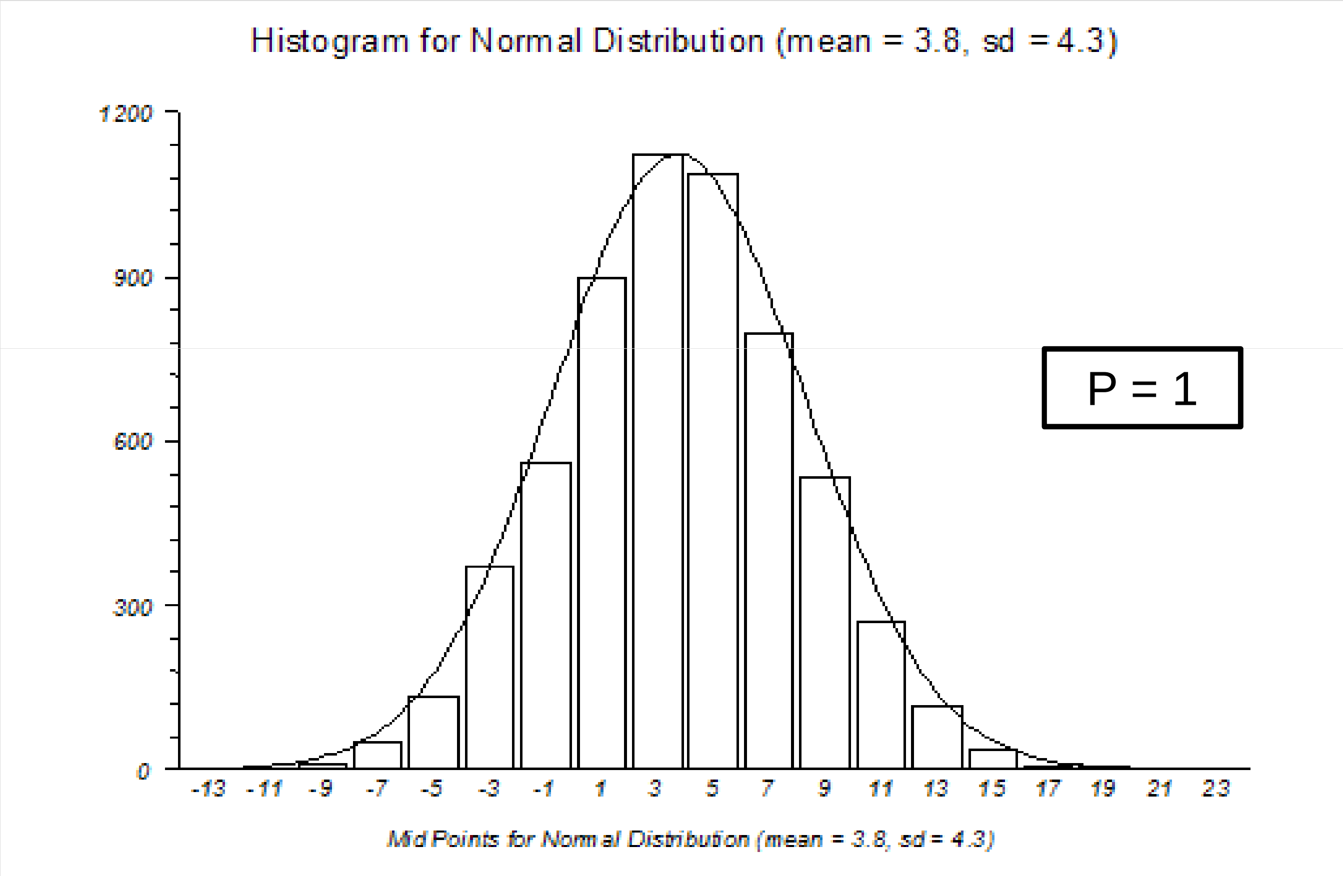
Analysis of interval data

Normal distribution



Normal distribution

- The center of the curve represents the mean (and median and mode)
- The curve is symmetrical around the mean
- The tails meet the x-axis in infinity
- The curve is bell-shaped
- The area under the curve is 1 (by definition)

Central limit theorem

The mean of several sample means is normally distributed (and approximates the sample mean), even if the distribution in the true population is not normally distributed.

