Measure of Dispersion

The following table shows the distribution of 100 families according to their expenditure per week. Number of families corresponding to the expenditure groups Tk. 10-20 (hundreds) and Tk. 30-40 (hundreds) are missing from the table. The median and mode are given to be Tk. 25 (hundreds) and 24 (hundreds). Calculate the missing frequencies and the calculate the arithmetic mean of the data

Expenditure (hundreds)	Number of Families
0 – 10	14
10 – 20	?
20 – 30	27
30 – 40	?
40 – 50	15

$$Mode = LB + \left(\frac{\Delta 1}{\Delta 1 + \Delta 2}\right)(i)$$

$$\text{Median} = \text{LB} + \left(\frac{\frac{N}{2} - cf}{f}\right)(i)$$

Class Limits	f	Cf
0 – 10	14	14
10 – 20	f1	14+f1
20 – 30	27	41+f1
30 – 40	f2	41+f1+f2
40 – 50	15	56+f1+f2
	100	

$$25 = 20 + \frac{\frac{56 + f1 + f2}{2} - (14 + f1)}{27}(10)$$

and

$$24 = 20 + \frac{27 - f1}{27 - f1 + 27 - f2}(10)$$

Dispersion

- ✓ **The difference between the actual value and the average value.**

Measures of Dispersion
Range
Average Deviation
Variance
Standard Deviation

Measures of Location
Quartiles
Deciles
Percentiles
Midhinge
Interquartile Range
Quartile Deviation

Range

- ✓ The difference of the highest value and the lowest value in the data set.

Example: The daily rates of a sample of eight employees at GMS Inc. are 550, 420, 560, 500, ₱700, 670, 860, 480. Find the range.

Solution:

