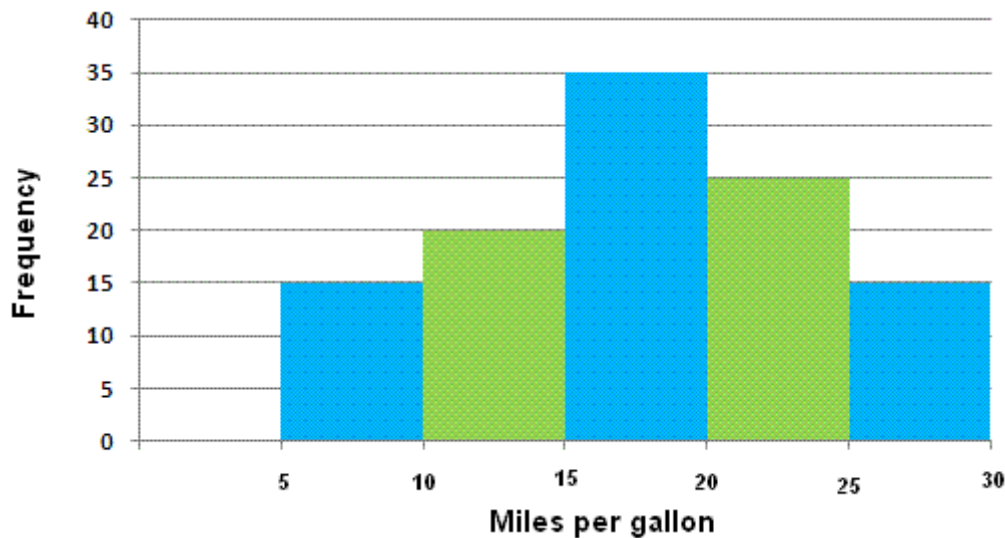


Histogram Exercise (Equal Widths)

The histogram below shows the efficiency level (in miles per gallons) of 110 cars.

- How many cars have have an efficiency between 15 and 20 miles per gallon?
- How many cars have have an efficiency more than 20 miles per gallon?
- What percentage of cars have have an efficiency less than 20 miles per gallon?



Solutions

- 35 cars
- $25 + 15 = 40$ cars
- $(15 + 20 + 35) / 110 = 0.636 = 63.6\%$

Histogram with Different Widths Example

When constructing a histogram with non-uniform (unequal) class widths, we must ensure that the areas of the rectangles are proportional to the class frequencies.

Remember that the histogram differs from a bar chart in that it is the area of the bar that denotes the value, not the height. This means that we would need to consider the widths in order to determine the height of each rectangle.

Example

The following frequency distribution gives the masses of 48 objects measured to the nearest gram. Draw a histogram to illustrate the data.

Mass (g)	10 – 19	20 – 24	25 – 34	35 – 50	51 – 55
Frequency	6	4	12	18	8

Solution:

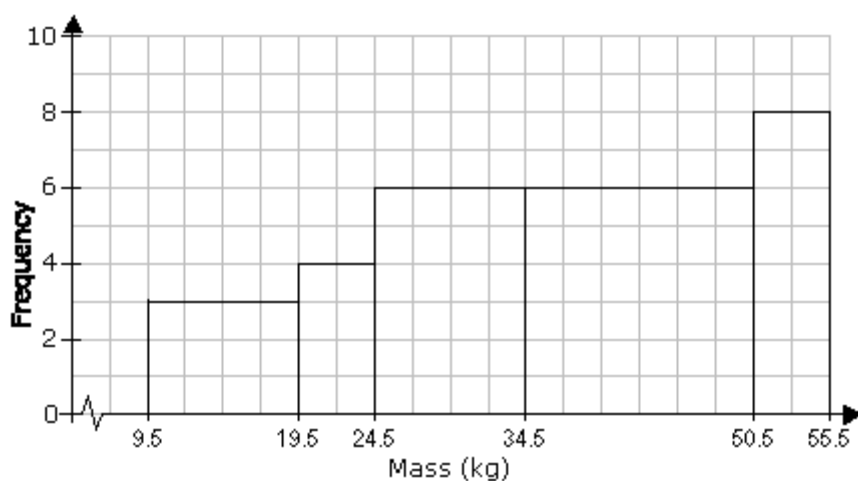
Evaluate each class widths.