Central limit theorem

	X_1	X_2	X_3	X_4	M
Sample 1	6	2	5	6	4.75

Central limit theorem

	X_1	X_2	X_3	X_4	M
Sample 1	6	2	5	6	4.75
Sample 2	2	3	1	6	3

Central limit theorem

	X_1	X_2	X_3	X_4	M
Sample 1	6	2	5	6	4.75
Sample 2	2	3	1	6	3
Sample 3	1	1	4	6	3

Central limit theorem

	X_1	X_2	X_3	X_4	M
Sample 1	6	2	5	6	4.75
Sample 2	2	3	1	6	3
Sample 3	1	1	4	6	3
Sample 4	6	2	2	1	2.75

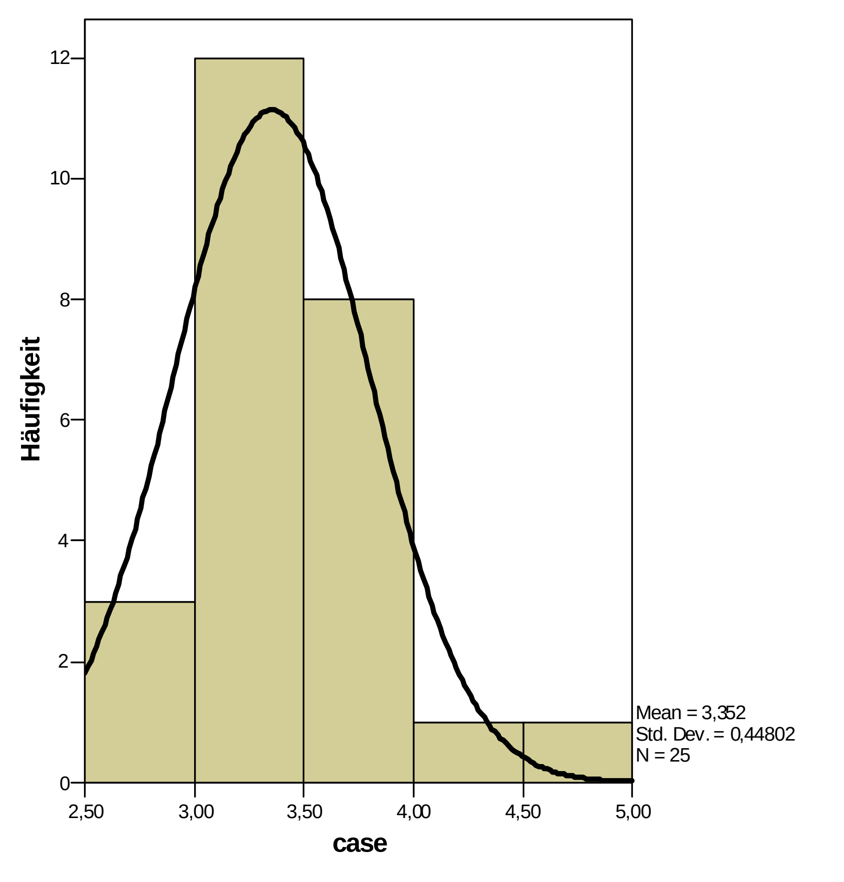
Central limit theorem

	X_1	X_2	X_3	X_4	M
Sample 1	6	2	5	6	4.75
Sample 2	2	3	1	6	3
Sample 3	1	1	4	6	3
Sample 4	6	2	2	1	2.75
Sample 5	1	5	1	3	2.5

Mean of the sample means

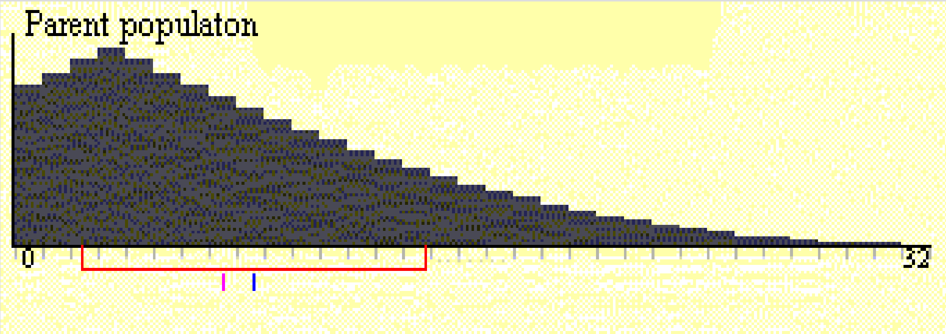
$$\frac{4.75 + 3.0 + 3.0 + 2.75 + 2.5}{5} = 3.2$$