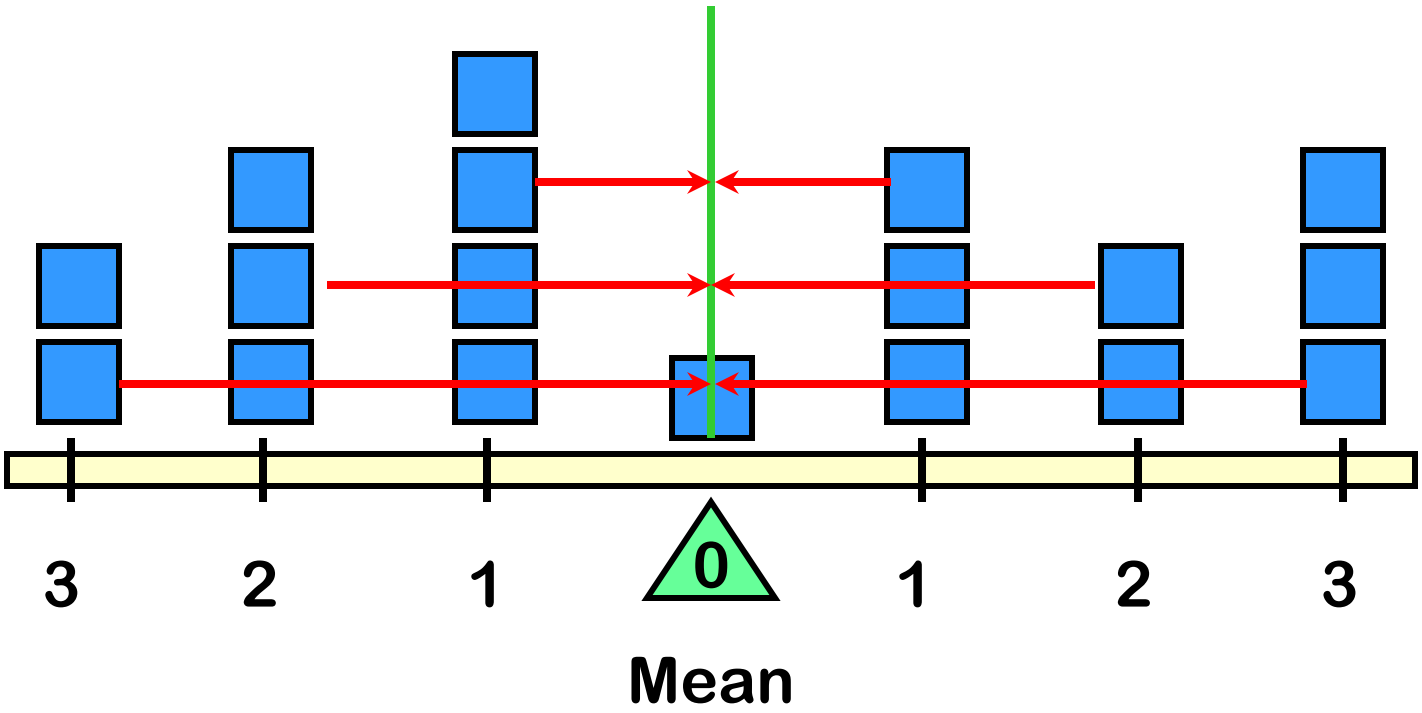
420, 480, 500, 550, 560, 670, 700, 860

HV = 860 and LV = 420

Range = HV – LV = 860 – 420 = 440

Average Deviation

- ✓ **It is the absolute difference between the element and a given point.**



Average Deviation for Ungrouped Data

$$AD = \frac{\sum |X - \mu|}{N}$$

Individual value

Sample mean

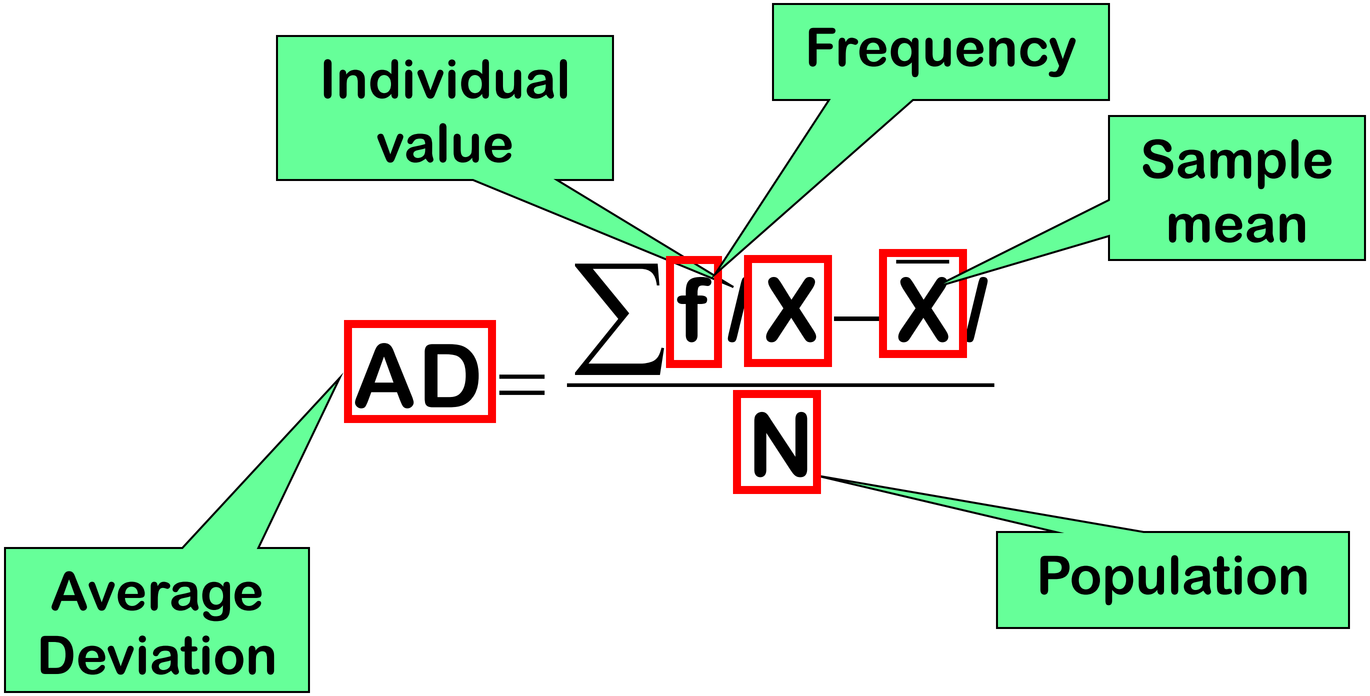
Average Deviation

Population

Sample AD:

$$AD = \frac{\sum |X - \bar{X}|}{n}$$

Average Deviation for Grouped Data



Sample AD:

$$AD = \frac{\sum f\,|X - \bar{X}|}{n}$$

Example 1 for Average Deviation

The daily rates of a sample of eight employees at GMS Inc. are 550, 420, 560, 500, 700, 670, 860, 480. Find the average deviation.

Solution:

Compute for the mean

$$\bar{X} = \frac{\sum X}{n} = \frac{550 + 420 + 560 + 500 + 700 + 670 + 860 + 480}{8}$$

$$= \frac{4,740}{8} = \boxed{592.50}$$

Solution: AD Ungrouped Data

550 – 592.5

X	$X-\bar{X}$	$/X-\bar{X}/$
550	–42.5	42.5
420	–172.5	172.5
560	–32.5	32.5
500	–92.5	92.5
700	107.5	107.5
670	77.5	77.5
860	267.5	267.5
480	–112.5	112.5
$\sum X = 4{,}740$	$\sum (X-\bar{X}) = 0$	$\sum /X-\bar{X}/ = 905$

$$AD = \frac{\sum /X-\bar{X}/}{n} = \frac{905}{8} = 113.125 \approx 113.13$$

Example: AD Grouped Data

The data below shows the frequency distribution of the amounts of electric consumption of a typical household in Batangas City for the month of January 2009. Find the average deviation.

Amount of Electric Bill	Number of Families
700 – 849	2
850 – 999	9
1,000 – 1,149	15
1,150 – 1,299	9
1,300 – 1,499	5

Solution: AD Grouped Data

Class Limits	f	X	fX
700 – 849	2	774.5	1,549.00
850 – 999	9	924.5	8,320.50
1,000 – 1,149	15	1,074.5	16,117.50
1,150 – 1,299	9	1,224.5	11,020.50
1,300 – 1,499	5	1,374.5	6,872.50
	40		$\sum fX = 43,880$

Midpoint

Add

$$\bar{X} = \frac{\sum fX}{n} = \frac{43,880}{40} = 1,097$$

Solution (continuation)

774.5 – 1,097

$f|X-\bar{X}|$ Mean

Class Limits	f	X	$X-\bar{X}$	$\lvert X-\bar{X}\rvert$	$f\lvert X-\bar{X}\rvert$
700 – 849	2	774.5	–322.5	322.5	645.0
850 – 999	9	924.5	–172.5	172.5	1,552.5
1,000 – 1,149	15	1,074.5	–22.5	22.5	337.5
1,150 – 1,299	9	1,224.5	127.5	127.5	1,147.5
1,300 – 1,499	5	1,374.5	277.5	277.5	1,387.5
	40			$\sum f\lvert X-\bar{X}\rvert = 5{,}070$	

$$AD = \frac{\sum f|X-\bar{X}|}{n} = \frac{5{,}070}{40} = 126.75$$

Standard Deviation

- ✓ It is a statistical term that provides a good indication of volatility.

Measure of risk

- ✓ It measures how widely values are dispersed from the average.

- ✓ It is calculated as the square root of variance.

Range Rule of Thumb

✓ It is used to approximate or to give a rough estimate of the standard deviation

Standard Deviation

$$s \approx \frac{\text{range}}{4}$$

Example: The daily rates of a sample of eight employees at GMS Inc. are 550, 420, 560, 500, 700, 670, 860, 480. Find the range.

Solution:

$$s \approx \frac{\text{range}}{4} = \frac{HV - LV}{4} = \frac{860 - 420}{4} = \frac{440}{4} = \boxed{110}$$

Variance

- ✓ It is a mathematical expectation of the average squared deviations from the mean.

Sample Variance:

For ungrouped data

$$s^2 = \frac{\sum (X - \bar{X})^2}{n-1}$$

or

$$s^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}$$

For grouped data

$$s^2 = \frac{\sum f(X - \bar{X})^2}{n-1}$$

or

$$s^2 = \frac{\sum fX^2 - \frac{(\sum fX)^2}{n}}{n-1}$$

Variance

For ungrouped data

$$s = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}}$$

or

$$s = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}}$$

For grouped data

$$s = \sqrt{\frac{\sum f(X - \bar{X})^2}{n-1}}$$

or

$$s = \sqrt{\frac{\sum fX^2 - \frac{(\sum fX)^2}{n}}{n-1}}$$

Population Variance and SD

Sample Variance

Sample Variance

Individual value

Sample mean

$$s^2 = \frac{\sum (X - \bar{X})^2}{n - 1}$$

Sample Population

Sample Standard Deviation

Sample SD

$$s = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}}$$

Example: SD & Variance Ungrouped Data

The daily rates of a sample of eight employees at GMS Inc. are 550, 420, 560, 500, 700, 670, 860, 480. Find the average deviation.

Solution:

Compute for the mean

$$\bar{X} = \frac{\sum X}{n} = \frac{550 + 420 + 560 + 500 + 700 + 670 + 860 + 480}{8}$$

$$= \frac{4{,}740}{8} = \boxed{592.50}$$

Solution 1: SD & Variance Ungrouped Data

X	$X-\bar{X}$	$(X-\bar{X})^2$
550	–42.5	1,806.25
420	–172.5	29,756.25
560	–32.5	1,056.25
500	–92.5	8,556.25
700	107.5	11,556.25
670	77.5	6,006.25
860	267.5	71,556.25
480	–112.5	12,656.25
$\sum X = 4{,}740$	$\sum(X-\bar{X}) = 0$	$\sum / X-\bar{X} / = 142{,}950$

550 – 592.5

$$s^2 = \frac{\sum(X-\bar{X})^2}{n-1} = \frac{142{,}950}{8-1} = 20{,}421.43$$

$$s = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}} = \sqrt{\frac{142{,}950}{8-1}} = \sqrt{20{,}421.43} = 142.90$$

Solution 2: SD & Variance Ungrouped Data

X	X^2
550	302,500
420	176,400
560	313,600
500	250,000
700	490,000
670	448,900
860	739,600
480	230,400
$\sum X = 4{,}740$	$\sum X^2 = 2{,}951{,}400$

$$s^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}$$

$$= \frac{2{,}951{,}400 - \frac{(4{,}740)^2}{8}}{8-1}$$

$$= \frac{2{,}951{,}400 - 2{,}808{,}450}{7}$$

$$= \boxed{20{,}421.43}$$

$$s = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}} = \sqrt{20{,}421.43} = \boxed{142.90}$$

Example: Variance & SD Grouped Data

Using the example provided in Chapter 2 on SJS Travel Agency. Determine the variance and standard of the frequency distribution on the ages of 50 people taking travel tours.