Mass (g)	10 – 19	20 – 24	25 – 34	35 – 50	51 – 55
Frequency	6	4	12	18	8
Class width	10	5	10	15	5

Since the class widths are not equal, we choose a convenient width as a standard and adjust the heights of the rectangles accordingly. We notice that the smallest width size is 5. We can choose 5 to be the standard width. The other widths are then multiples of the standard width.

The table below shows the calculations of the heights of the rectangles.

Mass (g)	10 – 19	20 – 24	25 – 34	35 – 50	51 – 55
Frequency	6	4	12	18	8
Class widths	10	5	10	15	5
	2 × standard	standard	2 × standard	3 × standard	standard
Rectangle's height in histogram	$6 \div 2 = 3$	4	$12 \div 2 = 6$	$18 \div 3 = 6$	8



Pie Chart

When we represent data in a pie chart we must remember we are using 360 degrees.

Sales by store



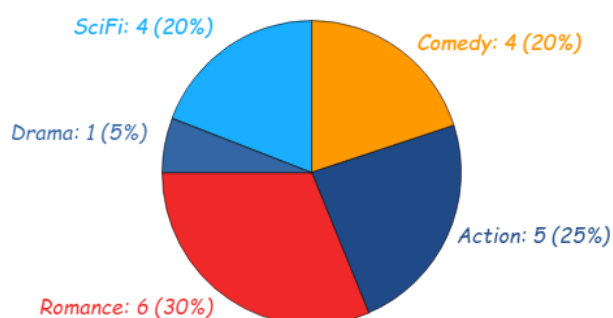
Example

Table: Favorite Type of Movie

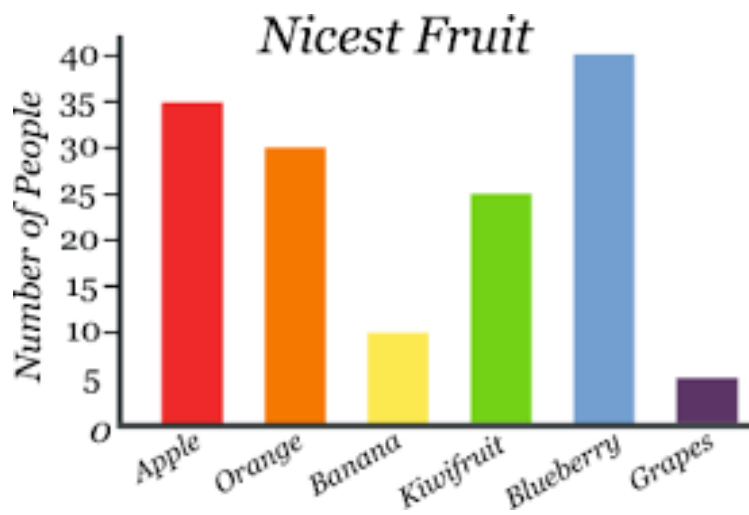
Comedy	Action	Romance	Drama	SciFi
4	5	6	1	4

Comedy	Action	Romance	Drama	SciFi	TOTAL
4	5	6	1	4	20
20%	25%	30%	5%	20%	100%
$\frac{4}{20} \times 360^\circ = 72^\circ$	$\frac{5}{20} \times 360^\circ = 90^\circ$	$\frac{6}{20} \times 360^\circ = 108^\circ$	$\frac{1}{20} \times 360^\circ = 18^\circ$	$\frac{4}{20} \times 360^\circ = 72^\circ$	360°