The mean of the sample means approximates the population mean

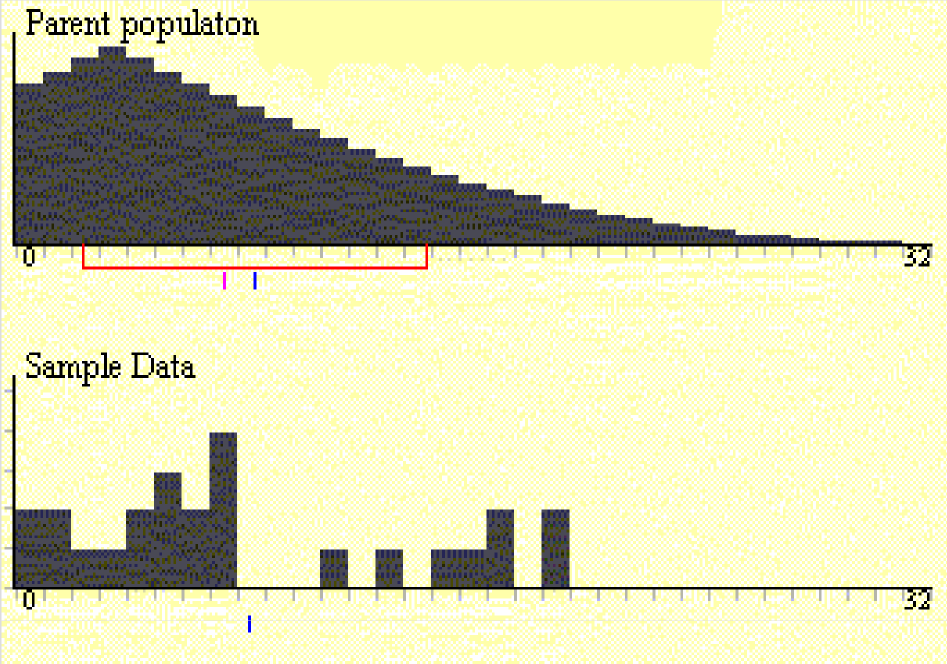


The sample means are normally distributed even if the phenomenon in the parent population is not normally distributed.