Class Limits	Frequency
18 – 26	3
27 – 35	5
36 – 44	9
45 – 53	14
54 – 62	11
63 – 71	6
72 – 80	2

Solution 1:

Class Limits	f	X	fX
18 – 26	3	22	66
27 – 35	5	31	155
36 – 44	9	40	360
45 – 53	14	49	686
54 – 62	11	58	638
63 – 71	6	67	402
72 – 80	2	76	152
Total	50		$\sum fX = 2{,}459$

$$\bar{X} = \frac{\sum fX}{n} = \frac{2{,}459}{50} = \boxed{49.18}$$

Solution 1 (continuation)

Class Limits	f	X	$X-\bar{X}$	$(X-\bar{X})^2$	$f(X-\bar{X})^2$
18 – 26	3	22	–27.18	738.7534	2,216.2572
27 – 35	5	31	–18.18	330.5142	1,652.5620
36 – 44	9	40	–8.18	84.2724	758.4516
45 – 53	14	49	–0.18	0.0324	0.4536
54 – 62	11	58	8.82	77.7924	855.7164
63 – 71	6	67	17.82	317.5524	1,905.3144
72 – 80	2	76	26.82	719.3124	1,438.6248
Total	50				$\sum f(X-\bar{X})^2 = 8{,}827.3800$

Variance: $$s^2 = \frac{\sum f(X-\bar{X})^2}{n-1} = \frac{8{,}827.38}{50-1} = \boxed{180.15}$$

Standard Deviation: $$s = \sqrt{\frac{\sum f(X-\bar{X})^2}{n-1}} = \sqrt{180.15} = \boxed{13.42}$$

Solution 2:

Class Limits	f	X	fX	fX²
18 – 26	3	22	66	66
27 – 35	5	31	155	155
36 – 44	9	40	360	360
45 – 53	14	49	686	686
54 – 62	11	58	638	638
63 – 71	6	67	402	402
72 – 80	2	76	152	152
Total	50		$\sum fX = 2{,}459$	$\sum fX^2 = 129{,}761$

$$s^2 = \frac{\sum fX^2 - \frac{(\sum fX)^2}{n}}{n-1} = \frac{129{,}761 - \frac{(2{,}459)^2}{50}}{50-1} = \frac{129{,}761 - 120{,}933.92}{49} = \boxed{180.15}$$

$$s = \sqrt{\frac{\sum fX^2 - \frac{(\sum fX)^2}{n}}{n-1}} = \sqrt{180.15} = \boxed{13.42}$$

Population Variance and SD

Population Variance

Population Variance

$$\sigma^2 = \frac{\sum (X - \mu)^2}{N}$$

Individual value

Population mean

Population

Population Standard Deviation

Population SD

$$\sigma = \sqrt{\frac{\sum (X - \mu)^2}{N}}$$

Example for Variance & SD

The monthly incomes of the five research directors of Recoletos schools are: 55,000, 59,500, 62,500, 57,000, and 61,000. Find the variance and standard deviation.

Solution:

Compute for the mean

$$\mu = \frac{\sum X}{N} = \frac{55{,}000 + 59{,}500 + 62{,}500 + 57{,}000 + 61{,}000}{5}$$

$$= \frac{295{,}000}{5} = \boxed{59{,}000}$$

Solution for Variance & SD

X	$X-\mu$	$(X-\mu)^2$
55,000	–4,000	16,000,000
59,500	500	250,000
62,500	3,500	12,250,000
57,000	–2,000	4,000,000
61,000	2,000	4,000,000
$\sum X = 295{,}000$	$\sum (X-\mu) = 0$	$\sum (X-\mu)^2 = 36{,}500{,}000$

$$\sigma^2 = \frac{\sum (X-\mu)^2}{N} = \frac{36{,}500{,}000}{5} = 730{,}000$$

$$\sigma = \sqrt{\frac{\sum (X-\mu)^2}{N}} = \sqrt{730{,}000} = 2{,}701.85$$

Quartiles, Deciles & Percentiles

	Ungrouped	Grouped
Quartiles	$Q_k = \frac{k(N+1)}{4}$	$Q_k = LB + \left(\frac{\frac{kN}{4} - cf}{f} \right)(i)$
Deciles	$D_k = \frac{k(N+1)}{10}$	$D_k = LB + \left(\frac{\frac{kN}{10} - cf}{f} \right)(i)$
Percentiles	$P_k = \frac{k(N+1)}{100}$	$P_k = LB + \left(\frac{\frac{kN}{100} - cf}{f} \right)(i)$

Example for Quartiles

Find the 1st, 2nd, and 3rd quartiles of the ages of 9 middle-management employees of a certain company. The ages are 53, 45, 59, 48, 54, 46, 51, 58, and 55.

Solution:

$$Q_1 = \frac{1(N+1)}{4} = \frac{1(9+1)}{4} = \frac{10}{4} = 2.5$$

$$Q_2 = \frac{2(N+1)}{4} = \frac{2(9+1)}{4} = \frac{2(10)}{4} = 5$$

$$Q_3 = \frac{3(N+1)}{4} = \frac{3(9+1)}{4} = \frac{3(10)}{4} = 7.5$$

Example for Quartiles

45, 46, 48, 51, 53, 54, 55, 58, 59

2.5th 5th 7.5th

$$Q_1 = \frac{46+48}{2} = \frac{94}{2} = 47$$

$$Q_3 = \frac{55+58}{2} = \frac{113}{2} = 56.5$$

Therefore, $Q_1 = 47$, $Q_2 = 53$, and $Q_3 = 56.5$.

Example for Q_k, D_k, P_k

Using the example provided in on SJS Travel Agency. Determine the Q_1, Q_2, Q_3, D_7, P_{22} of the frequency distribution on the ages of 50 people taking travel tours.