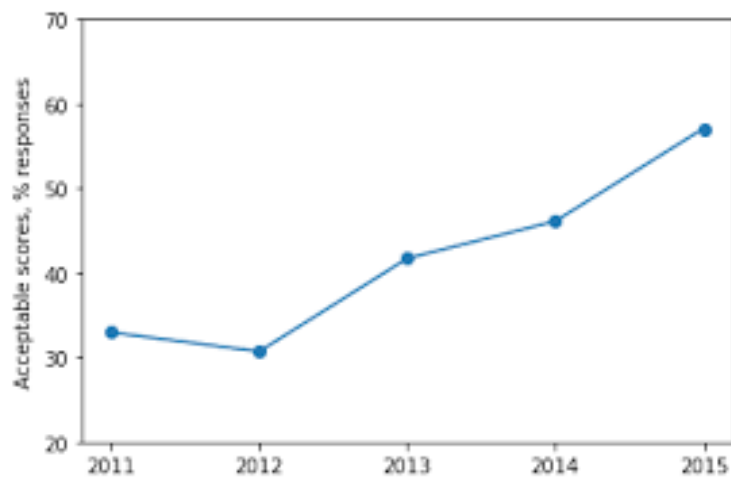
Favorite Type of Movie



Bar Chart



Trend Graph



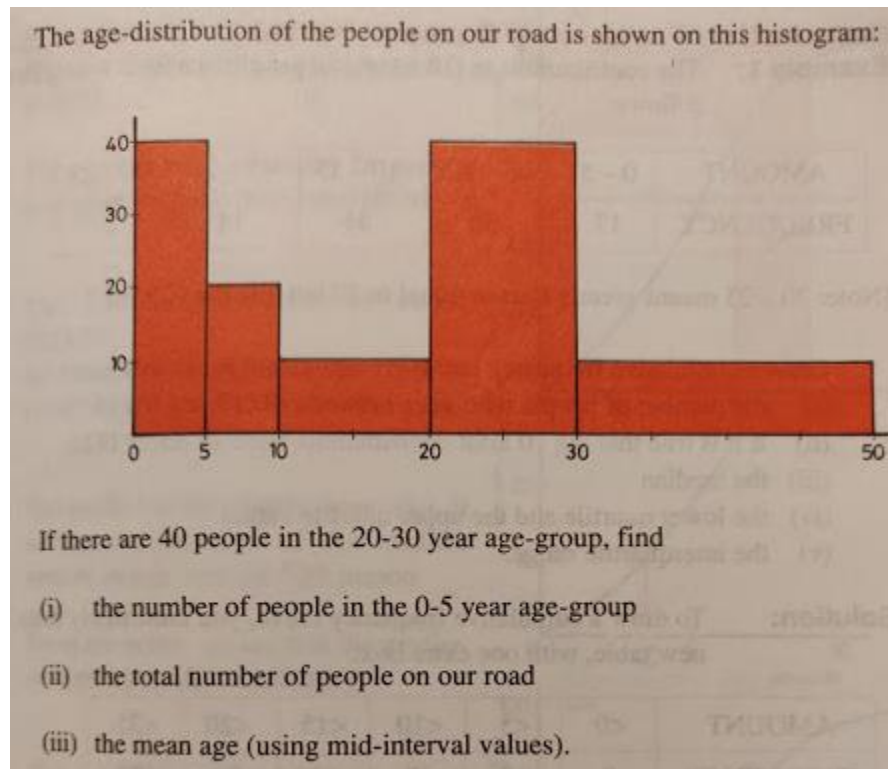
Exercise

Represent the following data of student grades in a pie chart, bar chart and trend graph.

A	B	C	D
4	12	10	2

Histograms

Question



Answer

(i) 20

(ii) 100

Mid-interval value	2.5	7.5	15	25	40	
x	0_5	5_10	10_20	20_30	30_50	
f	20	10	10	40	20	
	50	75	150	1000	800	2075
	20	10	10	40	20	100

Mean = 20.75

(iii) 20.75

Data

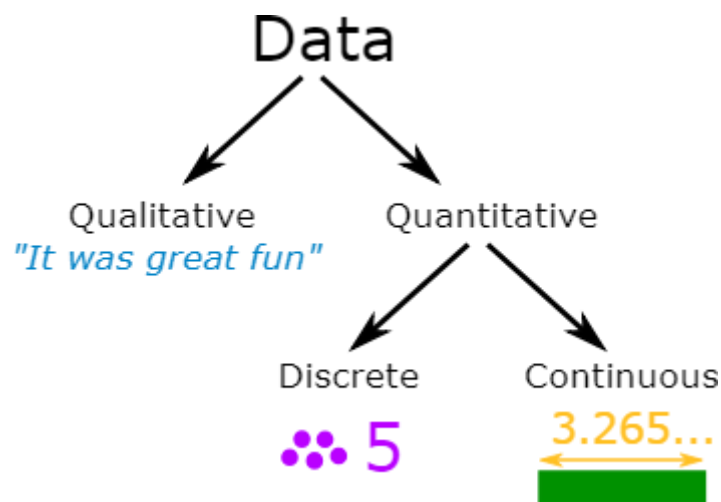
Data is a collection of facts, such as numbers, words, measurements, observations or even just descriptions of things. Data can be Descriptive (like "high" or "fast") or Numerical (numbers). Numerical Data can be Discrete or Continuous:

Population - Sample Population

In stats, a sample is a part of a population. A population is a whole, it's every member of a group. A population is the opposite to a sample, which is a fraction or percentage of a group. Sometimes it's possible to survey every member of a group. A classic example is the Census in Ireland. Note: if you do manage to survey everyone, it is called a census:

Qualitative vs Quantitative

Data can be qualitative or quantitative. Qualitative data is descriptive information (it describes something). Quantitative data is numerical information (numbers). Quantitative data can be Discrete or Continuous:



Discrete Data

Discrete Data can only take certain values. Discrete data is counted. Discrete data can only take certain values (like whole numbers)

Examples:

- The number of students in a class. We can't have half a student!
- The result of rolling a die - Only has the values 1, 2, 3, 4, 5 or 6

Continuous Data

Continuous data is measured. Continuous data can take any value (within a range)

Examples:

- A person's height: could be any value (within the range of human heights), not just certain fixed heights
- Time in a race: you could even measure it to fractions of a second
- A dog's weight
- The length of a leaf

Continuous Variables

A variable is a quantity that has a changing value; the value can vary from one example to the next. A continuous variable is a variable that has an infinite number of possible values. In other words, any value is possible for the variable. A continuous variable is the opposite of a discrete variable, which can only take on a certain number of values. A continuous variable doesn't have to have every possible number (like -infinity to +infinity), it can also be continuous between two numbers, like 1 and 2. For example, discrete variables could be 1,2 while the continuous variables could be 1,2 and everything in between: 1.00, 1.01, 1.001, 1.0001...

Examples of continuous variables: (Heights, weights are examples of continuous variables)

- Time it takes a computer to complete a task. You might think you can count it, but time is often rounded up to convenient intervals, like seconds or milliseconds. Time is actually a continuum: it could take 1.3 seconds or it could take 1.333333333333333... seconds.
- A person's weight. Someone could weigh 180 pounds, they could weigh 180.10 pounds or they could weigh 180.1110 pounds. The number of possibilities for weight are limitless.
- Income. You might think that income is countable (because it's in dollars) but who is to say someone can't have an income of a billion dollars a year? Two billion? Fifty nine trillion? And so on...
- Age. So, you're 25 years-old. Are you sure? How about 25 years, 19 days and a millisecond or two? Like time, age can take on an infinite number of possibilities and so it's a continuous variable.
- The price of gas. Sure, it might be \$4 a gallon. But one time in recent history it was 99 cents. And give inflation a few years it will be \$99. not to mention the gas stations always like to use fractions (i.e. gas is rarely \$4.47 a gallon, you'll see in the small print it's actually \$4.47 9/10ths)

Discrete Variables

A variable is a quantity that has changing values. A discrete variable is a variable that can only take on a certain number of values. In other words, they don't have an infinite number of values. If you can count a set of items, then it's a discrete variable. The opposite of a discrete variable is a continuous variable. Continuous variables can take on an infinite number of possibilities.

Examples of discrete variables:

- Number of quarters in a purse, jar, or bank. Discrete because there can only be a certain number of coins (1,2,3,4,5...). Coins don't come in amounts of 2.3 coins or 10 1/2 coins, so it isn't possible for there to be an infinite number of possibilities. In addition, a purse or even a bank is restricted by size so there can only be so many coins.
- The number of cars in a parking lot. A parking lot can only hold a certain number of cars.
- Points on a 10-point rating scale. If you're graded on a 10-point scale, the only possible values are 1,2,3,4,5,6,7,8,9, and 10.
- Ages on birthday cards. Birthday cards only come in years...they don't come in fractions. So there are a finite amount of possibilities (presumably, about one hundred).

Standard Deviation

Example 1

Calculate the mean and standard deviation of the following:

9, 2, 5, 4, 12, 7, 8, 11, 9, 3, 7, 4, 12, 5, 4, 10, 9, 6, 9, 4

Mean (μ) = $9+2+5+4+12+7+8+11+9+3+7+4+12+5+4+10+9+6+9+4 = 140/20 = 7$ Mean=7

$$(x_i - \mu)^2 \Rightarrow 4, 25, 4, 9, 25, 0, 1, 16, 4, 16, 0, 9, 25, 4, 9, 9, 4, 1, 4, 9$$

$$= 4+25+4+9+25+0+1+16+4+16+0+9+25+4+9+9+4+1+4+9 = 178$$

Variance

$$\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$$

$$\text{Variance} = (1/20) \times 178 = 8.9$$

Standard Deviation

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

$$\text{Standard Deviation } \sigma = \sqrt{(8.9)} = 2.983$$

Example 2

Occurance(x):(in separated lines)

