The Normal Curve and its Applications

WHAT IS A NORMAL CURVE?

The literal meaning of the term normal is average. We make use of this term while computing the average in data related to education, psychology or sociology. In these areas, those who are able to reach a particular fixed level in qualification or characteristics are termed as normal, while there who are above or below this point are abnormal. Nature has been kind enough to distribute quite equally most of the things and attributes like wealth, beauty, intelligence, height, weight and the like. As a result, a majority among us possess average beauty, wealth, intelligence, height and weight. There are quite a few persons who deviate noticeably from average, either be above or below it. This is equally true for data in terms of achievement scores, intelligence scores, rating scores, etc. collected through tests, surveys and experiments performed in education, psychology and sociology on a randomly selected sample or population.

If we plot such a distribution of data on graph paper (Figure 8.1),

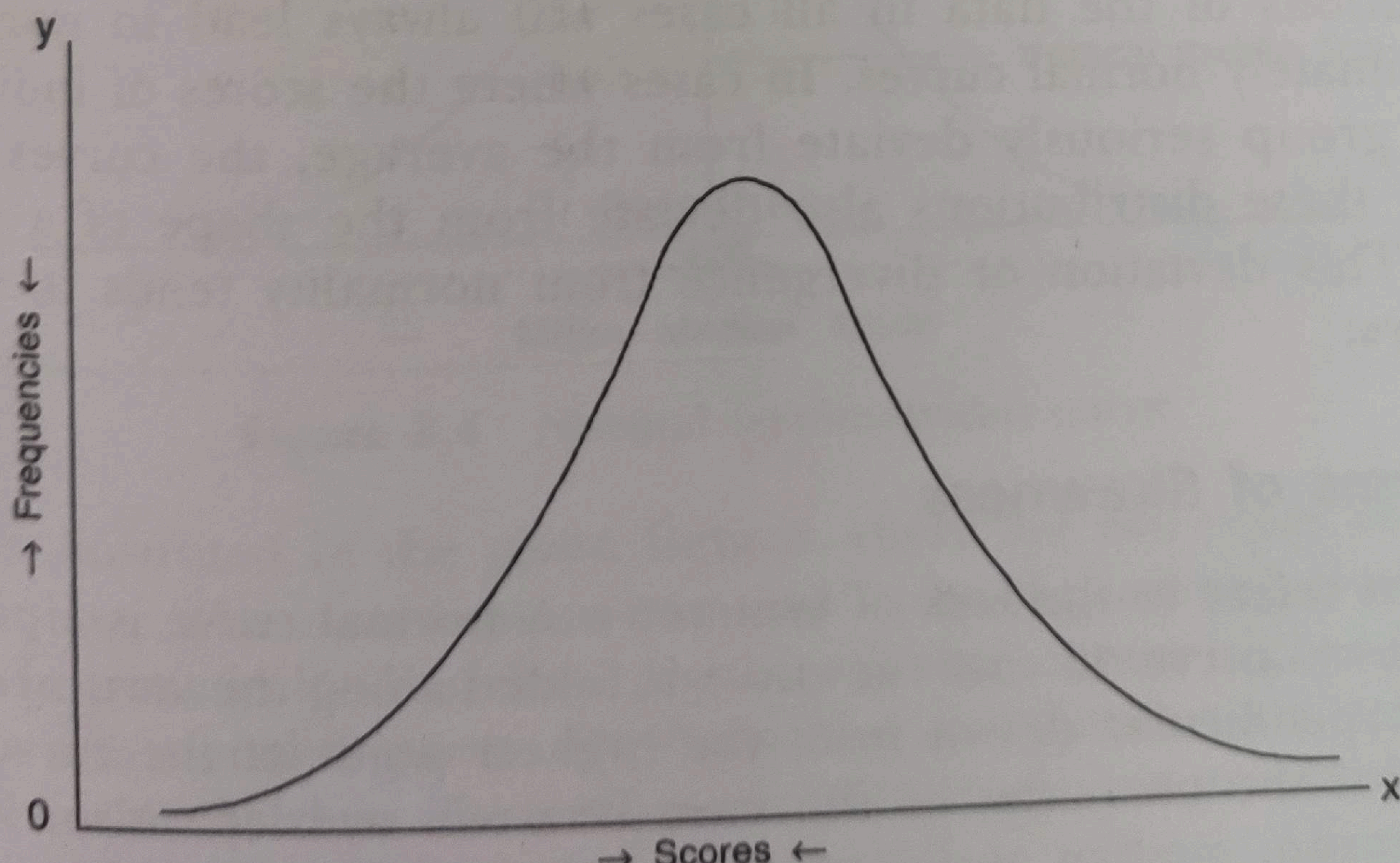

Figure 8.1 Normal curve.

we would get an interesting typical curve often resembling a vertical cross-section of a bell. This bell-shaped curve is called a *normal curve.*

The data from a certain coin or a dice-throwing experiment involving a chance success or probability, if plotted on a graph paper, give a frequency curve which closely resembles the normal curve. It is because of this reason and because of its origin from a game of chance that the normal curve is often called the normal probability curve.