Class Limits	Frequency
18 – 26	3
27 – 35	5
36 – 44	9
45 – 53	14
54 – 62	11
63 – 71	6
72 – 80	2

Solution for Q_1

$$Q_1(\text{Ranked Value}) = \frac{N}{4} = \frac{50}{4} = 12.5$$

Class Limits	f	cf
18 – 26	3	3
27 – 35	5	8
36 – 44	9	17
45 – 53	14	31
54 – 62	11	42
63 – 71	6	48
72 – 80	2	50
Total	50	

LB = 36 – 0.5
= 35.5

cf

f

Q_1 Class

$$Q_1 = LB + \left(\frac{\frac{N}{4} - cf}{f}\right)(i) = 35.5 + \left(\frac{\frac{50}{4} - 8}{9}\right)(9) = 40$$

Solution for Q_2

$$Q_2(\text{Ranked Value}) = \frac{2N}{4} = \frac{2(50)}{4} = 25$$

Class Limits	f	cf
18 – 26	3	3
27 – 35	5	8
36 – 44	9	17
45 – 53	14	31
54 – 62	11	42
63 – 71	6	48
72 – 80	2	50
Total	50	

LB = 45 – 0.5
= 44.5

cf

f

Q_2 Class

$$Q_2 = LB + \left(\frac{\frac{2N}{4} - cf}{f} \right)(i) = 44.5 + \left(\frac{\frac{2(50)}{4} - 17}{14} \right)(9) = 49.64$$

Solution for Q_3

$$Q_3(\text{Ranked Value}) = \frac{3N}{4} = \frac{3(50)}{4} = 37.5$$

Class Limits	f	cf
18 – 26	3	3
27 – 35	5	8
36 – 44	9	17
45 – 53	14	31
54 – 62	11	42
63 – 71	6	48
72 – 80	2	50
Total	50	

LB = 54 – 0.5 = 53.5

cf

f

Q_3 Class

$$Q_3 = LB + \left(\frac{\frac{3N}{4} - cf}{f} \right)(i) = 53.5 + \left(\frac{\frac{3(50)}{4} - 31}{11} \right)(9) = 58.82$$

Solution for D_7

$$D_7(\text{Ranked Value}) = \frac{7N}{10} = \frac{7(50)}{10} = 35$$

Class Limits	f	cf
18 – 26	3	3
27 – 35	5	8
36 – 44	9	17
45 – 53	14	31
54 – 62	11	42
63 – 71	6	48
72 – 80	2	50
Total	50	

LB = 54 – 0.5
= 53.5

cf

D_7 Class

f

$$D_7 = LB + \left(\frac{\frac{7N}{10} - cf}{f}\right)(i) = 53.5 + \left(\frac{\frac{7(50)}{10} - 31}{11}\right)(9) = 56.77$$

Solution for P_{22}

$$P_{22}(\text{Ranked Value}) = \frac{22N}{100} = \frac{22(50)}{100} = 11$$

Class Limits	f	cf
18 – 26	3	3
27 – 35	5	8
36 – 44	9	17
45 – 53	14	31
54 – 62	11	42
63 – 71	6	48
72 – 80	2	50
Total	50	

LB = 36 – 0.5
= 35.5

cf

f

P_{22} Class

$$P_{22} = LB + \left(\frac{\frac{22N}{100} - cf}{f}\right)(i) = 35.5 + \left(\frac{\frac{22(50)}{100} - 8}{9}\right)(9) = 38.5$$

Midhinge

It is the mean of the first (Q_1) and third (Q_3) quartiles in the data set. It is used to overcome potential problems introduced by extreme values (or outliers) in the data set.

$$\text{Midhinge} = \frac{Q_1 + Q_3}{2}$$

Example for Midhinge

Find the midhinge of the ages of 9 middle-management employees of a certain company. The ages are 53, 45, 59, 48, 54, 46, 51, 58, and 55.

Solution:

Recall that $Q_1 = 47$ and $Q_3 = 56.5$.

$$\text{Midhinge} = \frac{Q_1 + Q_3}{2} = \frac{47 + 56.5}{2} = \frac{103.5}{2} = 51.75$$

The midhinge age is 51.75.

Interquartile Range (IQR)

A is a measure of statistical dispersion, being equal to the difference between the third and first quartiles.

It is also called the midspread or middle fifty.

$$\text{Interquartile Range (IQR)} = Q_3 - Q_1$$

Example for Interquartile Range

Find the interquartile range of the ages of 9 middle-management employees of a certain company. The ages are 53, 45, 59, 48, 54, 46, 51, 58, and 55.

Solution:

Recall that $Q_1 = 47$ and $Q_3 = 56.5$.

Interquartile Range (IQR) $= Q_3 - Q_1 = 56.5 - 47 = 9.5$

The IQR is 9.5.

Quartile Deviation (QD)

It is slightly better measure of absolute dispersion than the range.

It ignores the observation on the tails.

The difference samples from a population & calculate their quartile deviations, their values are quite likely to be sufficiently different (called sampling fluctuation).

It is calculated from the sample data does not help us to draw any conclusion (inference) about the quartile deviation in the population.

$$\text{Quartile Deviation (QD)} = \frac{Q_3 - Q_1}{2}$$

Example for Quartile Deviation

Find the quartile deviation of the ages of 9 middle-management employees of a certain company. The ages are 53, 45, 59, 48, 54, 46, 51, 58, and 55.

Solution:

Recall that $Q_1 = 47$ and $Q_3 = 56.5$.

$$\text{Quartile Deviation (QD)} = \frac{Q_3 - Q_1}{2}$$

$$= \frac{56.5 - 47}{2} = \frac{9.5}{2} = 4.75$$

The QD is 4.75.

Coefficient of Variation (CV)

It is used when one is interested to compare standard deviations of two different units, coefficient of variations can be applied.

Sample Mean:

$$CV = \frac{s}{\overline{X}}(100\%)$$

Population Mean:

$$CV = \frac{\sigma}{\mu}(100\%)$$

Example 1 for Coefficient of Variations

The average age of the engineers at VSAS Pipeline Corporation is 33 years, with a standard deviation of 3; the average monthly salary of the engineers is 45,000, with standard deviation of 3,150. Determine the coefficient of variations of age and salary.

Solution:

$$CV = \frac{s}{\bar{X}}(100\%) = \frac{3}{33}(100\%) = 10\%$$ Age

$$CV = \frac{s}{\bar{X}}(100\%) = \frac{3{,}150}{45{,}000}(100\%) = 7\%$$ Salary

Since the coefficient of variation is larger for age, the ages are more variable than the salary.

Example 2 for Coefficient of Variations

The mean of commissions of Educational Insurance over 1-year period is 12,500, and the standard deviation is 1,350. The mean of the number of sales is 60, and the variance is 56.25. Compare the variations of the two.

Solution:

$$CV = \frac{s}{\overline{X}}(100\%) = \frac{1,350}{12,500}(100\%) = 10.8\%$$ **Commission**

$$CV = \frac{s}{\overline{X}}(100\%) = \frac{\sqrt{56.25}}{60}(100\%) = 12.5\%$$ **Sales**

Since the coefficient of variation is larger for sales, the sales are more variable than the commissions.

Chebychev's Theorem

For ay set of observations, the proportion of the values that lie within k standard deviations of the mean is at least $1-\frac{1}{k^2}$, where k is any constant greater than 1.

Example for Chebychev's Theorem

The mean price of laptop computer is 25,500 and the standard deviation is 2,500. Find the price range for which at least 88.89% of the laptop will sell.

Example 1 for Chebychev's Theorem

Chebychev's theorem states that 88.89%, of the data values will fall within 3 standards of the mean.

25,500 + 3(2,500) = 25,500 + 7,500 = 33,000

25,500 – 3(2,500) = 25,500 – 7,500 = 18,000

Therefore, at least 88.89% of all laptop sold will have a price range from 18,000 and 33,000

Example for Chebychev's Theorem

A survey conducted by Commission on Higher Education (CHED) found that the mean amount of training allowance for department heads of colleges and universities was 25,000. The standard deviation was 1,500. Using Chebychev's theorem, find the minimum percentage of the data values that will fall between 22,000 and 28,000.

Example 2 for Chebychev's Theorem

Subtract the mean from the larger value.