1
2
3
4
5
6
7

Frequency of Occurance(f)

5
12
8
3
0
0
1

Occurance(X)	Frequency(f)	Freq*X	(X-mean)	(X-mean) ²	f*(X-mean) ²
1	5	5	-1.483	2.199	10.993
2	12	24	-0.483	0.233	2.797
3	8	24	0.517	0.268	2.14
4	3	12	1.517	2.302	6.906
5	0	0	2.517	6.337	0
6	0	0	3.517	12.371	0
7	1	7	4.517	20.405	20.405
Total ->	29	72	-	-	43.241

Mean = 2.483

Standard Deviation = 1.221

Variance = 1.491

Standard Deviation from list of numbers

$$\sigma = \sqrt{\frac{\sum [x - \bar{x}]^2}{n}}$$

σ = lower case sigma

\sum = capital sigma

\bar{x} = x bar

Standard Deviation from Frequency Distribution Table

$$\sqrt{\frac{\sum (X - \bar{X})^2}{N}}$$

$$\text{Standard Deviation (population)} = \sqrt{\frac{\sum_{i=1}^n (x_i - \text{mean})^2}{n}}$$

$$\text{Standard Deviation (sample)} = \sqrt{\frac{\sum_{i=1}^n (x_i - \text{mean})^2}{n-1}}$$

Frequency Distribution of a Discrete Variable

Since, a discrete variable can take some or discrete values within its range of variation, it will be natural to take a separate class for each distinct value of the discrete variable as shown in the following example relating to the daily number of car accidents during 30 days of a month.

3 4 4 5 5 3

4 3 5 7 6 4

4 3 4 5 5 5

5 5 3 5 6 4

5 4 4 6 5 6

Table No. 2: Showing frequency distribution for daily number of car accidents during a month.

Number of car accidents	Frequency
3	5
4	9
5	11
6	4
7	1
Total	30

Frequency Distribution of a Continuous Variable

For a continuous variable if we take a class for each distinct value of the variable, the number of classes will become unduly large, thus defeating the purpose of tabulation. In fact, since a continuous variable can assume an infinite number of values within its range of variation, the classification or sub-division of such data is necessarily artificial. Some guidelines that should be followed while dividing continuous data into classes are as follows:

1. The classes should be mutually exclusive, i.e., non-overlapping. No two classes should contain the same interval of values of the variable.
2. The classes should be exhaustive, i.e., they must cover the entire range of the data.
3. The number of classes and the width of each class should neither be too small nor too large. In other words, there should be relatively fewer classes if the difference between the least value of the variable and its highest value is small and relatively more classes if the same difference is large. This difference between the least value of the variable and the greatest value of the variable is called the range of the variable or the data set.
4. The classes should, preferably, be of equal width.

Let us consider the following example regarding daily maximum temperatures in $^{\circ}C$ in a city for 50 days.

28 28 31 29 35 33 28 31 34 29

25 27 29 33 30 31 32 26 26 21

21 20 22 24 28 30 34 33 35 29

23 21 20 19 19 18 19 17 20 19

18 18 19 27 17 18 20 21 18 19

Minimum Value= 17

Maximum Value=35

Range=35-17=18

Number of classes=5 (say)

\therefore width of each class=4

Table No. 3: Showing frequency distribution of temperature in a city for 50 days.

Class Intervals(Temperatures in $^{\circ}C$)	Frequency
17-20	17
21-24	7
25-28	10
29-32	9
33-36	7
Total	50

Standard Deviation

Random Sample

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{N}}$$

Calculate the standard deviation of the following numbers: 4, 2, 5, 8, 6.

Example of Frequency Distribution Table

Frequency Distribution

$$\sigma = \sqrt{\frac{\sum f \cdot (x_i - \mu)^2}{\sum f}}$$

Calculate the standard deviation from the following:

	f	$x - \bar{x}$	$(x - \bar{x})^2$	$(x - \bar{x})^2 f$
100	1	11.2	125.44	125.44
98	1	9.2	84.64	84.64
95	2	6.2	38.44	76.88
94	1	5.2	27.04	27.04
85	3	-3.8	14.44	43.32
77	1	-11.8	139.24	139.24
74	1	-14.8	219.04	219.04
totals	10			715.60

Standard Deviation Practice Problems (with answers)

1. Consider the following three data sets A, B and C.

$$A = \{9, 10, 11, 7, 13\}$$

$$B = \{10, 10, 10, 10, 10\}$$

$$C = \{1, 1, 10, 19, 19\}$$

- Calculate the mean of each data set.
- Calculate the standard deviation of each data set.
- Which set has the largest standard deviation?

