1. The stages of a malignant disease (cancer) is recorded using the symbols 0, I, II, III, IV. We say that the scale used is:
   A. Alphanumeric
   B. Numerical
   C. Ordinal
   D. Nominal

2. The fundamental statistical indicators are:
   A. Mean
   B. Median
   C. Variance
   D. Standard deviation

3. The average of a series of numerical values is:
   A. The sum of the values divided by their number
   B. Lower than the minimum value in the series
   C. Lower than the maximum value in the series
   D. An indicator of central tendency for the values of the series

4. Standard deviation:
   A. is the square root of variance
   B. is measured using the unit of the variable
   C. is measured using the squared unit of the variable
   D. has values generally comparable with the average value

5. If the average of a series of values is 10 and their variance is 4, then the coefficient of variation (= the ratio standard deviation / average) is:
   A. 40%
   B. 20%
   C. 80%
   D. 10%
   *Standard deviation is square root of variance $\sqrt{4} = 2$, so the coefficient of variation is $2/10 = 0.2 = 20%$

6. Having two sets of data, we can compare their scattering as follows:
   A. For approximately equal average values, the one with a higher standard deviation is more scattered
   B. For approximately equal standard deviation values, the one with a higher average is more scattered
   C. For approximately equal standard deviation values, the one with a lower average is more scattered
   D. If both the averages and standard deviations differ much between the series, we can compare scattering using the coefficient of variation

7. The median of a series of numerical values is:
   A. Equal to the average
   B. A graph or chart
   C. A number
   D. A frequency table
8. The median of a series of numerical values is:
   A. A value for which half of the values are higher and half of the values are lower
   B. The value located exactly midway between the minimum and maximum of the series
   C. The most commonly encountered values among the series
   D. A measure of the eccentricity of the series

9. If a series of values consists of 21 numbers, then, for finding the median, we ordered the series ascending and we use:
   A. The 11th value in the ordered series
   B. The mean between the 10th and 11th values
   C. The mean between the 11th and 12th values
   D. The 10th value in the ordered series

10. The first quartile of a series of values is:
    A. The value in the ordered series located at 25% of the number of values in the series
    B. The value of the ordered series located at 75% of the number of values in the series
    C. The numeric value for which a quarter of the series’ values are lower
    D. The numeric value for which a quarter of the series’ values are higher

11. Relative risk:
    A. Shows the relationship between a factor assumed to influence the occurrence of disease, and the disease
    B. is the ratio of the risk of disease for those exposed and those not exposed to that risk factor
    C. Cannot be greater than 1
    D. is expressed as a percentage

12. If on a group of 457 patients, for a risk factor we calculated a Relative Risk RR= 12.74, the possibility of developing the disease being investigated is:
    A. very high when exposed to the factor
    B. very small when exposed to the factor
    C. the same in the case of exposure in the case of non-exposure
    D. lower in the exposed than in the unexposed, RR being less than 100

13. If on a group of 457 patients, for a risk factor we calculated an Odds Ratio OR= 12.74, the possibility of developing the disease being investigated is:
    A. very high when exposed to the factor
    B. very small when exposed to the factor (protective factor)
    C. the same in the case of exposure in the case of non-exposure
    D. lower in the exposed than in the unexposed, OR being less than 100

14. The Sensitivity (SN) of a clinical trial
    A. is the ratio of sick patients, diagnosed as positive, and the total number of sick patients.
    B. is the ratio of healthy subjects, diagnosed as negative, and the total number of healthy subjects
    C. is the ratio of sick patients, diagnosed as negative, and the total number of patients.
    D. is the ratio of sick patients, diagnosed as negative, and the total number of healthy persons

15. A clinical trial is more valuable when:
    A. Sensitivity and Specificity have higher values
    B. Sensitivity is higher than specificity
    C. Specificity is higher than Sensitivity
    D. The sensitivity and specificity values are close, even equal, regardless of their values
16. In a contingency table that shows data from a clinical trial is good to have high values for:
   A. sick subjects, diagnosed as negative
   B. sick subjects, diagnosed as positive
   C. healthy subjects, diagnosed as negative
   D. healthy subjects, diagnosed as positive