N = 25

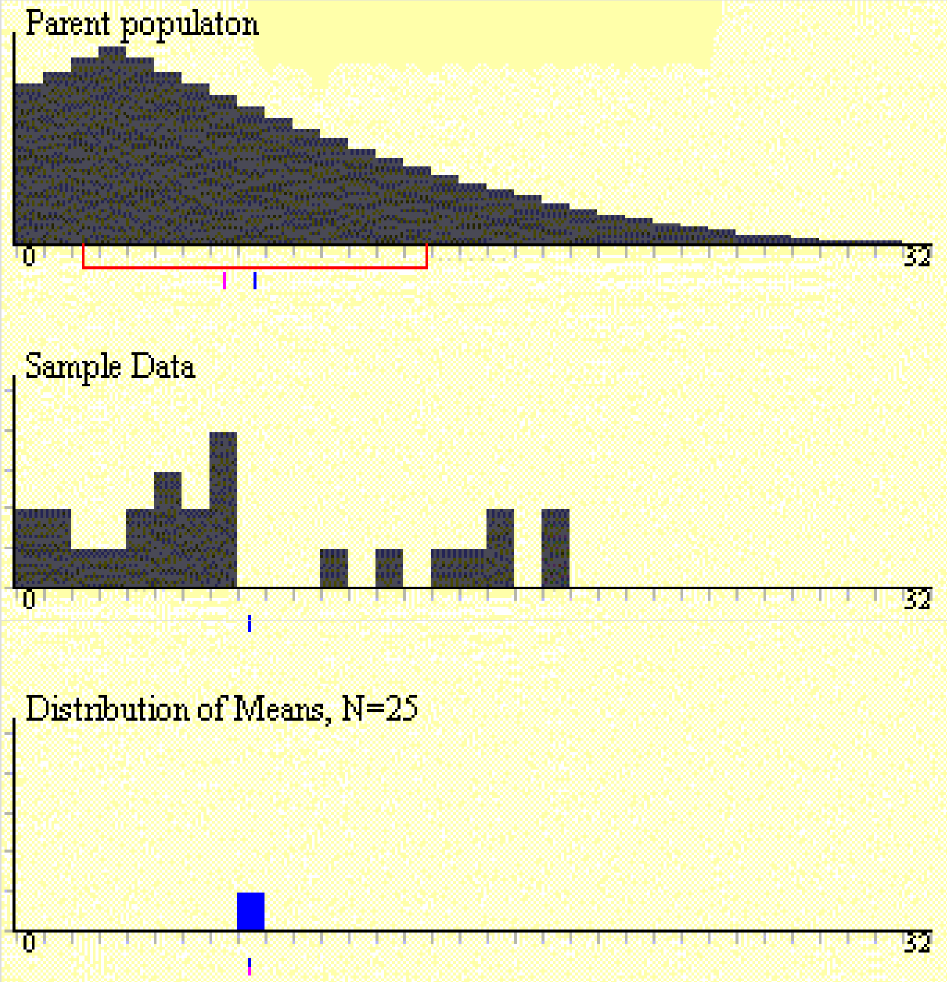


population



population

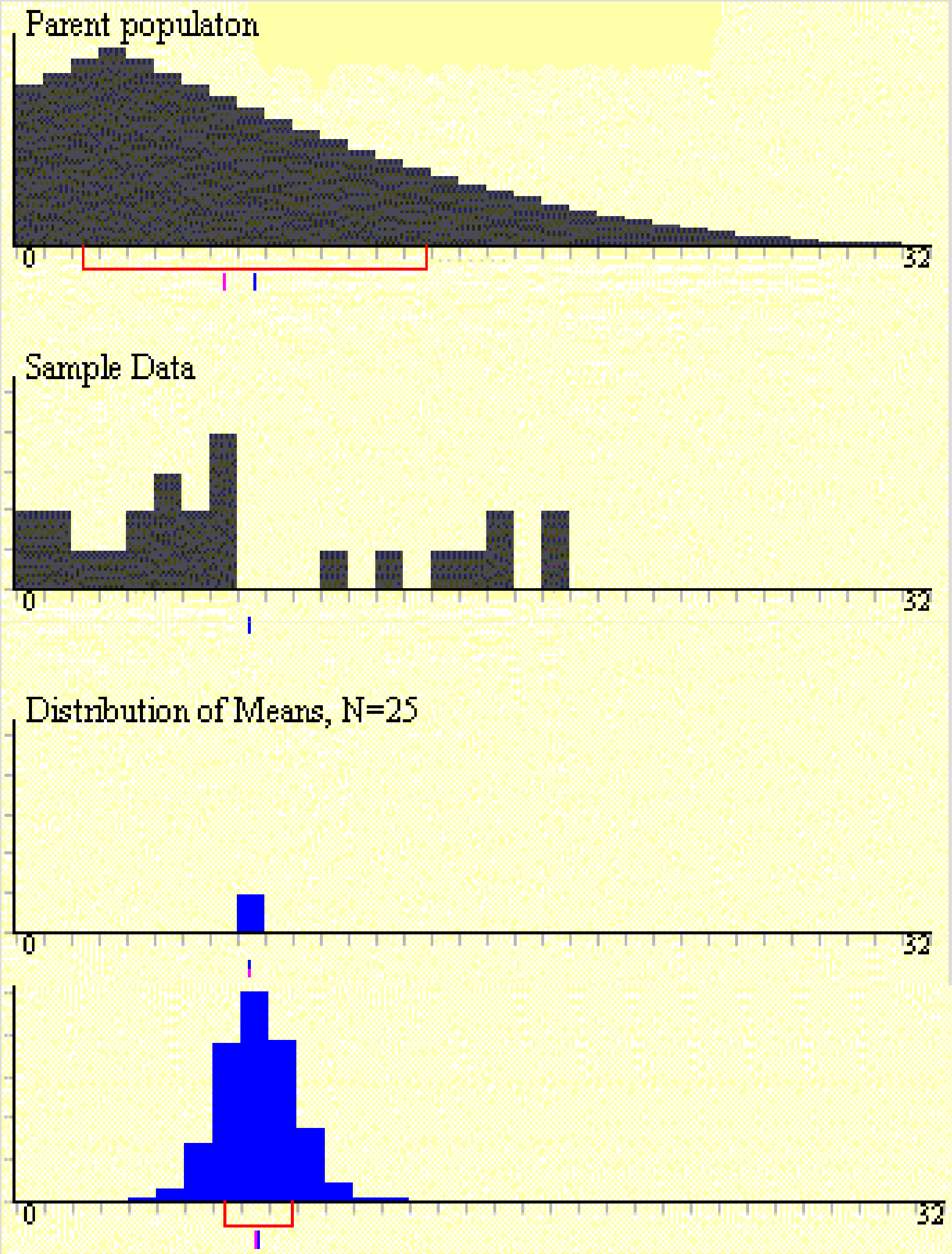
sample



population

sample

mean of this sample



population

sample

mean of this sample

distribution of many sample means

Are your data normally distributed?