2. The frequency table of the monthly salaries of 20 people is shown below.

Salary (in \$)	Number of people with this salary
3500	5
4000	8
4200	5
4300	2

- Calculate the mean of the salaries of the 20 people.
- Calculate the standard deviation of the salaries of the 20 people.

ANSWERS:

1.

a. mean of Data set A = $(9+10+11+7+13)/5 = 10$

mean of Data set B = $(10+10+10+10+10)/5 = 10$

mean of Data set C = $(1+1+10+19+19)/5 = 10$

b.

Standard Deviation Data set A

$$= \sqrt{[(9-10)^2 + (10-10)^2 + (11-10)^2 + (7-10)^2 + (13-10)^2]/5} = 2$$

Standard Deviation Data set B

$$= \sqrt{[(10-10)^2 + (10-10)^2 + (10-10)^2 + (10-10)^2 + (10-10)^2]/5} = 0$$

Standard Deviation Data set C

$$= \sqrt{[(1-10)^2 + (1-10)^2 + (10-10)^2 + (19-10)^2 + (19-10)^2]/5} = 8.05$$

c. Data set C has the largest standard deviation.

2.

a. Mean= \$3955

b. standard deviation= 282 (rounded to the nearest unit)

Exercise – Student Ages Data

A survey was carried out in a Further Education college where students were asked what age they were on the first day of college. The results were counted and presented in a grouped frequency table as shown below:

Age	18_28	28_38	38_48	48_58	58_68	68_78
No. of Students	131	172	152	32	12	1

You are required to answer the following:

1. What is the modal class?
2. Calculate the mean.
3. Calculate the standard deviation.

Solution

Mean

Mid interval
value

23 33 43 53 63 73

Age	18_28	28_38	38_48	48_58	58_68	68_78
No. of Students	131	172	152	32	12	1

Sum of f -> 500

f by mid interval 3013 5676 6536 1696 756 73 Sum of fx -> 17750 Mean(μ)= 35.5

$$\text{Mean} = \frac{\sum fx}{\sum f}$$

Standard Deviation

x	f	Mean	(x-M)^2	f(x-M)^2
23	131	35.5	156.25	20468.8
33	172	35.5	6.25	1075
43	152	35.5	56.25	8550
53	32	35.5	306.25	9800
63	12	35.5	756.25	9075
73	1	35.5	1406.25	1406.25
	500			50375

St.
Dev(σ)= 10.0374

Standard Deviation formula $\sigma = \sqrt{\frac{\sum f(x-\mu)^2}{\sum f}}$ (μ represents the Mean)

Answers

1. Modal class = 28 to 38 years old
2. Mean = 35.5 years old
3. Standard deviation = 10.0374 years old

Microsoft Excel

A	B	C	D	E	F	G	H	I	J	K
1	A									
2										
3	x	7	8	9	10	11	12	13		
4	f	1	3	5	7	5	3	1		
5										
6	Mean => f X x	7	24	45	70	55	36	13	250	<= sum c
7		1	3	5	7	5	3	1	25	<= sum c
8										
9	Standard deviation =>(x-M)	-3	-2	-1	0	1	2	3		
10	(x-M)^2	9	4	1	0	1	4	9	28	<= sum c
11	f(x-M)^2	9	12	5	0	5	12	9	52	<= sum c M)^2
12										
13	Standard Deviation=>	1.44222051								
14										
15	FORMULA									
16	x	7	8	9	10	11	12	13		
17	f	1	3	5	7	5	3	1		
18										
19	Mean => f X x	=C16*C17	=D16*D17	=E16*E17	70	55	36	13	=SUM(C19:I19)	<= sum c
20		=C17	=D17	=E17	7	5	3	1	=SUM(C20:I20)	<= sum c
21										
22	Standard deviation =>(x-M)	=C16-\$M\$6	=D16-\$M\$6	=E16-\$M\$6	0	1	2	3		
23	(x-M)^2	=C22^2	=D22^2	=E22^2	0	1	4	9	=SUM(C23:I23)	<= sum c
24	f(x-M)^2	=C23*C17	=D23*D17	=E23*E17	0	5	12	9	=SUM(C24:I24)	<= sum c M)^2
25										
26	Standard Deviation =>	=SQRT(M24)								