Laplace and Gauss (1777–1855) derived this curve independently. They worked on experimental errors in Physics and Astronomy and found the errors to be distributed normally. This is the reason why the curve representing the normal characteristics is also named as normal curve of error or simply the curve of error where 'error' is used in the sense of a deviation from the true value. Because of its discovery by Laplace and Gauss, the curve is also named as Gaussian curve in the honour of Gauss.

The normal curve takes into account the law which states that the greater a deviation from the mean or average value in a series, the less frequently it occurs. This is satisfactorily used for describing many distributions which arise in the fields of education, psychology and sociology.

However, it is not at all essential for a normal distribution to be described by an exactly perfect bell shaped curve as shown in Figure 8.1. Such a perfect symmetrical curve rarely exists in our actual dealings as we usually cannot measure an entire population. Instead, we work on representative samples of the population. Therefore, in actual practice, the slightly deviated or distorted bell-shaped curve is also accepted as the normal curve on the assumption of normal distribution of the characteristics measured in the entire population.

From the above account, it should not be assumed that the distributions of the data in all cases will always lead to normal or approximately normal curves. In cases where the scores of individuals in the group seriously deviate from the average, the curves representing these distributions also deviate from the shape of a normal curve. This deviation or divergence from normality tends to vary in two ways:

In Terms of Skewness

Skewness refers to the lack of symmetry. A normal curve is a perfectly symmetrical curve. In case the curve is folded along the vertical middle line (perpendicular drawn from the highest point of the curve to its base line) the two sides of the base line will overlap. Also, for this curve, mean, median and mode are the same. In many distributions which deviate from the normal, the values of mean, median and mode are different and there is no symmetry between the right and the left halves of the curve. Such distributions are said to be skewed, being inclined more towards the left or the right to the centre of the curve as shown in Figures 8.2 and 8.3. For comparison, a normal (symmetrical) curve is also provided in Figure 8.4.

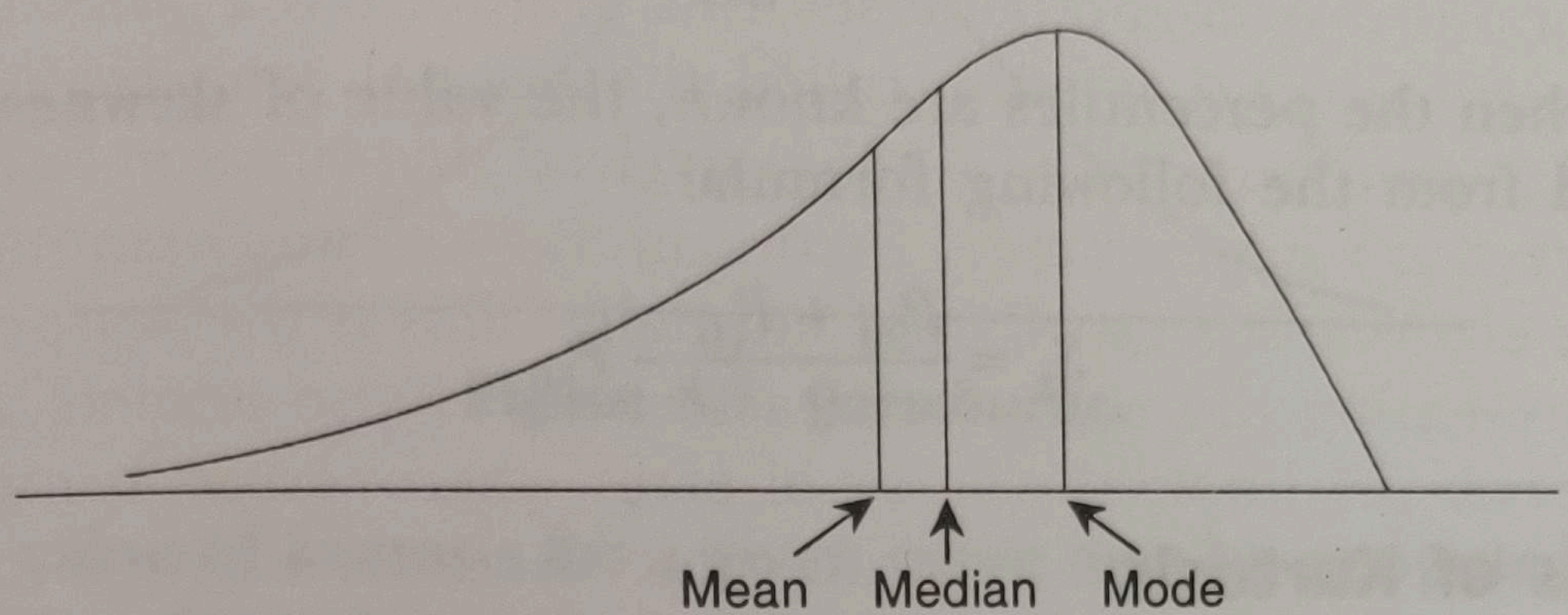

Figure 8.2 Negative skewness (the curve inclines more to the left).

