

Статистически анализ на данни и моделиране със STATISTICA и R
Statistical data analysis and modeling using STATISTICA and R

Лектор: проф. дн Нейко Матеев Нейков, НИМХ

Prof. Neyko M. Neykov, DSc.,

Tel: 02 462 45 97; 0884038744

E-mail: neyko.neykov@meteo.bg

Хорариум: 30 учебни часа

Анотация: Основната цел на курса е да запознае слушателите с някои често използвани методи за статистически анализ на данни и моделиране като: първични статистически характеристики, параметрични и непараметрични тестове за еднородност на едномерни и многомерни извадки, разкриване на несъгласувани наблюдения в данните, линеен множествен и нелинеен регресионен анализ; регресионен анализ на бинарни, категорийни (номинални, ординални), бройни, гамма, Вайбул и други асиметрично разпределени данни, квантилна регресия, дисперсионен анализ, анализ на главните компоненти и канонична корелация, дискриминантен анализ, клъстерни методи, моделиране със смес от вероятностни разпределения. Ще бъде изяснена връзката на различните типове данни с основните вероятностни (дискретни и непрекъснати) разпределения. Упражненията ще се провеждат с анализ на данни от областта на метеорологията, хидрологията и околната среда със статистическите програмни пакети STATISTICA и R.

Abstract: The main objective of this course is to introduce some widely used methods for statistical data analysis and modeling such as: data description, exploratory data analysis, parametric and nonparametric tests for homogeneity of univariate and multivariate data, outlier detection, linear and nonlinear regression analysis, binary, nominal and ordinal logistic regression, Poisson regression for count data, regression for asymmetric data such as gamma and Weibull and other asymmetric distributions, quantile regression, univariate one and two way analysis of variance, principal component analysis, canonical correlation, discriminant analysis, cluster methods and mixture of distributions.

Програма на лекциите:

1. Типове данни в статистическите изследвания и връзката им с основните вероятностни разпределения. Първична обработка на данни. Числови характеристики на едномерни и многомерни извадки. Емпирична функция на разпределение. Графично представяне на данните. Вероятностни и квантилни плотове **(3 часа)**

2. Параметрични статистически методи. Проверка на хипотези за еднородност на две извадки. Проверка на хипотези за равенство на средните на две извадки. Доверителни интервали. Еднофакторен и двуфакторен дисперсионен анализ с фиксирани ефекти. Процедури за множествени сравнения на средните **(3 часа)**

3. Непараметрични методи. Тестове на Уилкоксон, Ман-Уитни и Ман-Кендал. Тестове на Колмогоров-Смирнов и хи-квадрат за проверки на хипотези за вида на разпределението. Непараметричен еднофакторен и двуфакторен дисперсионен анализ по Кръскал-Уолис и Фридмен. Тест на Петит **(3 часа)**

4. Линеен и нелинеен регресионен анализ. Оценки по метода на най-малките квадрати. Проверка на хипотези. Доверителни интервали. Анализ на регресионните остатъци. Процедури за избор на значими предиктори в регресионни модели. Информационни критерии на Акайке (AIC) и Шварц (BIC). Техники за разкриване на несъгласувани наблюдения в данните **(4 часа)**

5. Квантилна регресия **(1 час)**

6. Регресионни модели на бинарни, номинални, ординални и бройни данни. Оценки по метод на максималното правдоподобие (ММП) и проверка на хипотези **(5 часа)**

7. Регресионни модели на асиметрично (експоненциално, гама, Вайбул) разпределени данни. Оценки по ММП и проверка на хипотези **(3 часа)**

8. Многомерни статистически методи. Проверка на хипотези за еднородност на две многомерни извадки. Процедури за разкриване на несъгласувани наблюдения в многомерни данни. Метод на главните компоненти. Редуциране на размерността на пространството на данните по метода на главните компоненти. Факторен анализ. Канонична корелация **(3 часа)**

9. Методи за класификация и дискриминация. Линеен дискриминантен анализ. Клъстерни методи **(3 часа)**

10. Статистически анализ на нееднородни данни със смеси от вероятностни разпределения **(2 часа)**

Литература:

1. Afifi, A., May, S. and Clark, V. A. (2011). Practical Multivariate Analysis, Fifth Edition, CRC Press.
2. Dobson. A. (2002). An introduction to Generalized Linear Models. 2nd edition, Chapman and Hall.
3. Faraway, J. (2009). Linear Models with R. 2nd edition, Chapman and Hall, CRC Press.
4. Helsel, D. R. and Hirsch, R. M. (2002). Statistical Methods in Water Resources. Publication available at: <http://water.usgs.gov/pubs/twri/twri4a3/>
5. Wilks, D. S. (2011). Statistical Methods in the Atmospheric Sciences. Academic Press. 3rd ed.

Начин на оценяване: Курсов проект