28,000 – 25,000 = 3,000

Divide the difference by the sd to obtain k

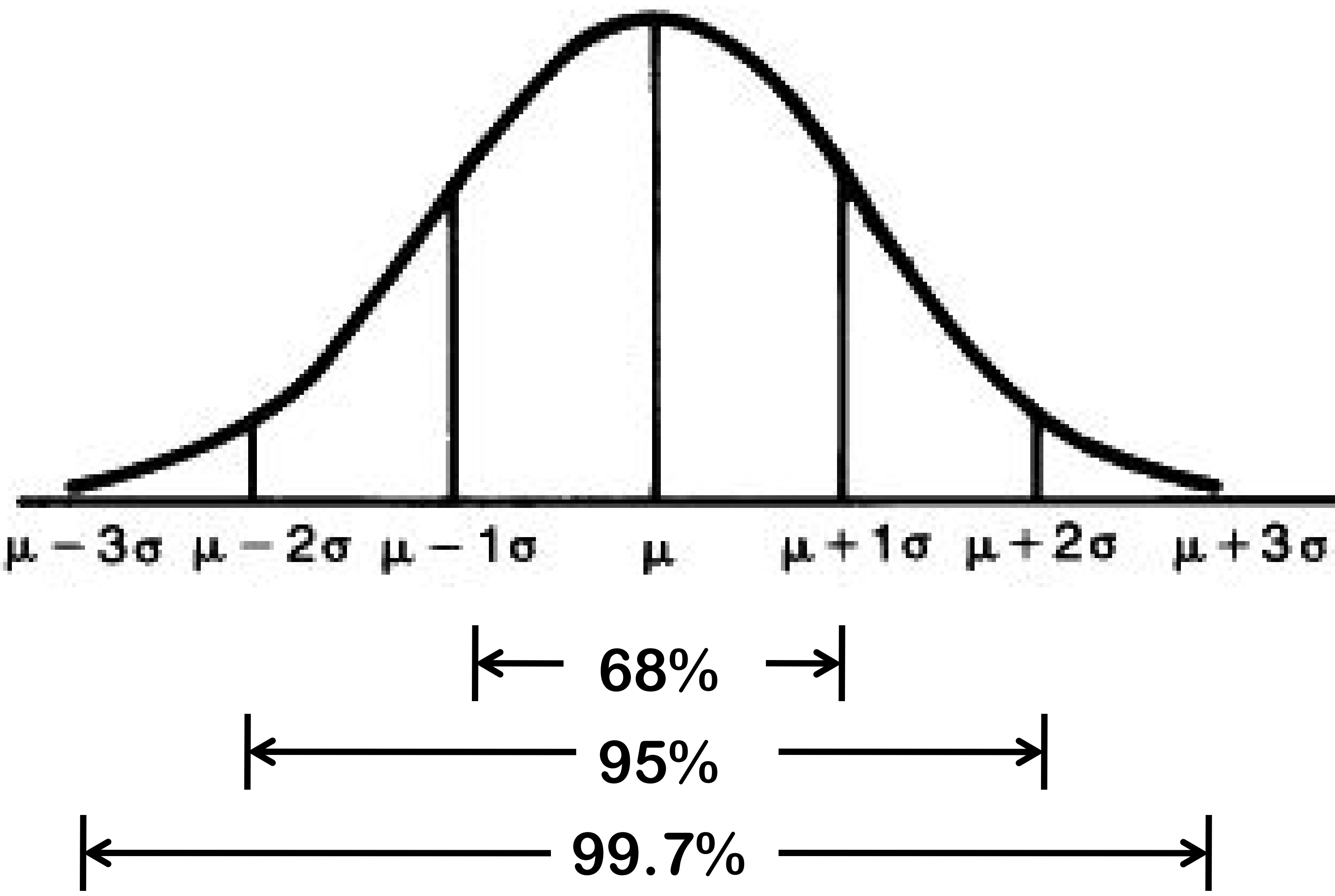
$$k = \frac{3{,}000}{1{,}500} = 2.0$$

Use Chebychev's theorem.

$$1 - \frac{1}{k^2} = 1 - \frac{1}{2.0^2} = 1 - \frac{1}{4} = 1 - 0.25 = 0.75$$

Therefore, at least 75% of the data value will fall between 22,000 and 28,000.

Empirical Rule



Kurtosis

Statistical measure used to describe the distribution of observed data around the mean. It measures the relative peakedness or flatness of a distribution (as compared to the normal distribution, which shows a kurtosis of zero)

$$\text{kurt} = \left\{ \left[\frac{n(n+1)}{(n-1)(n-2)(n-3)} \right] \left[\sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{s} \right)^4 \right] \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

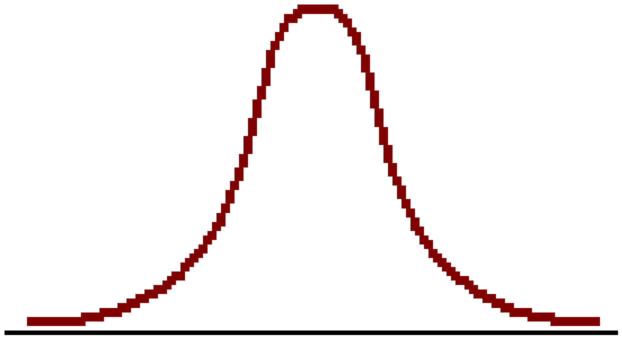
Three Types of Kurtosis

Leptokurtic are distributions where values clustered heavily or pile up in the center. ($k > 0$)

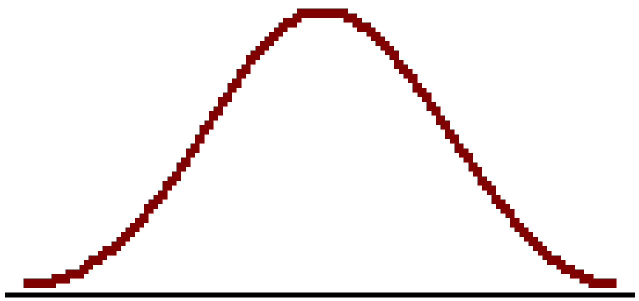
Mesokurtic are intermediate distribution w/c are neither too peaked nor too flat. ($k = 0$).

Platykurtic are flat distributions with values more evenly distributed about the center with broad humps and shot tails. ($k < 0$)

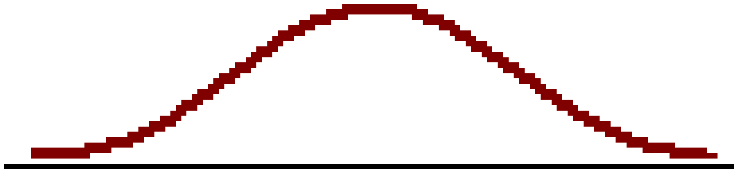
Three Types of Kurtosis



Leptokurtic



Mesokurtic



Platykurtic

Example for Kurtosis

Example: The daily rates of a sample of eight employees at GMS Inc. are 550, 420, 560, 500, ₱700, 670, 860, 480. Find the range.

