

# Методы Прикладной Статистики в Биологии

1. Планирование статистического эксперимента. Эксперимент и наблюдательное исследование. Простая случайная выборка. Отклонения в выборках и их классификация.
2. Описательные статистики. Способы графического изображения выборок: гистограмма, бокс-плот, дот-плот, и другие. Способы представления двумерных данных. Визуализация данных с использованием программы R-statistics и пакета графических грамматик ggplot2.
3. Обзор элементарных сведений из теории вероятностей. Условная вероятность. Формула Байеса. Формула полной вероятности. Случайные величины. Математическое ожидание, дисперсия. Ковариация и корреляция. Геометрический смысл.
4. Дискретные распределения. Равномерное, биномиальное, геометрическое и Пуассоновское распределения. Примеры дискретных распределений в биологических задачах.
5. Приближение биномиального распределения Пуассоновским. Выборки с возвращением и без возвращения. Гипергеометрическое распределение.
6. Непрерывные распределения. Нормальное распределение. Распределение пропорций и его связь с биномиальным распределением. Центральная предельная теорема. Аппроксимация биномиального распределения нормальным и поправка на непрерывность. Применимость аппроксимаций.
7. Теория точечного оценивания. Несмещенность и эффективность оценок. Среднеквадратичное отклонение. Теорема Штейнера. Математическое ожидание и дисперсия выборочного среднего. Примеры несмещенных и эффективных оценок.
8. Интервальное оценивание. Доверительные интервалы и их интерпретация. Уровень доверия. Стандартная ошибка. Поправка на конечный размер популяции.
9. Интервальное оценивание пропорций и разности пропорций. Оценки для среднего и разности средних в случае известных и неизвестных стандартных отклонений. Распределение Стьюдента и условия его применимости. Случай равных дисперсий и оценка дисперсии объединенной (pooled) выборки.
10. Проверка гипотез. Ошибки первого и второго рода. Уровень значимости и сила теста. Р-значение и его интерпретация. Тестирование гипотез с использованием доверительных интервалов и тестовых статистик. Критические значения. Согласованность гипотез и доверительных интервалов. Биологические задачи.

11. Распределение хи-квадрат. Условия применимости в задачах. Критерий согласия Пирсона. Таблицы сопряженности. Выборки без возвращения и точный тест Фишера. Биологические задачи.
12. Свойства выборочного стандартного отклонения и интервальное оценивание дисперсии. Проверка гипотез с использованием распределения хи-квадрат. Распределение Фишера. Свойства квантилей. Проверка гипотез с использованием распределения Фишера.
13. Модель парной регрессии. Метод наименьших квадратов. Интерпретация коэффициента наклона. Остатки и их свойства. Выбросы и влиятельные значения. Обобщенные линейные модели.
14. Стандартная ошибка коэффициента наклона в парной регрессии. Доверительные интервалы и тестирование гипотез в модели регрессии. Доверительные и прогнозные интервалы.
15. Коэффициент корреляции и коэффициент детерминации, их интерпретация. Разложение суммы квадратов вычетов. Статистические свойства оценок коэффициента корреляции. Применение к анализу дисперсии зависимых выборок.
16. Однофакторный и двухфакторный дисперсионный анализ (ANOVA). Проверка гипотез в дисперсионном анализе. Предположения. Доверительные интервалы для одновременного оценивания разности средних. Разложение суммы квадратов.
17. Множественная линейная регрессия. Многомерное нормальное распределение и предположения о нормальности. Остаточная дисперсия, коэффициент детерминации, и его интерпретация. Мультиколлинеарность.
18. Обобщенные линейные модели. План анализа, матрица эксперимента, вектор контрастов и функция связи. Эспоненциальное семейство распределений. Частные случаи факторного анализа и многомерной регрессии. Смешанные модели (ANCOVA).
19. Методы редукции размерности. Анализ главных компонент. Матрица ковариации. Счета и нагрузки. Сингулярное разложение. Интерпретация результатов анализа главных компонент. Многомерное шкалирование. Другие методы кластеризации данных.

# Methods of Applied Statistics in Biology

1. Experiment planning. Experiment vs. observational study. Simple random sample. Sources of bias and their classification. Descriptive statistics. Graphical representation of samples by histograms, box-plots etc.
2. Descriptive statistics. Graphical representation of samples and sample statistics: histogram, box-plot, dot-plot, and others. Graphical representation of paired datasets. Data visualization by using R-statistics and graphical grammars package ggplot2.
3. Review of elementary concepts from probability theory. Conditional probability. Bayes formula and full probability. Random variable, expected value and variance. Covariance and correlation. Geometric meaning.
4. Discrete probability distributions. Uniform, binomial, geometric, and Poisson distributions. Examples of discrete distributions in biological problems.
5. Poisson approximation to binomial. Sampling with and without replacement. Hypergeometric distribution.
6. Continuous distributions. Normal distribution. Distribution of proportions and its relation to binomial distribution. Central limit theorem. Normal approximation to binomial distribution and continuity correction. Limitations of approximations.
7. Point estimation theory. Unbiased and effective estimators. Mean squared error. Steiner's theorem. Expectation and variance of the sample mean. Examples of unbiased and effective estimators.
8. Interval estimation theory. Confidence intervals and their interpretation. Confidence level. Standard error. Correction for finite population size.
9. Interval estimation of proportions and proportion differences. Estimates for the mean and difference of means in case of known and unknown variance. Student t-distribution and its applicability. Equal variances and estimation of variance in the pooled sample.
10. Hypothesis testing. Type I and type II errors. Significance level and power of the test. P-value and its interpretation. Hypothesis testing by using confidence intervals and test statistics. Critical values. Correspondence between hypotheses and confidence intervals. Biological problems.
11. Chi-square distribution and its applicability. Pearson goodness of fit test. Contingency tables. Sampling without replacement and Fisher exact test. Biological problems.

12. Properties of sample standard deviation and interval estimates. Chi-square distribution in hypothesis testing. Fisher distribution. Properties of quantiles. Fisher distribution in hypothesis testing.

13. Pairwise regression model. Method of least squares. Interpretation of the regression slope. Residuals and their properties. Outliers and influential scores. Generalized linear models.

14. Standard error for the regression slope. Confidence intervals and hypotheses testing in linear regression. Prediction intervals and confidence intervals.

15. Correlation coefficient, determination coefficient, and their interpretation. Decomposition of sum of squares of residuals. Statistical properties of estimators of correlation coefficient. Applications to the analysis of variance in dependent samples.

16. One-way and two-way analysis of variance (ANOVA). Hypothesis testing in ANOVA, assumptions. Confidence intervals for simultaneous estimation of population means. Decomposition of sum of squares of residuals.

17. Multiple regression. Multi-dimensional normal distribution and normality assumptions. Residual variance, determination coefficient, and its interpretation. Multicollinearity.

18. Generalized linear models. Analysis plan, experiment matrix, contrasts vector and link function. Exponential family of distributions. Particular cases of factor analysis and multiple regression. Mixed models (ANCOVA).

19. Dimension reduction methods. Principal component analysis. Covariance matrix. Scores and loadings. Singular decomposition. Interpretation of principal component analysis. Multidimensional scaling. Other data clusterization methods.