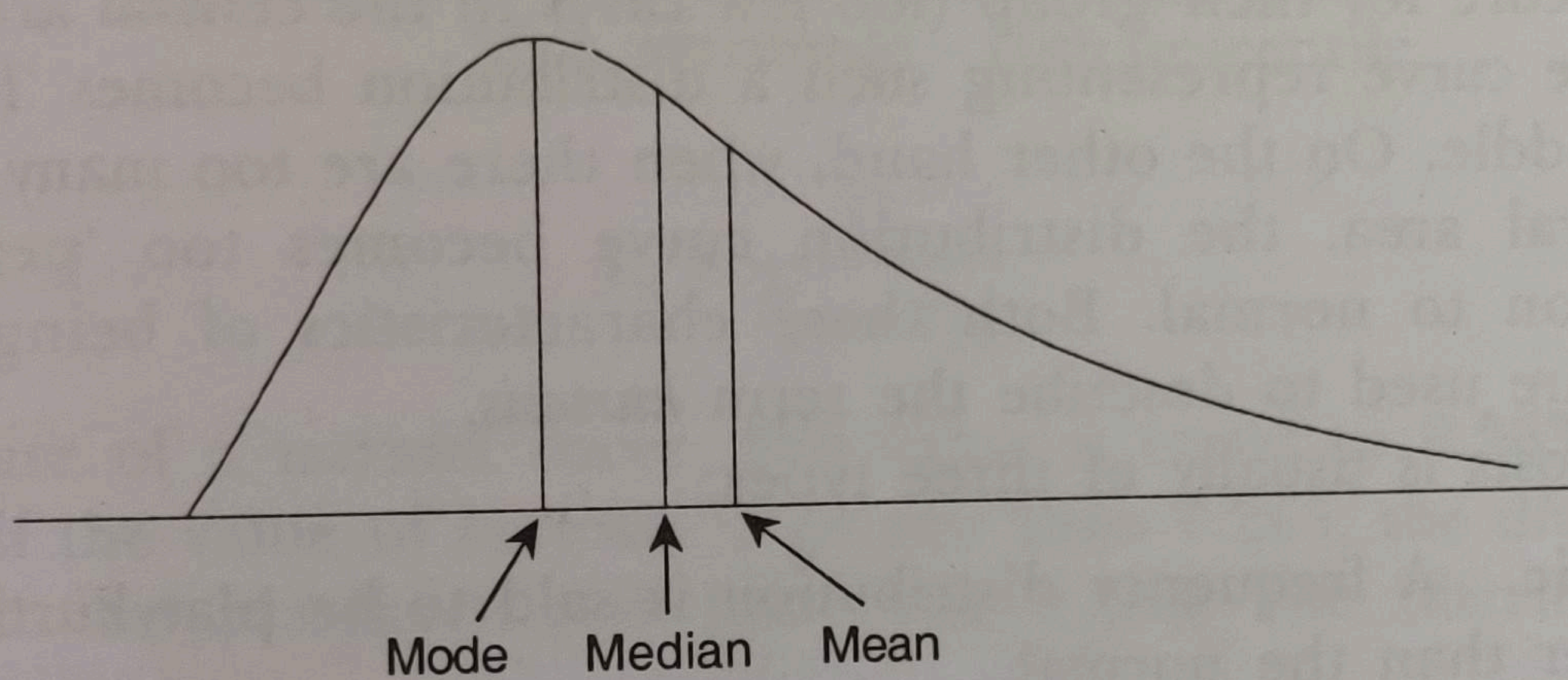

Figure 8.3 Positive skewness (the curve inclines more to the right).