17. For a clinical trial, the Sensitivity is Sn = 0.562 and Specificity is Sp = 0.893. This means that:
   A. The test is a valuable test because both indicators are more than 50%
   B. The test is a worthless test, since it gives errors when detecting both sick and healthy subjects
   C. The test is a worthless test, because the sensitivity is too low (lower than 75%)
   D. a perfect test

18. A regression line is a straight line which:
   A. is located as close as possible to all the points of a scatter chart
   B. is defined by an equation having 2 parameters: the slope and the intercept
   C. provides an approximate relationship between the values of two parameters
   D. is parallel to one of the coordinate axes

19. Pearson correlation coefficient, denoted by r, measures:
   A. The scattering strength of data for a statistical series
   B. The strength of the correlation between the mean and median
   C. The strength of the correlation between two numerical parameters
   D. The tendency of simultaneous increase or decrease, or inverse evolution, for two numerical parameters

20. The correlation coefficient computed for two parameters measured in 429 patients is r=0.829. This means that:
   A. The two parameters are directly correlated, and the link is weak - r is positive and close to 0
   B. The two parameters are inversely correlated, and the link is strong - r is negative and close to 1
   C. The two parameters are directly correlated, and the link is strong - r is positive and close to 1
   D. There are too few cases (under 30) and we do not trust this coefficient’s value

21. For a Histogram chart the following statements are true:
   A. Each bar (class or column) is the same width
   B. The height of the bars is proportional to that class’s absolute frequency (number of individuals in the class)
   C. The width of the bars (classes) is obtained by dividing the difference between the maximum and the minimum values in the series we represent to the number of desired classes
   D. We do not lose any information of the original data series by making such a chart

22. A Frequency Polygon is:
   A. A statistical indicator that shows the scattering of a series of values
   B. A graph representing by a broken line the absolute frequencies of classes of a data series
   C. A graph that contains exactly the same information as the corresponding histogram
   D. A graph that contains less information than the corresponding histogram

23. A Gauss curve is a curve which is:
   A. Symmetrical to the mean
   B. Symmetrical to the vertical axis, which passes through 0
   C. Has a maximum where the average of the series of values is located
   D. Tends to 0 towards plus infinity and minus infinity (i.e. very low and very high values)
24. A Gauss curve, the curve of a normal distribution, has the following features (where \( m = \text{mean}, \ s = \text{standard deviation} \)):

A. in the interval \([m – 1s; m + 1s]\) about \(2/3 (~ 68\%)\) of the series’ values are located
B. in the interval \([m – 2s; m + 2s]\) about \(95\%\) of the series’ values are located
C. in the interval \([m – 3s; m + 3s]\) about \(99\%\) of the series’ values are located
D. in the interval \([m – 1s; m + 1s]\) about 50% of the series’ values are located

25. The Confidence Interval for the mean, calculated for a series of values, has the interpretation:

A. The true mean, the one that approximates the population’s mean, is almost certain inside the confidence interval
B. The true variance is almost certain inside the confidence interval
C. The true median is almost certain inside the confidence interval
D. It is an interval that contains almost all the values of the series

26. The Student's t test is:

A. a parametric test
B. a nonparametric test
C. a test for comparing averages
D. a test for comparing variances

27. Which of the following tests are parametric tests:

A. ANOVA
B. Student
C. Wilcoxon
D. Kruskal-Wallis

28. The result of a statistical test, denoted \( p \), shall be interpreted as follows:

A. the null hypothesis \( H_0 \) is rejected if \( p < 0.05 \)
B. the null hypothesis \( H_0 \) is rejected if \( p > 0.05 \)
C. the alternate hypothesis \( H_1 \) is rejected if \( p > 0.05 \)
D. the null hypothesis \( H_0 \) is accepted if \( p < 0.05 \)

29. The null hypothesis (\( H_0 \)) when comparing two means shall be interpreted as:

A. Data do not support the Hypothesis that the populations' means are different
B. The compared values are different
C. The two sampling averages do not differ significantly
D. The two populations, from which the compared values were sampled, do not differ

30. If, after performing a Student test for comparison of means, we obtain \( p = 0.0256 \), then:

A. We reject \( H_0 \) and accept \( H_1 \)
B. We accept \( H_0 \)
C. We reject \( H_1 \)
D. We cannot decide