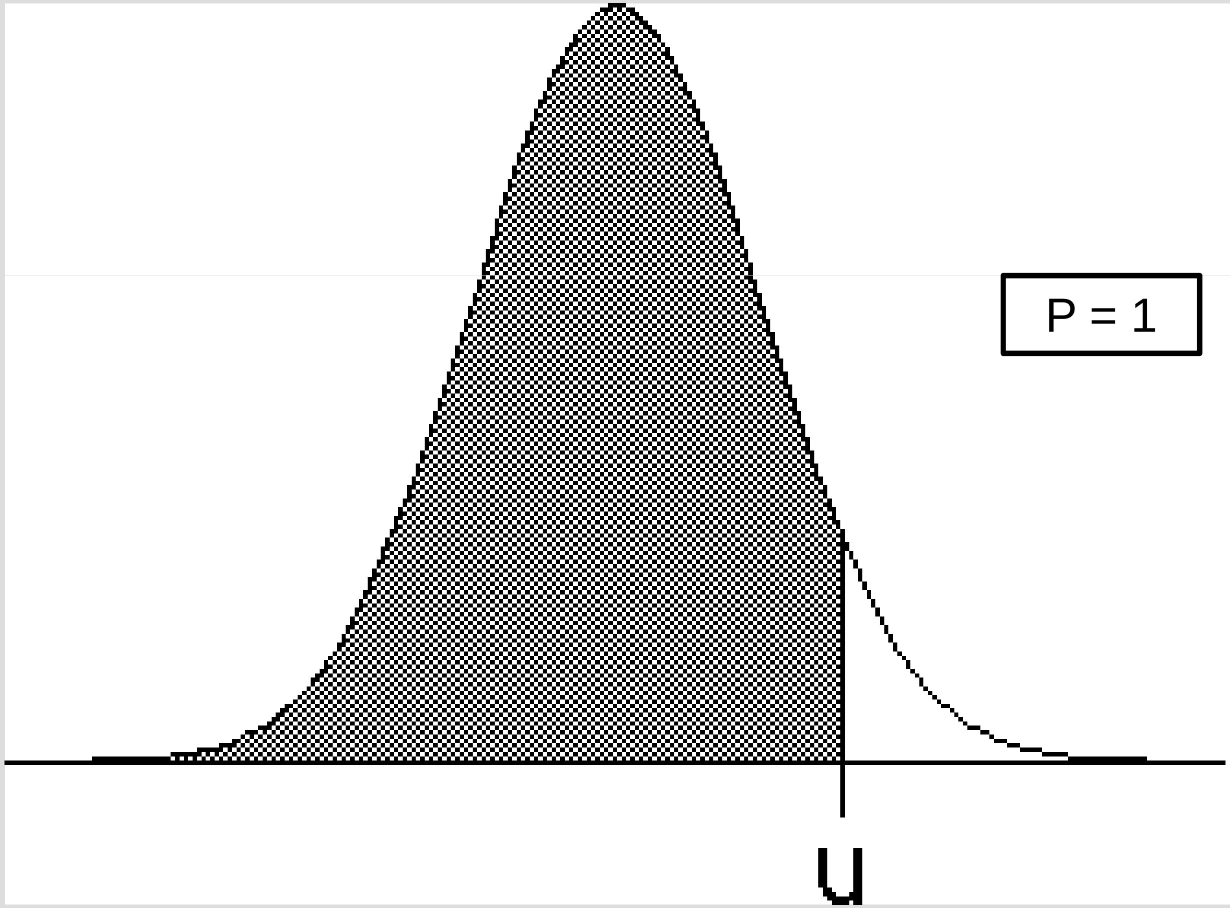
How many Ns do we need in order to assume that the distribution of sample means is normally distributed?