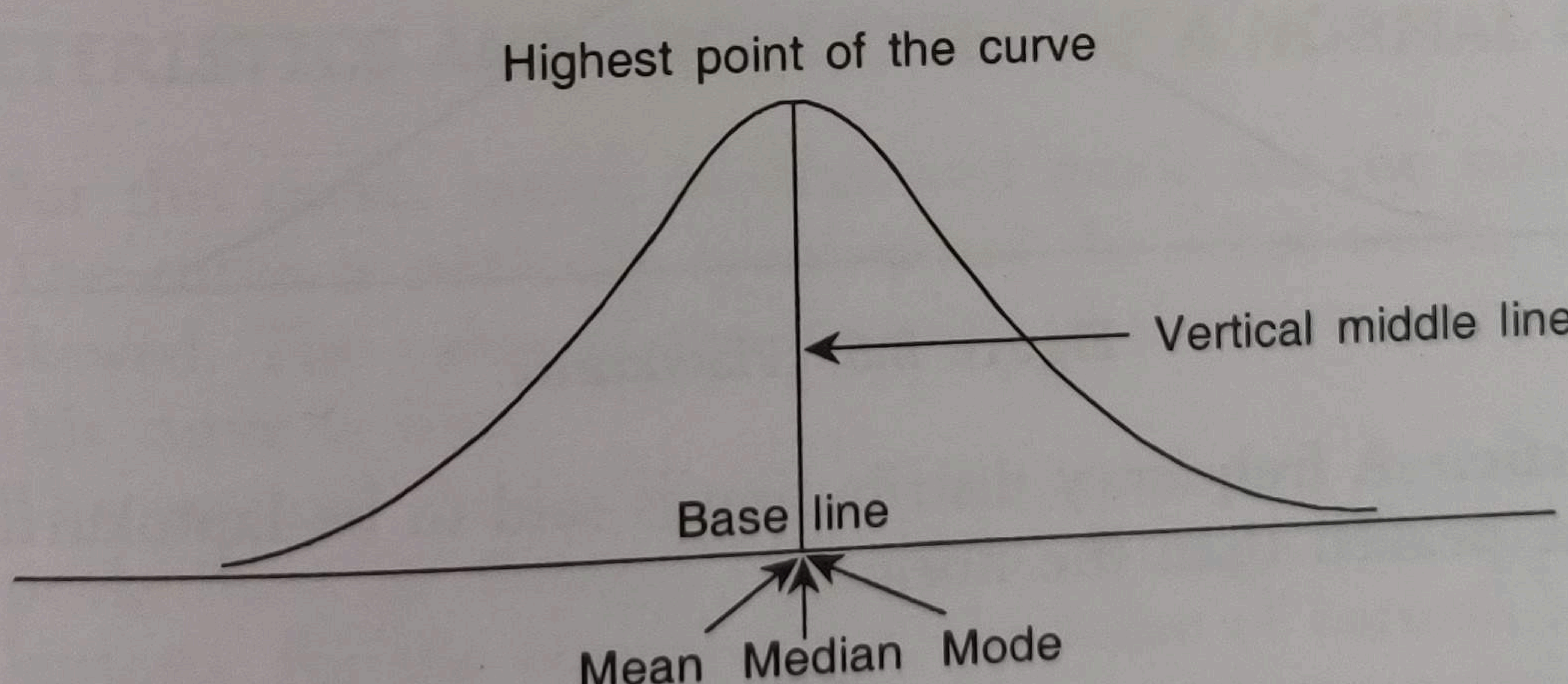

Figure 8.4 Normal (symmetrical) curve.

As illustrated in the given figures, there are two types of skewed distributions. The distributions are said to be skewed negatively when there are many individuals in a group with their scores higher than the average score of the group. Similarly, the distributions are said to be skewed positively when there are more individuals in a group who score less than the average score for their group.

Skewness in a given distribution may be computed by the following formula:

$$\text{Skewness} = \frac{3\,(\text{Mean} - \text{Median})}{\text{Standard deviation}}$$

or

$$S_k = \frac{3\,(M - M_d)}{\text{SD}}$$

In case when the percentiles are known, the value of skewness may be computed from the following formula:

$$S_k = \frac{P_{90} + P_{10}}{2} - P_{50}$$

In Terms of Kurtosis

When there are very few individuals whose scores are near to the average score for their group (too few cases in the central area of the curve) the curve representing such a distribution becomes 'flattened' in the middle. On the other hand, when there are too many cases in the central area, the distribution curve becomes too 'peaked' in comparison to normal. Both these characteristics of being flat or peaked, are used to describe the term *kurtosis*.

Kurtosis is usually of three types:

Platykurtic. A frequency distribution is said to be platykurtic, when it is flatter than the normal.