Solution:

Compute for the mean

$$\overline{X} = \frac{\sum X}{n} = \frac{550 + 420 + 560 + 500 + 700 + 670 + 860 + 480}{8}$$

$$= \frac{4{,}740}{8} = \boxed{592.50}$$

Solution…

X	$X-\bar{X}$	$(X-\bar{X})^2$
550	–42.5	1,806.25
420	–172.5	29,756.25
560	–32.5	1,056.25
500	–92.5	8,556.25
700	107.5	11,556.25
670	77.5	6,006.25
860	267.5	71,556.25
480	–112.5	12,656.25
$\sum X = 4{,}740$	$\sum (X-\bar{X}) = 0$	$\sum / X-\bar{X} / = 142{,}950$

$$s = \sqrt{\frac{\sum (X-\bar{X})^2}{n-1}} = \sqrt{\frac{142{,}950}{8-1}} = \sqrt{20{,}421.43} = \boxed{142.90}$$

Solution…

X	$X-\bar{X}$	$(X-\bar{X})^2$	$\left(\frac{X-\bar{X}}{s}\right)^4$
550	–42.5	1,806.25	0.0078
420	–172.5	29,756.25	2.1232
560	–32.5	1,056.25	0.0027
500	–92.5	8,556.25	0.1755
700	107.5	11,556.25	0.3202
670	77.5	6,006.25	0.0865
860	267.5	71,556.25	12.2779
480	–112.5	12,656.25	0.3841
4,740	0	142,950	15.3779

Solution…

$$\text{kurt} = \left\{\left[\frac{n(n+1)}{(n-1)(n-2)(n-3)}\right]\left[\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{s}\right)^4\right]\right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

$$= \left\{\left[\frac{8(8+1)}{(8-1)(8-2)(8-3)}\right][15.3779]\right\} - \frac{3(8-1)^2}{(8-2)(8-3)} = 0.3724$$

The distribution on the daily rates of employee is somewhat leptokurtic.

Coefficient of Skewness

Measures the general shape of the distribution or the lack of symmetry of a distribution.

Ranges from –3 to +3.

It relates the difference between the mean and the median to the standard deviation.

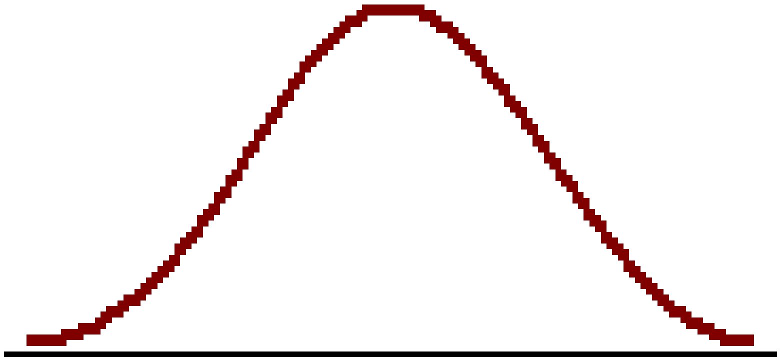
The direction of the long tail of the distribution points the direction of the skewness.

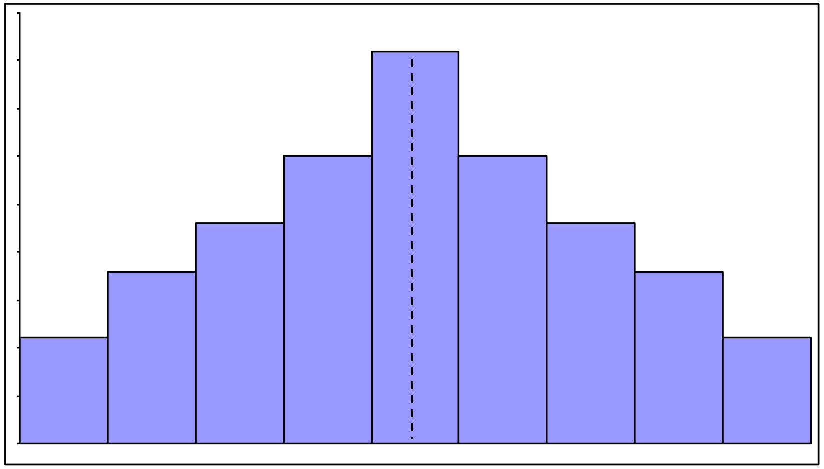
Types of Distribution

Symmetrical. Data values are evenly distributed.

The distribution is unimodal.

The mean, median, and mode are similar & are at the center of the distribution.





Mean = Median = Mode

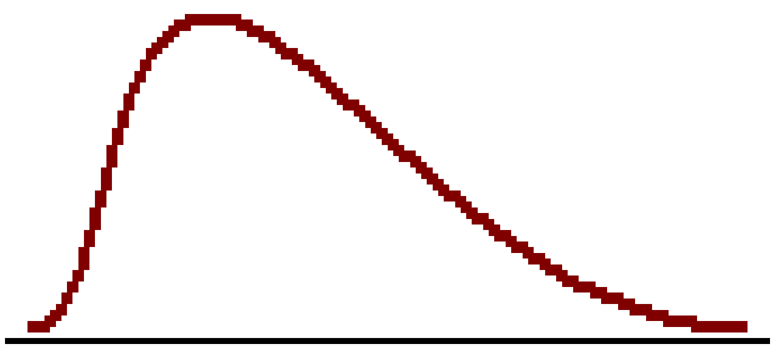
Types of Distribution

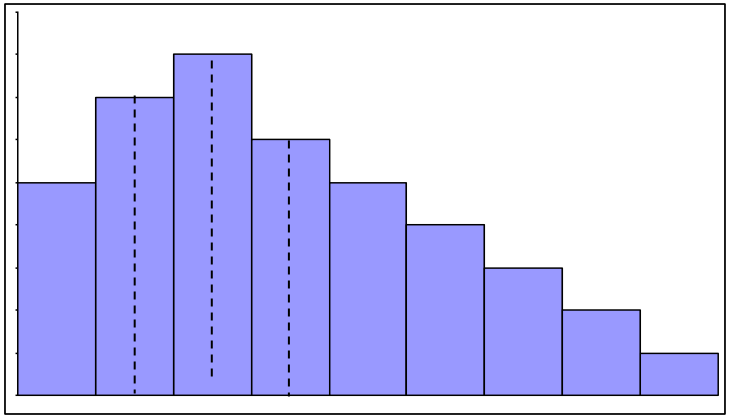
Positively Skewed (or Right-Skewed).

Most of the values in the data fall to the left of the mean and group at the lower end of the distribution.

The tail is to the right.

