



Detailed description of the services provided by the the Centre of Biostatistics and Bioinformatics Analysis

Detailed description of the services provided

The Centre provides a wide range of comprehensive services in the field of advanced biostatistical and bioinformatic analysis. The services are offered primarily to scientists conducting clinical, epidemiological, population, molecular and translational research. The services offered by the Centre are based on statistical methods and computational techniques included in the list below:

Planning of experiments

- Formulating working and null hypotheses
- Selection of appropriate methods of statistical analysis and tests
- Selection of appropriate control group
- Sample size estimation
- Randomization methods

Basic descriptive statistics

- Descriptive statistics (mean, SD, median, IQR, range, etc.)
- Descriptive statistics (summary tables)
- Basic correlation matrices
- Contingency tables (chi² tests)
- Frequency tables
- Tests assessing the normality of distribution
- Area under the curve (AUC)
- Confidence intervals

Data distribution evaluation

- The Shapiro-Wilk W test
- The Kolmogorov-Smirnov test

Student's t-tests

- The Student's t-test for a single sample
- The Student's t-test for paired samples
- The Student's t-test for two independent samples

Significance tests for more than 2 samples: analysis of variance (ANOVA)

- One-way analysis of variance (ANOVA)
- Analysis of covariance (ANCOVA)
- Multivariate analysis of variance (MANOVA)
- Analysis of variance for unbalanced designs



- General linear model (GLM)
- Repeated measures analysis of variance
- Area under the curve (AUC)
- Significance tests for post-hoc multiple comparisons (the Student's t-test with Bonferroni correction, Fisher's LSD test, Scheffe's test, Tukey's test, Duncan's test, etc.)
- Significance tests for planned comparisons (contrast analysis, Dunnett's test, etc.)
- Tests to verify analysis assumptions (Cochran's C test, Hartley's test, Bartlett's test, Levene's test)

Evaluation of agreement between two methods

- Bland-Altman plot and analysis
- Tests based on χ^2 and z statistics

Significance tests for proportions based on χ^2 and z statistics

Correlation analysis

- Pearson, Spearman, Tau Kendall, Gamma correlations
- Partial correlations within the multiple regression analysis
- Correlation matrices
- Canonical functions correlation

Regression analysis

- Linear Regression
- Multiple regression and partial correlations
- Analysis of covariance (ANCOVA)
- General linear model (GLM)
- Logistic regression
- Cox Regression
- Probit regression
- Nonlinear regression

Curve fitting methods

- Distribution curve fitting methods (normal, exponential, gamma, log-normal, χ^2 , Weibull, binomial, Poisson, Bernoulli, etc.)
- Polynomial regression
- Nonlinear regression

Analysis of contingency tables

- Contingency tables (χ^2 tests)
- Frequency tables
- Q Cochran test
- Log-linear models
- Mantel-Haenszel test
- χ^2 tests for trend

Nonparametric methods



- The Mann-Whitney U Test
- The Kruskal-Wallis H test
- Friedman's test
- Cumulative frequency analysis
- Spearman, Tau Kendall, and Gamma correlations
- Contingency tables (Chi-Square Test)
- McNemar's test for dependent samples
- The Q Cochran test
- Wilcoxon tests
- Chi² compliance tests
- Kaplan-Meier curves (the log-rank tests)
- ROC and area under the curve analysis (AUC)

ROC curve analysis

- Analysis of a single ROC curve and determination of the cut-off value
- Comparison of two ROC curves
- Area Under the Curve (AUC)

Multivariate analyses

- Logistic regression
- General linear models
- Generalized linear and nonlinear models
- Factor analysis
- Canonical correlation analysis
- Discriminant function analysis
- Principal component analysis
- Cluster analysis
- Multivariate analysis of variance (MANOVA)
- Correspondence analysis
- Log-linear analysis of contingency tables
- Survival analysis
- Cox proportional hazard regression
- Nonlinear estimation and regression

Cluster analysis

- Fuzzy clustering methods
- Hierarchical methods/Dendrograms
- K-means clustering
- Double dendrograms (heatmaps)

Survival analysis

- Cumulative frequency curves
- Kaplan-Meier curves (log-rank tests)
- Cox regression



Meta-analyses

- Meta-analyses for RR, OR values
- Meta-analyses for mean values
- Meta-analyses for proportions

Data exploration, data mining and machine learning techniques

- CART and CHAID classification trees
- Random forest classification methods
- MARSplines regression
- Support Vector Machines techniques
- Neural networks

Graphical procedures

2D and 3D column charts

- univariate
- multivariate

Bland-Altman plot

Box (box-and-whiskers) plots

Histograms

- circular histograms
- column histograms
- circular histograms

Contour plots

3D surface plot

Dendrograms

- hierarchical dendrograms
- double dendrograms (heatmaps)

Scatter plots

Bar charts

Forest Plots for meta-analyses

Heatmaps

- double dendrograms
- correlation matrices



- 3D surface plots
- contour plots

Kaplan-Meier curves

ROC curves

Data scatter plots

- 2D and 3D scatter plots
- scatter plots with error bars

3D plots

- 3D data scatter plots
- 3D surface plots
- 3D column plots
- correspondence analysis charts

Data processing

- Data import
- Data export
- Data filtering and clearing
- Data transformation
- Data normalization
- Randomization
- Generating simulated data sets
- Merging data sets

Preparation of reports

- Summary (a brief presentation of the purpose of conducted analyses, the methods used, the most important results and conclusions)
- Summary of the analysed research problems and tested hypotheses
- Detailed description of the analytical procedure and methods used
- Detailed summary and discussion of obtained results of the analysis
- Detailed discussion of the conclusions drawn from the statistical analysis
- Emphasizing the strengths and weaknesses of the experiment and conducted analyses
- Suggestions and hints for further research

Statistical consultations