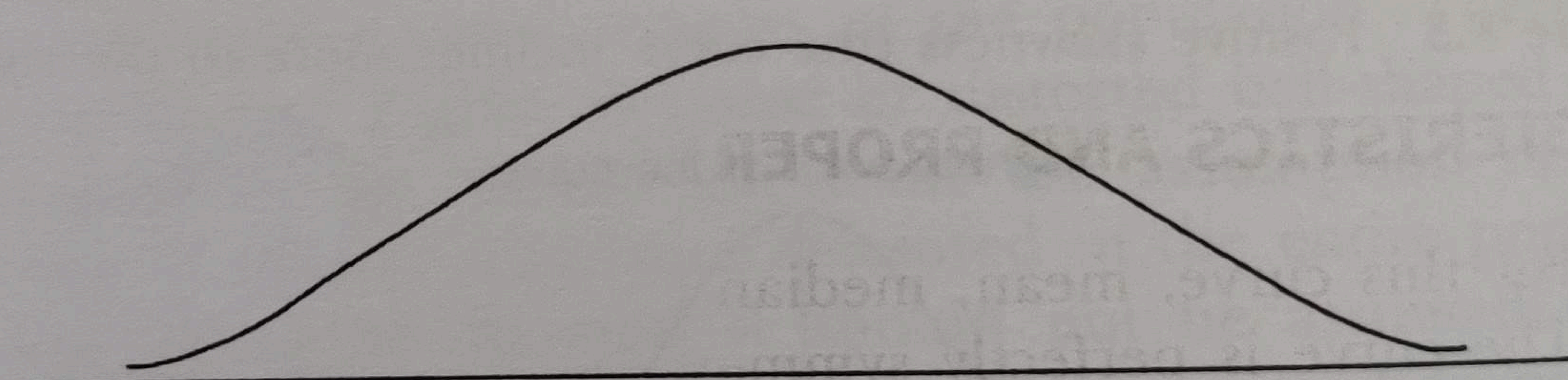

Figure 8.5 Platykurtic.

Leptokurtic. A frequency distribution is said to be laptokurtic, when it is more peaked than the normal.

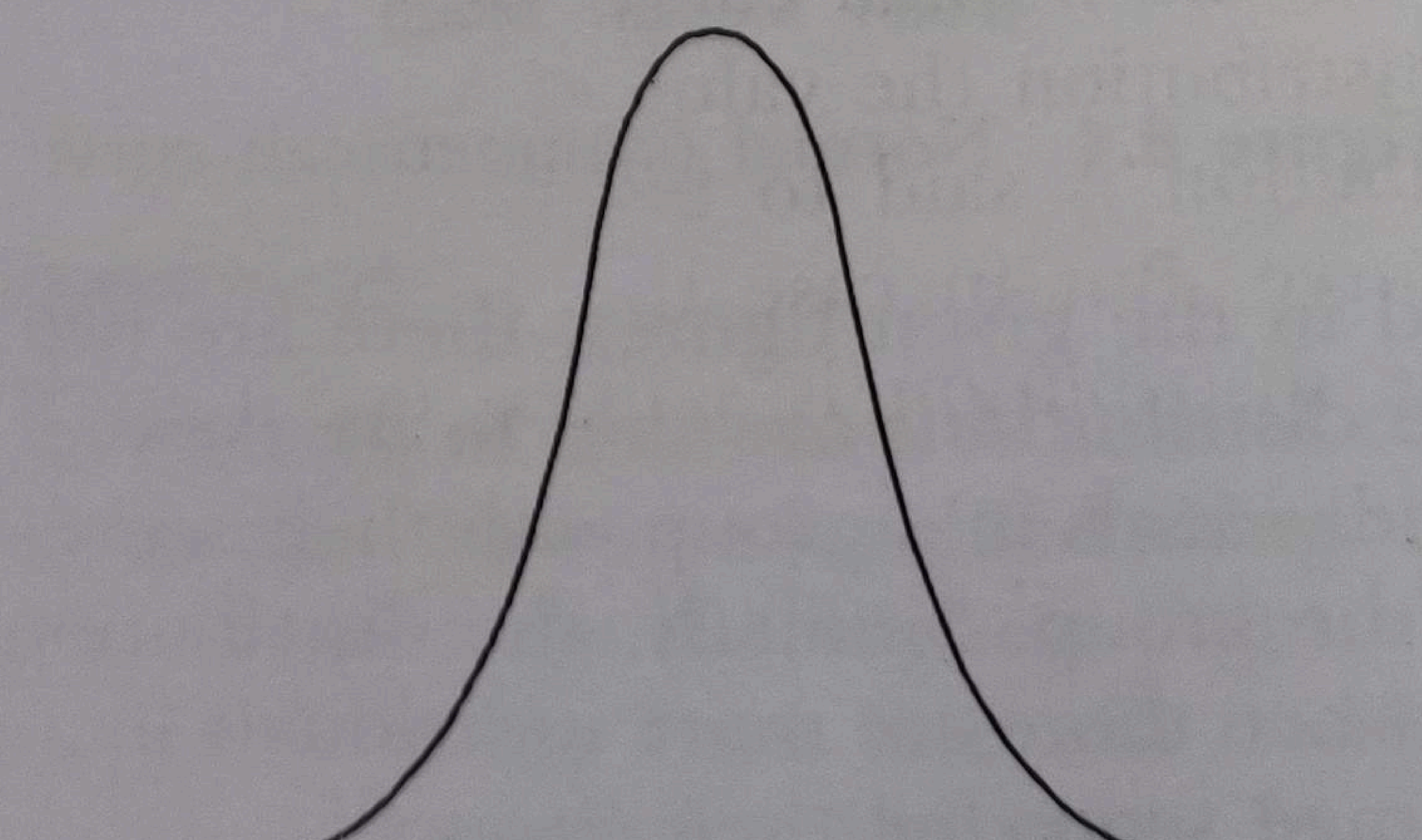

Figure 8.6 Leptokurtic.