The mean is to the right of the median, and the mode is to the left of the median





Mode Median Mean

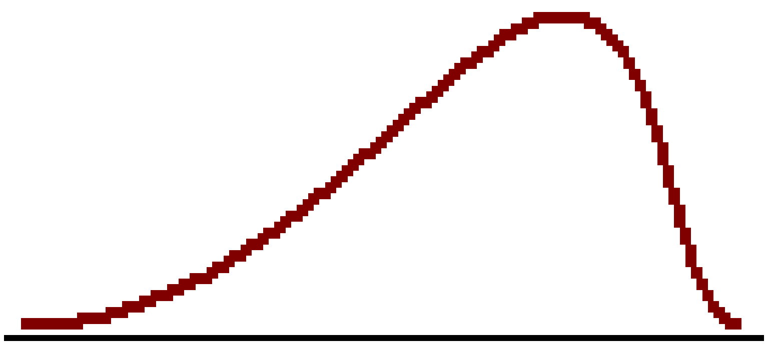
Types of Distribution

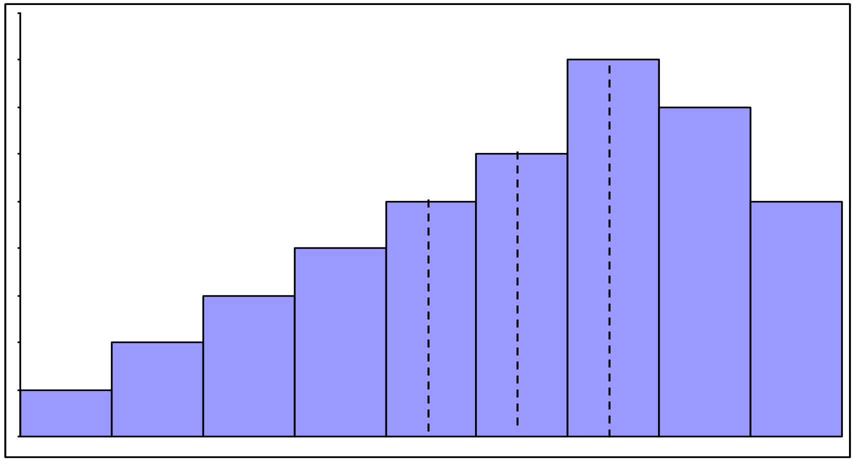
Negatively Skewed (or Left-Skewed).

The mass of the data values fall to the right of the mean and group at the upper end of the distribution.

The tail to the left.

The mean is to the left of the median, and the mode is to the right of the median.





Computation for Skewness

Pearson's Coefficient of Skewness

$$sk = \frac{3(\bar{X} - \text{median})}{s}$$

Software Coefficient of Skewness

$$sk = \frac{n}{(n-1)(n-2)}\left[\sum\left(\frac{X - \bar{X}}{s}\right)^3\right]$$

Example for Pearson's Coefficient of Skewness

Example: A motorcycle dealership pays its salesperson a salary plus a commission on sales. The mean monthly commission is 8,800, the median 9,000, and the standard deviation 1,200. Determine the coefficient of skewness. Comment on the shape of distribution.

Solution:

$$sk = \frac{3(\bar{X} - \text{median})}{s} = \frac{3(8{,}800 - 9{,}000)}{1{,}200} = \frac{3(-200)}{1{,}200} = -0.50$$

There is a slight negative skewness in the distribution of commission on sales.

Example for Software Coefficient of Skewness

Example: The daily rates of a sample of eight employees at GMS Inc. are 550, 420, 560, 500, ₱700, 670, 860, 480. Find the range.

Solution:

Compute for the mean

$$\overline{X} = \frac{\sum X}{n} = \frac{550 + 420 + 560 + 500 + 700 + 670 + 860 + 480}{8}$$

$$= \frac{4,740}{8} = \boxed{592.50}$$

Solution…

X	$X-\bar{X}$	$(X-\bar{X})^2$
550	–42.5	1,806.25
420	–172.5	29,756.25
560	–32.5	1,056.25
500	–92.5	8,556.25
700	107.5	11,556.25
670	77.5	6,006.25
860	267.5	71,556.25
480	–112.5	12,656.25
$\sum X = 4{,}740$	$\sum (X-\bar{X}) = 0$	$\sum / X-\bar{X} / = 142{,}950$

$$s = \sqrt{\frac{\sum (X-\bar{X})^2}{n-1}} = \sqrt{\frac{142{,}950}{8-1}} = \sqrt{20{,}421.43} = \boxed{142.90}$$

Solution....

X	$X-\bar{X}$	$(X-\bar{X})^2$	$\left(\frac{X-\bar{X}}{s}\right)^3$
550	–42.5	1,806.25	–0.03
420	–172.5	29,756.25	–1.76
560	–32.5	1,056.25	–0.01
500	–92.5	8,556.25	–0.27
700	107.5	11,556.25	0.43
670	77.5	6,006.25	0.16
860	267.5	71,556.25	6.56
480	–112.5	12,656.25	–0.49
4,740	0	142,950	4.59

Solution....

$$sk = \frac{n}{(n-1)(n-2)}\left[\sum\left(\frac{X-\overline{X}}{s}\right)^3\right]$$

$$= \frac{8}{(8-1)(8-2)}(4.59) = 0.19047619(4.59) = 0.87$$

The daily rates of employees are somewhat positively skewed.

Outliers

A data set should be checked for extremely high or extremely low values. These values are called outliers. Outliers can strongly affect the mean and standard deviation of a variable. One method in determining the outliers is when a data value in a data set is less Q_1 – 1.5(IQR) or greater than Q_3 + 1.5(IQR)

Outliers can strongly affect the mean and standard deviation of a variable.

One method in determining the outliers is when a data value in a data set is less Q_1 – 1.5(IQR) or greater than Q_3 + 1.5(IQR).

Example for Outliers

Check the data set of the ages of 9 middle-management employees of a certain company for outliers. The ages are 53, 45, 59, 48, 54, 46, 51, 58, and 55.

Solution:

Remember that $Q_1 = 47$, $Q_3 = 56.5$, and $IQR = 9.5$

$$Q_1 - 1.5(IQR) = 47 - 1.5(9.5) = 47 - 14.25 = 32.75$$

$$Q_3 + 1.5(IQR) = 56.5 + 1.5(9.5) = 56.5 + 14.25 = 70.75$$

Since, there are no values outside this interval; hence, there are no outliers in the data set.

Boxplot (Box-and-Whisker plot)

Introduced by John Tukey in 1970's..

It gives the following information:

- If the median is near the center of the box, the distribution is approximately symmetric.

- If the median falls to the right of the center of the box, the distribution is negatively skewed.

- If the median falls to the left of the center of the box, the distribution is positively skewed.

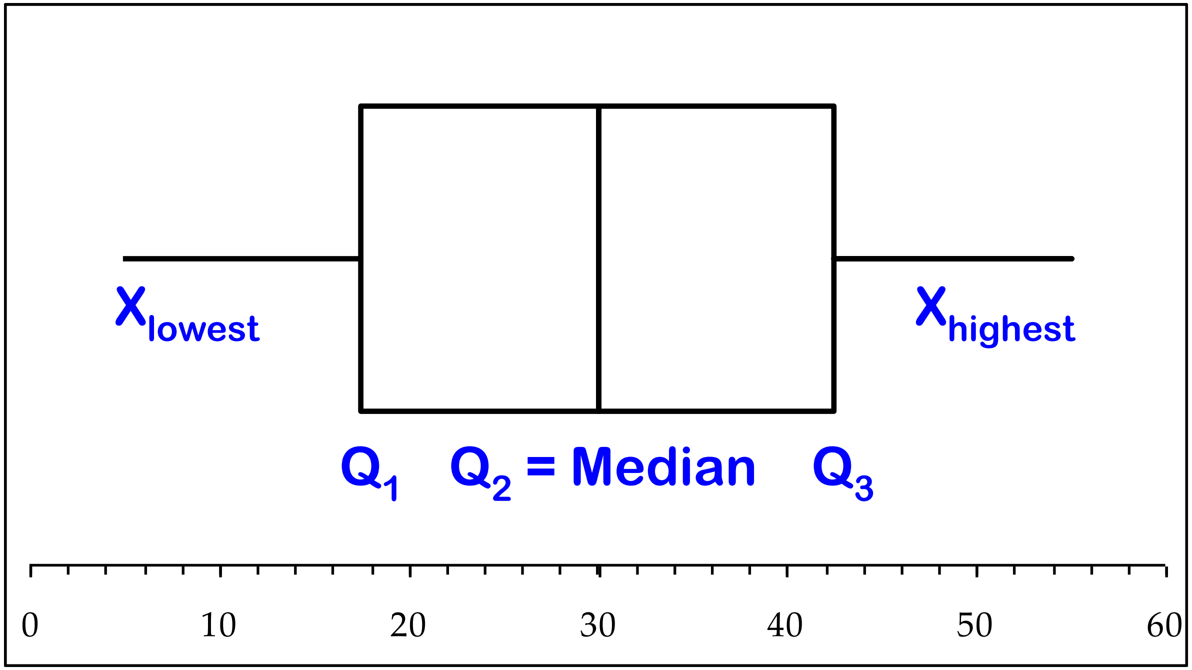
- If the lines are about the same length, the distribution is approximately symmetric.