N = 30

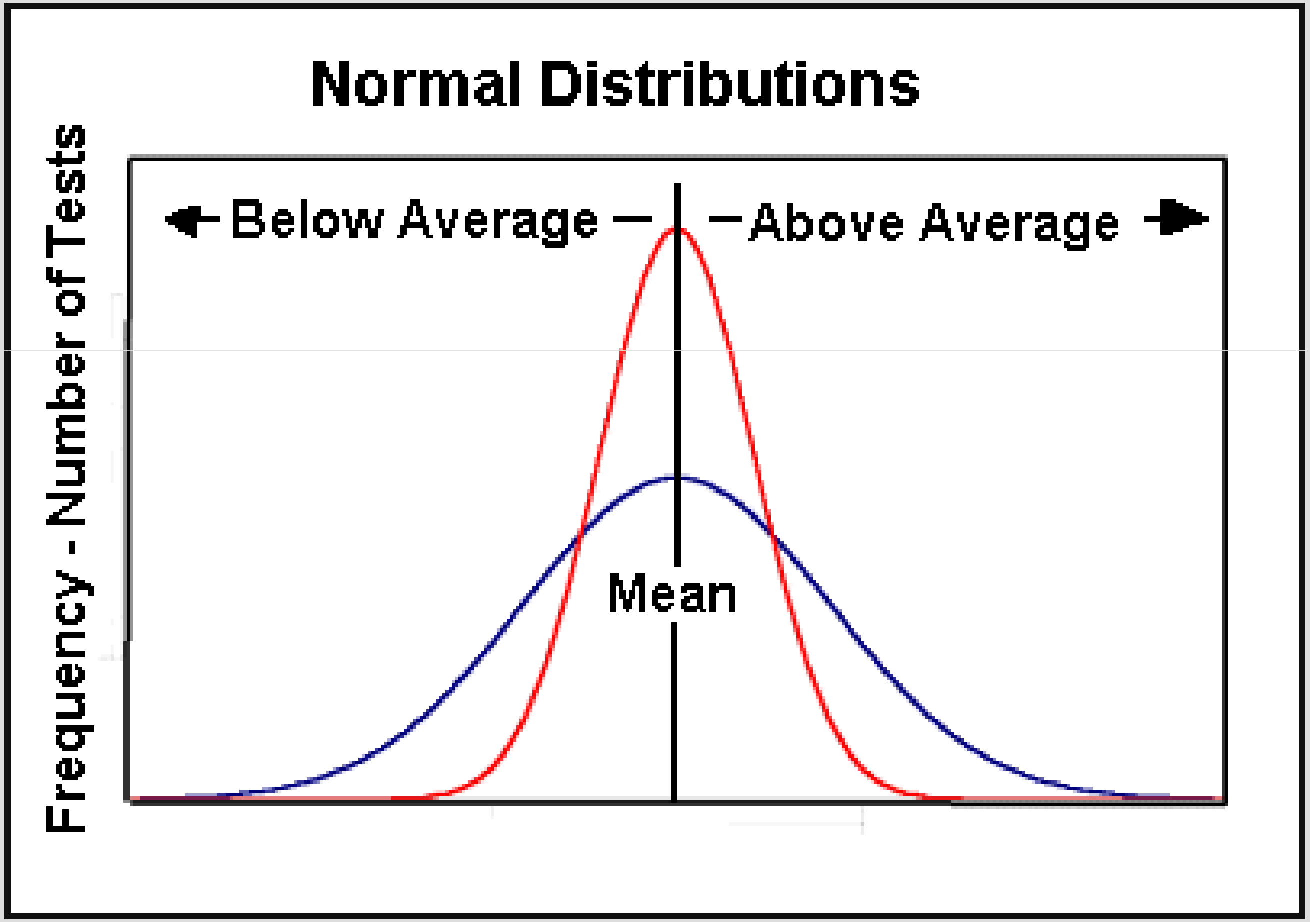
Three important factors:

- The distribution in the parent population (normal - skewed).
- The number of observations in individual samples.
- The total number of samples.

Probability



The standard normal distribution



The standard normal distribution

The x-axis of the standard normal distribution is defined in terms of z-scores, i.e. the distance between the mean and an individual data point.