Mesokurtic. A frequency distribution is said to be mesokurtic, when it almost resembles the normal curve (neither too flattened nor too peaked).

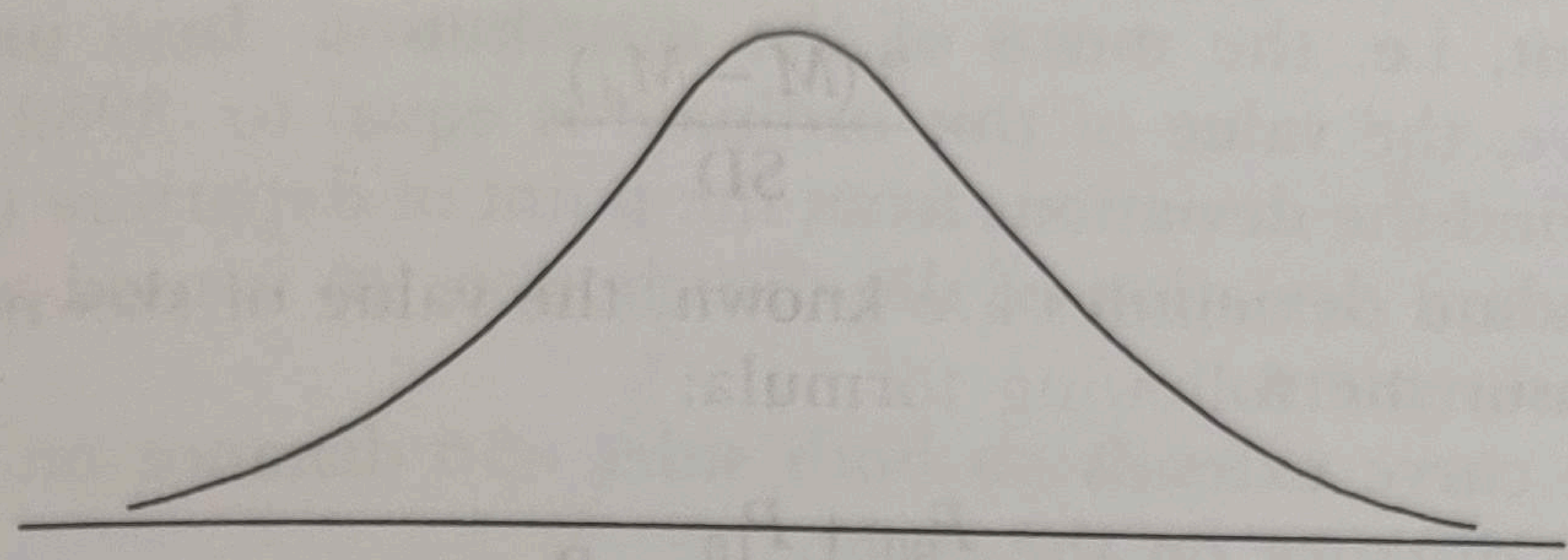

Figure 8.7 Mesokurtic.

The value of kurtosis for a given curve may be computed through the following formula:

$$\text{Kurtosis} = \frac{\text{Quartile deviation}}{\text{90th percentile} - \text{10th percentile}}$$

or

$$K_u = \frac{Q}{P_{90} - P_{10}}$$

In the case of a normal curve, this value is equal to 0.263. Consequently, if the value of kurtosis is greater than 0.263, the distribution is said to be platykurtic; if less than 0.263, the distribution is leptokurtic.

CHARACTERISTICS AND PROPERTIES OF A NORMAL CURVE

1. For this curve, mean, median and mode are the same.
2. The curve is perfectly symmetrical. In other words, it is not skewed. The value of the measure of skewness computed for this curve is zero.
3. The normal curve serves as a model for describing the peakedness or flatness of a curve through the measure of kurtosis. For the normal curve, the value of kurtosis is 0.263. If for a distribution the value of kurtosis is more than 0.263, the distribution is said to be more flat at the top than the normal curve. But in case the value of kurtosis is less than 0.263, the distribution is said to be more peaked than the normal.
4. The curve is asymptotic. It approaches but never touches the base line at the extremes because of the possibility of locating in the population, a case which scores still higher than our highest score or lower than our lowest score. Therefore, theoretically, it extends from minus infinity to plus infinity.

