) about 95% of the data lie within 2 standard deviations from the mean.
    c) about 99.7% of the data lie within 3 standard deviations from the mean.

Example 7

John got 70 in a Mathematics examination.
The marks of the examination are normally distributed with mean = 75 and standard deviation = 5.
If there are 100 students, how many students do better than John?

Solution

By symmetry, 50 students do better than 75 marks, 50 students do worse.
John's standard score is –1 as found in Example 1.
John is at 1 standard deviation below the mean.
There are 34 students lying within one s.d. below the mean.
Totally number of students with standard score greater than –1 is 50+34 = 84.
There are about 84 students doing better than John out of the 100 students.

Example 8

The life hours of 200 sample light bulbs are normally distributed with a mean of 1000 hours and the standard deviation is 150 hours. Find the number of light bulbs having life hours:

a) more than 700 hours;
b) less than 1150 hours.

Solution

$700 = 1000 - 2\ (\_\_\_)$
$= \bar{x} - 2\sigma$

Since 95% of light bulbs with life hours lie within $\bar{x} \pm 2\sigma$
The number of light bulbs with life hours lie between the mean and $\bar{x} - 2\sigma$ = 95%/2 = 47.5%
The number of light bulbs with life hours above the mean = 50%
The number of light bulbs with life hours more than 700 hours = 200 (47.5% + 50%) = 195

$1150 = 1000 + 1\ (150)$
$= \bar{x} + \sigma$

Since 68% of light bulbs with life hours lie within $\bar{x} \pm \sigma$
The number of light bulbs with life hours lies between the mean and $\bar{x} + \sigma$ = 68% / 2 = 34%
The number of light bulbs with life hours below the mean = 50%
The number of light bulbs with life hours less than 1150 hours = 200 (34% + 50%) = 168

Example 9

The cheesecake made by a cake factory are normally distributed with a mean of 500 gm and a standard deviation of 12 gm. How many per cent of cheesecake have weigh more than 524 gm?

Solution

$524 = 500 + 2\ (12)$
$= \bar{x} + 2\sigma$

The percentage of cheesecake with weigh more than 524 gm = (1 – 95%) /2 = 2.5%

Example 10

The length of a sample of 400 conduits are normally distributed with a mean of 2.92 m and a standard deviation of 2 cm. How many conduits in the sample have length between 2.9 m and 2.96 m?

Solution

$2.9 = 2.92 - 1$ (____)
$= \bar{x} - \sigma$

$2.96 = 2.92 + 2\ (0.02)$
$= \bar{x} + 2\sigma$

The number of conduits with length between 2.9m to 2.96m = 400 (34% +34% + ___%)
=

Example 11

The resistance of a sample of 100 resistors is normally distributed with a mean of 10kΩ. If all the resistors have a resistance of 10kΩ ± 20%, lie within 3 standard deviations from the mean. Find the standard deviation and the possible resistance lying within $\bar{x} \pm \sigma$

Solution

The possible highest resistance = 10 kΩ (1 + 20%) = 12 kΩ
$12\ k\Omega = 10\ k\Omega + 3\sigma$
$\sigma = (12 - 10) / 3 = 0.$____ kΩ

$\bar{x} + \sigma = 10 + 0.6667 = 10.6667\ k\Omega$
$\bar{x} - \sigma = 10 - 0.6667 = 9.3333\ k\Omega$