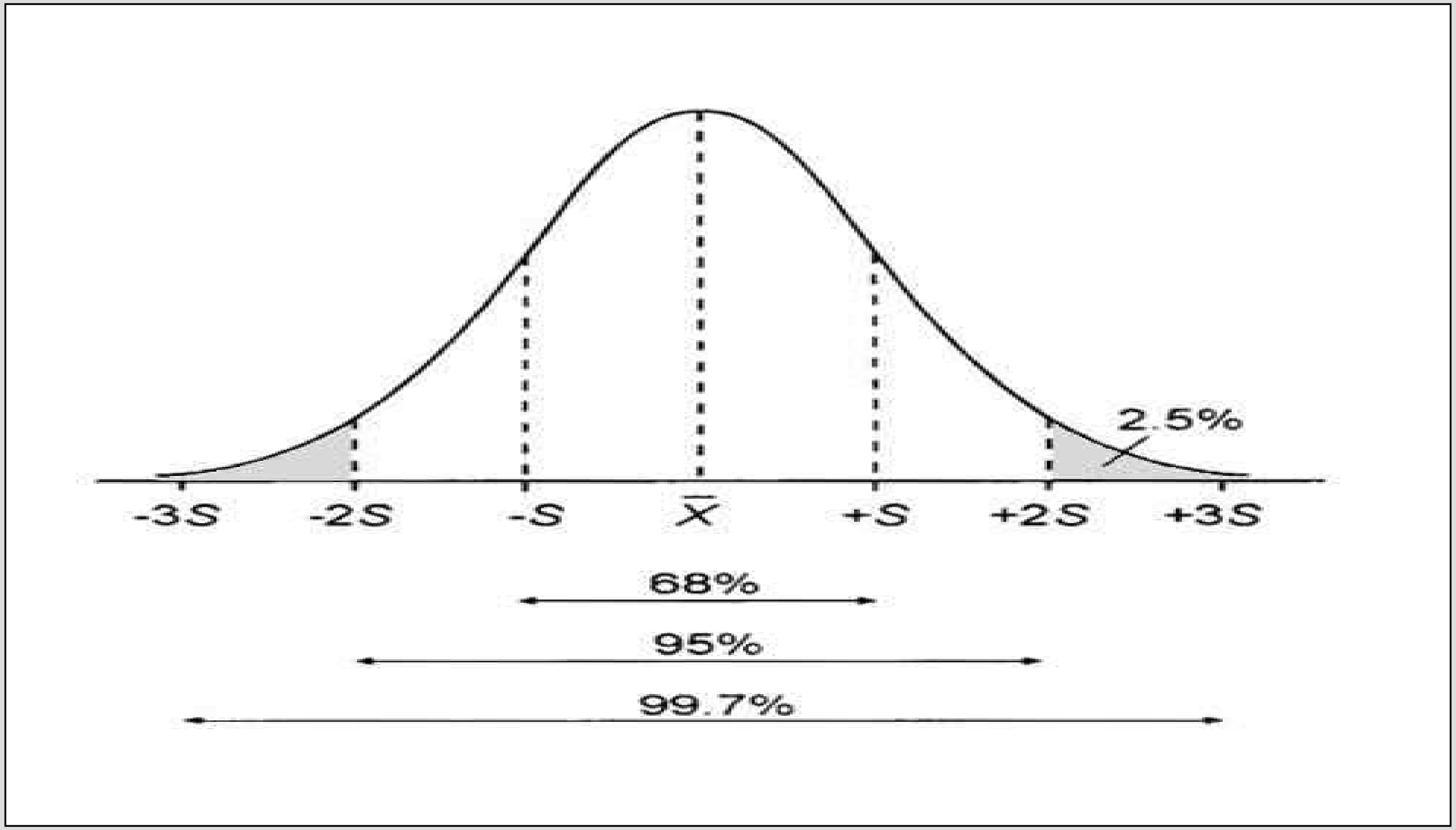
$$z = \frac{x - \mu}{\sigma}$$

x = individual data point

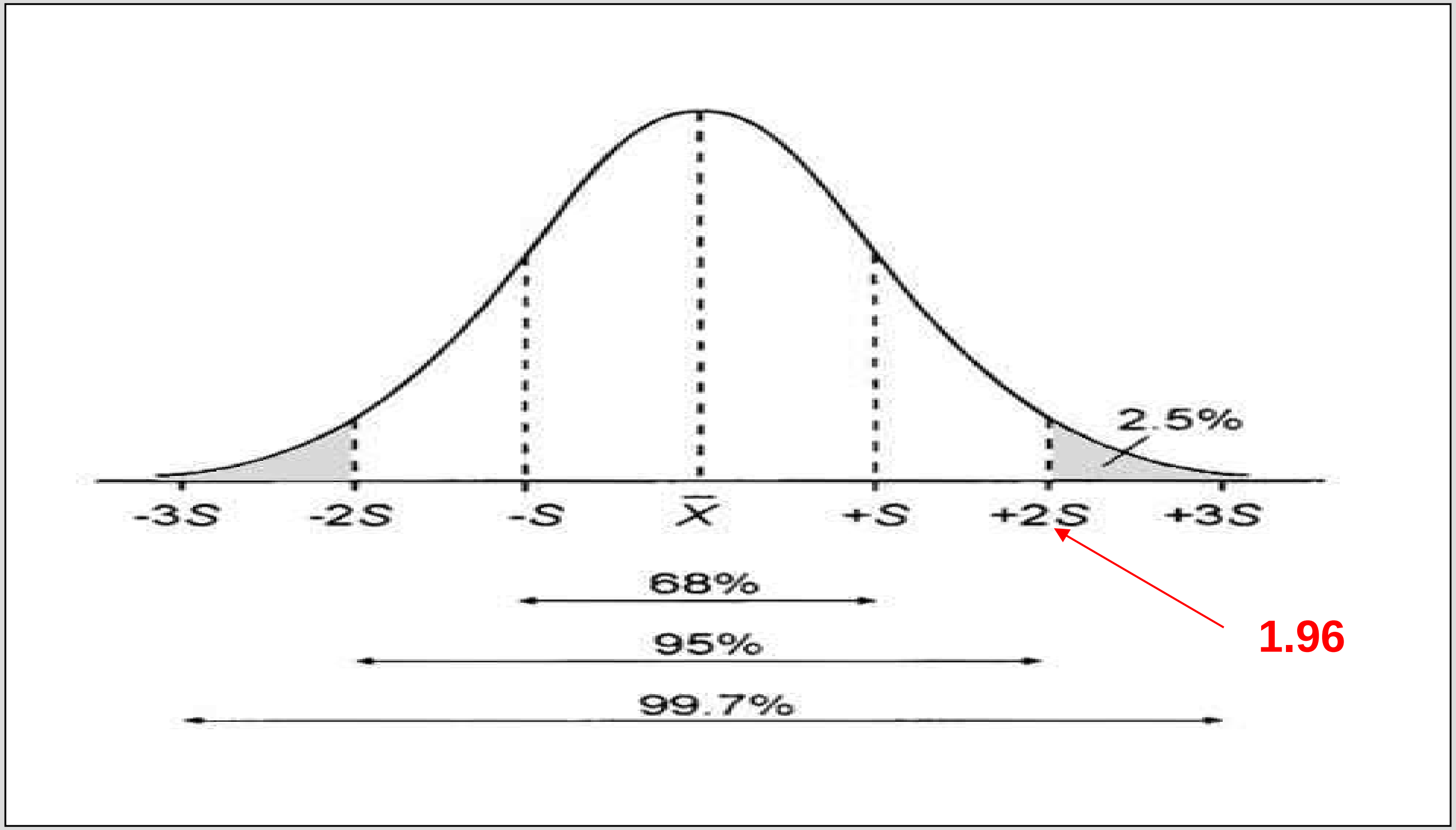
μ = mean

σ = standard deviation

The empirical rule



The empirical rule



Parametrical and non-parametrical tests

Non-parametrical tests are used …

- for ordinal data
- for interval data that are not normally distributed
- for small samples

Confidence intervals

Confidence intervals indicate a range within which the mean (or other statistical parameters) of the true population may lie given the values of your sample and assuming a certain probability.

In order to determine confidence intervals for the mean of the sample means we need:

- The mean of the sample means (given)
- The standard error
- A certain degree of confidence (e.g. 95%)

Standard error

The standard error is the equivalent of the standard deviation of the sample distribution.

The standard error indicates the degree to which the individual sample means deviate from the mean of the sample means.

Given that the mean of the sample means is a good approximation of the population mean, the standard error of the sample means will also tells us something about the spread of the mean in the true population.

Standard error

We can estimate the standard eror of the population mean from one sample:

$$SE = \frac{SD}{\sqrt{n}}$$

For any given sample, if we divide the sample SD by the square root of the sample size, we get a good approximation of the standard error.

Confidence intervals

degree of certainty $\times$ standard error = x

sample mean +/- x = confidence interval

Confidence intervals

95% degree of certainty = 1.96 [z-score]

$1.96 \times 0.2701 = 0.53$

1.66 +/- 0.53 = **0.97–2.03**

We can be 95% certain that the population mean lies in the range between 0.97 and 2.03.

Exercise: Given the sample (2, 5, 6, 7, 10, 12) determine the 95% confidence interval of the population mean.

Mean: 7

$$SD = \frac{(2-7)^2 + (5-7)^2 + (6-7)^2 + (7-7)^2 + (10-7)^2 + (12-7)^2}{6 - 1 = 3.58} = 3.58$$

$$SE = \frac{3.58}{\sqrt{6}} = 1.46$$

$$CI = 1.46 \times 1.96 = 2.86$$

CI = 7 +/– 2.86 = 4.14 – 9.86