5. As the curve does not touch the base line, the mean is used as the starting point for working with the normal curve.
6. The curve has its maximum height or ordinate at the starting point, i.e. the mean of the distribution. In a unit normal curve, the value of this ordinate is equal to .3989.
7. To find the deviations from the point of departure (i.e. mean), standard deviation of the distribution (σ) is used as a unit of measurement.
8. The curve extends on both sides -3σ distance on the left to $+3\sigma$ distance on the right.
9. The points of inflection of the curve occur at ±1 standard deviation unit ($\pm 1\sigma$) above and below the mean. Thus the curve changes from convex to concave in relation to the horizontal axis at these points.
10. The total area under the curve extending from -3σ to $+3\sigma$ is taken arbitrarily to be 10,000 because of the greater ease in the computation of the fractional parts of the total area found for the mean and the ordinates erected at various distances from the mean. The computation of such fractional parts of the total area for travelling desired σ distances from the mean may be conveniently made with the help of Table B given in the appendix of this textbook.
11. We may find that 3413 cases out of 10,000 or 34.13% of the entire area of the curve lies between the mean and $+1\sigma$ on the base line of the normal curve. Similarly, another 34.13% cases lie between the mean and -1σ on the base line. Consequently, 68.26% of the area of the curve falls within the limits ±1 Standard deviation ($\pm 1\sigma$) unit from the mean. Going further it may be found out that 95.44% cases lie from -2σ to $+2\sigma$ and 99.74% cases lie from -3σ to $+3\sigma$. Consequently only 26 cases in 10,000 (10,000 – 9974) should be expected to lie beyond the range $\pm 3\sigma$ in a large sample as shown in Figure 8.8.

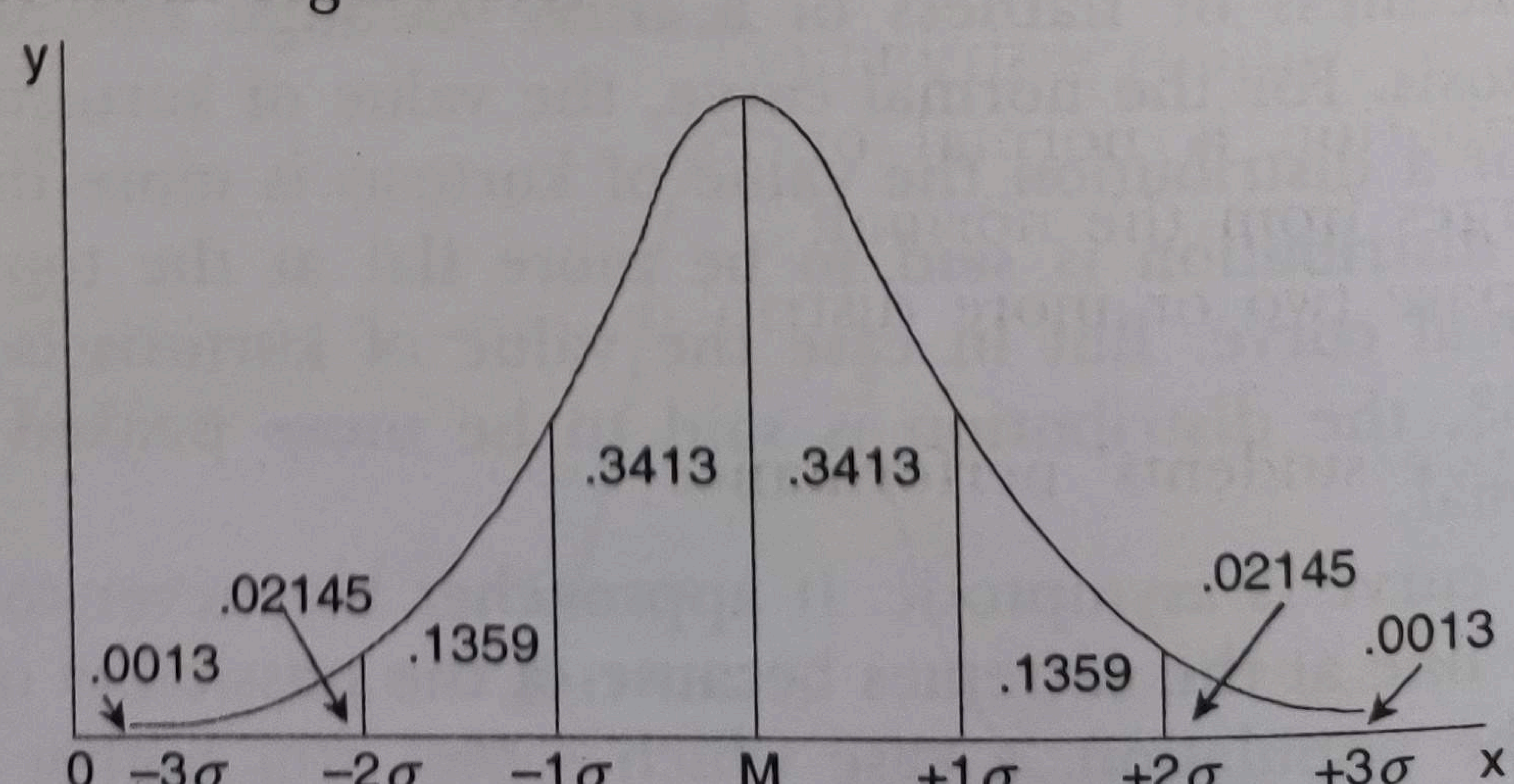

Figure 8.8 Normal curve showing areas at different distances from the mean.