Boxplot (Box-and-Whisker plot)

- **If the left line is larger than the right line, the distribution is negatively skewed**

- **If the right line is larger than the left line, the distribution is positively skewed.**

Figure: Boxplot



Example for Boxplot

Construct a boxplot for the data set of the ages of 9 middle-management employees of a certain company. The ages are 53, 45, 59, 48, 54, 46, 51, 58, and 55. What can we say about the distribution of the data set

Solution:

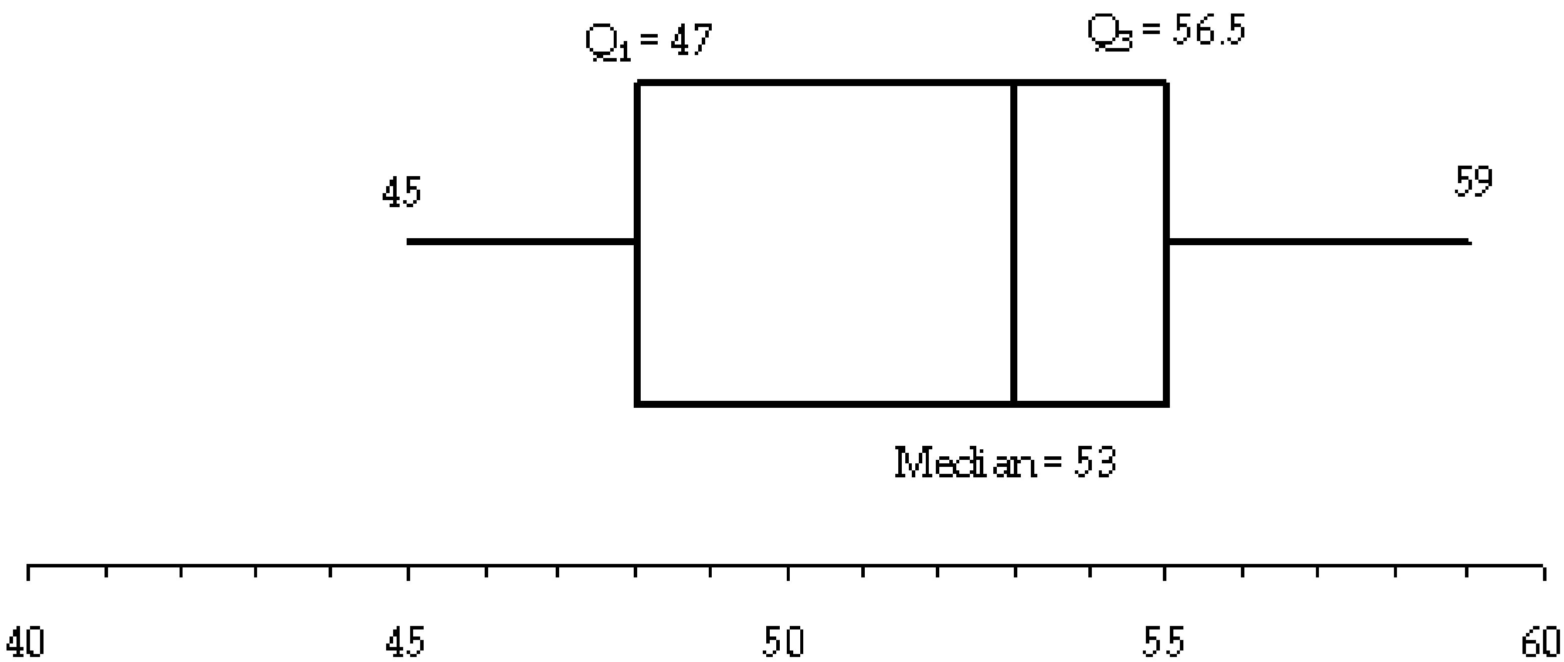
Recall that Q_1 = 47, Median = 53, and Q_3 = 56.5

Lowest Value = 45

Highest Value = 59

Solution for Boxplot

Figure: Employees Age



The data set of the distribution in negatively-skewed, since the median falls to the right of the center of the box

Effects of Changing Units

Let X_1, X_2, X_3, ... X_n be n observations. If we added a constant k to each observation or multiplied each observation by a number (h, h $\neq$ 0), then

Added Constant

Constant Multiplier

Summary Measure	$Y_i = X_i + k$	$Y_i = h X_i$
Range	Unaffected	Range (Y) = h Range (X)
Standard Deviation	Unaffected	s (Y) = /h/ s (X)
Quartiles	Q_i (Y) = Q_i (X) + k	Q_i (Y) = h Q_i (X)
IQR	Unaffected	IQR (Y) = h IQR (X)

Example 1

A business professor gave a test to a set of students. The test had 40 questions, each worth 2 points. The summary for the applicants' scores on the test are as follows:

Summary Statistics	
Range	22
Standard deviation	7.33
First quartile	64
Third quartile	72
Interquartile range	14

After grading the test, the professor realized that, because he had a typographical error in question no. 4, no students was able to answer the question. So he decided to adjust the students' scores by adding 2 points to each one. What will be the range, standard deviation, Q_1, Q_3 & IQR?

Solution

Summary Statistics	
Range	22 (unaffected)
Standard deviation	7.33 (unaffected)
First quartile	64 + 2 = 66
Third quartile	72 + 2 = 74
Interquartile range	14 (unaffected)

Example 2

The summary statistics for the monthly contribution of employees on cooperative are as follows:

Summary Statistics	
Range	700
Standard deviation	222.28
First quartile	1,100
Third quartile	1,400
Interquartile range	300

This month, employees voted to increase the contribution by 5 percent (5%), in order to support their new project. What will be the mean and median for the new increased of contributions.

Solution

Summary Statistics	
Range	1.05(700) = 7.35
Standard deviation	1.05(222.28) = 233.39
First quartile	1.05(1,100) = 1.155
Third quartile	1.05(1,400) = 1,470
Interquartile range	1.05(300) = 315

Statistics maybe defined as “a body of methods for making wise decisions in the face of uncertainty.”

– W. A. Wallis