12. In this curve, the limits of the distances $\pm 1.96\sigma$ include 95% and the limits $\pm 2.58\sigma$ include 99% of the total area of the curve, 5% and 1% of the area, respectively falling beyond these limits as shown in Figure 8.9.

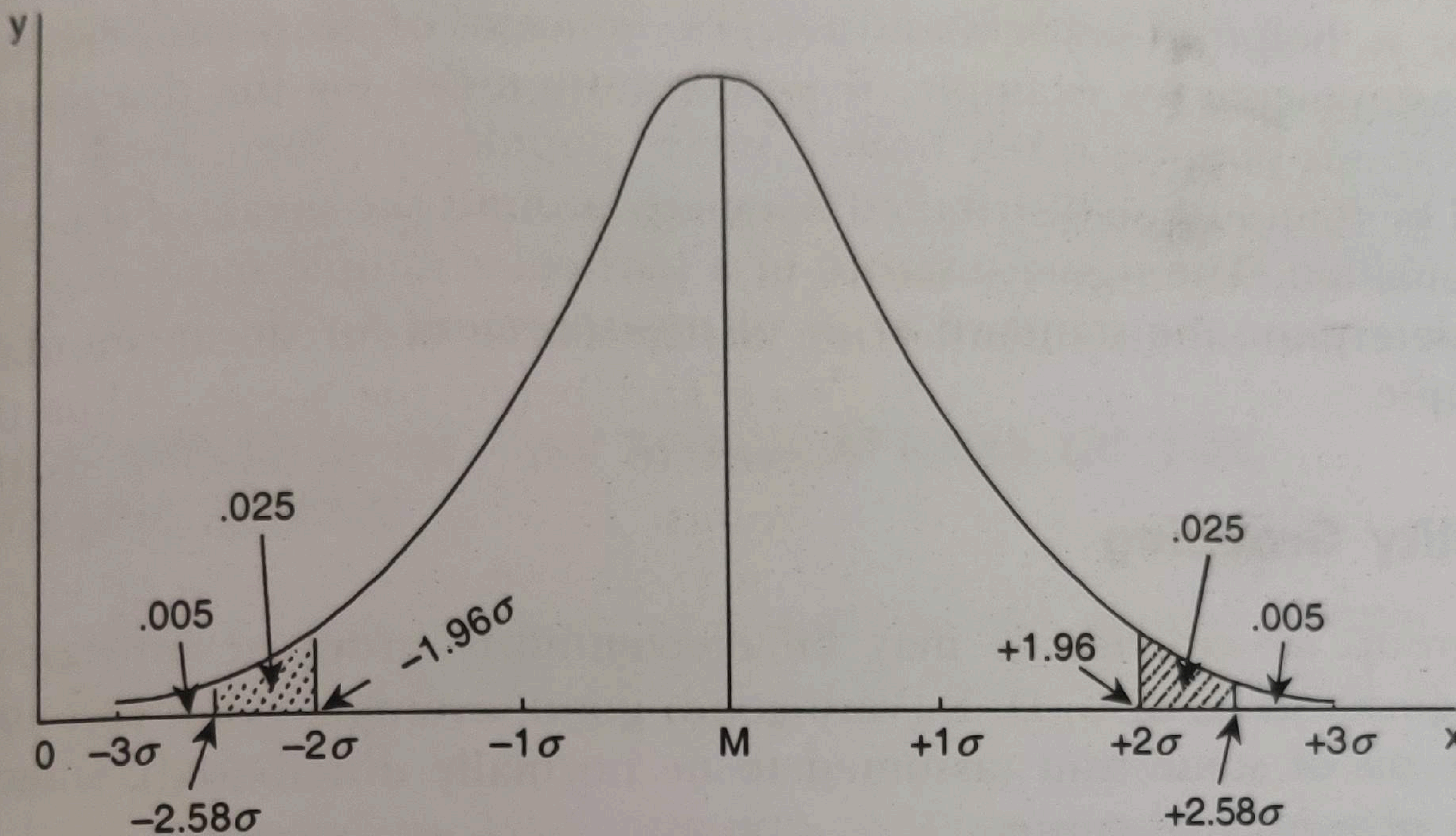

Figure 8.9 Normal curve showing 5% and 1% cases lying beyond the limits $\pm 1.96\sigma$ and $\pm 2.58\sigma$, respectively.