Topics for Primary seminar, 2018-19

1. Welch two-sample t-test and beyond (Aidan Rocke)
Explaining the standard (exact as well as asymptotic) two-sample t-test – the model, the null and the alternative hypothesis, construction of the test statistic.
Introducing the Welch version of the t-test
A detailed explanation of the choice of the degrees of freedom.
Illustrating on the real data set, implementation in R software.
Generalization to more than two samples.

Literature:
Anděl J.: Statistické metody, Matfyzpress, Praha, 1998.
Kulich, M.: *NMSA 331, Poznámky k přednášce*,
http://msekce.karlin.mff.cuni.cz/~omelka/Soubory/nmsa331/ms1.pdf
Kurková M (2006). Dvouvýběrový t-test v případě nestejných rozptylů. Bachelor thesis. Faculty of mathematics and physics. Charles University.
Miller Jr, R. G. (1997). Beyond ANOVA: basics of applied statistics. Chapman and Hall/CRC. Chapter 2.3
Satterthwaite F. E.: An Approximate Distribution of Estimates of Variance Components, Biometrics Bulletin 2 (1946), 110–114.
Welch B. L.: The Significance of the Difference Between Two Means when the Population Variances are Unequal, Biometrika 29 (1938), 350– 362.

2. Two sample Wilcoxon test and beyond (Aleksandra Duda)
Explaining the standard two sample Wilcoxon test – the model, the null and the alternative hypothesis, construction of the test statistic.
Mann-Whitney rank test
Generalizing Mann-Whitney rank test to more general hypothesis
Illustrating what happens if under the altternative hypothesis the distribution does not change only in location.
Dealing with ties

Literature:
Chung, E., & Romano, J. P. (2016). Asymptotically valid and exact permutation tests based on two-sample U-statistics. Journal of Statistical Planning and Inference, 168, 97-105.
Šlampiak, T. (2016). Wilcoxonův dvouvýběrový test. Bachelor thesis. Faculty of mathematics and physics. Charles University.

3. Q-Q plots (Viktor Dolnik)
Normal Q-Q plot – construction, implementation in R, examples of normal distributions as well distributions that are not normal. Comparison with histogram. Use in linear regression.
General Q-Q plots
Q-Q plots to compare two samples

Literature: Zvára: Regrese, kapitola 8.7, MATFYZPRESS, 2008 (in Czech),
Wilk, M.B., Gnanadesikan, R. (1968), "Probability plotting methods for the analysis of data",

Biometrika, 55 (1): 1–17

4. Tests of normality (Martina Nováková)
- Shapiro-Wilk test, D'Agostino test
- implementation in R – software
Literature: Zvára: Regrese, kapitola 9.3; (in Czech)
D'Agostino, Ralph B.; Albert Belanger; Ralph B. D'Agostino, Jr (1990). "A suggestion for using powerful and informative tests of normality." The American Statistician. 44 (4): 316–321.

5. Testing the equality of variances in several independent samples (Petra Zahrádková)
- Bartlett test, Levene test
- explaining ideas of the tests
- discussion of advantages and disadvantages
- implementation in R – software

Literature:
Anděl J.: Statistické metody, Matfyzpress, Praha, 1998.
Brown, M. B., & Forsythe, A. B. (1974). Robust tests for the equality of variances. Journal of the American Statistical Association, 69(346), 364-367.
Hrochová, M. (2014). Leveneův test shodnosti rozptylů. Bachelor thesis. Faculty of mathematics and physics. Charles University.

6. Pearson's correlation coefficient (Denisa Dočekalová)
Pearson's correlation coefficient – population and sample version
Graphical illustration – what type of dependence is measured?
Tests of independence based on Pearson's correlation coefficient
Pearson correlation coefficient vs. slope in the simple linear regression model
Consider the following special situations:
* one of the variables is zero-one (i.e. it takes only values 0 and 1)
* both of the variables are zero-one (i.e. it takes only values 0 and 1)

Literature:
Kulich, M.: NMSA 331, Poznámky k přednášce,
http://msekce.karlin.mff.cuni.cz/~omelka/Soubory/nmsa331/ms1.pdf

7. Kernel density estimation (Andrej Uhliarik)
- what are we trying to estimate
- comparison with the histogram
- illustration of the different choice of bandwidths
- theoretical and practical approaches to the choice of the bandwidth
- implementation in R

Literature: Course notes to NMST434, Chapter 9, Available at
http://msekce.karlin.mff.cuni.cz/~omelka/Soubory/nmst434/nmst434_course-notes.pdf
Wand, M. P. and Jones, M. C. (1995). Kernel Smoothing. Chapman & Hall.

Silverman, B. W. (1986). Density estimation for statistics and data analysis. CHAPMAN/CRC.

8. Kernel regression estimation (Iakov Klyuchevskiy)
- introducing model; what are we trying to fit?
- explanation of the local polynomial estimation
- illustrative examples
- nearest-neighbour bandwidth choice
- LOWESS (Locally weighted scatterplot smoothing)

Literature: Course notes to NMST434, Chapter 10, Available at
http://msekce.karlin.mff.cuni.cz/~omelka/Soubory/nmst434/nmst434_course-notes.pdf
Fan, J. and Gijbels, I. (1996). Local Polynomial Modelling and Its Applications. Chapman & Hall/CRC, London.

9. Monte Carlo principle (Milan Šedivý)
The idea of Monte Carlo principle in general.
Monte Carlo principle and testing hypothesis. Interesting applications. Comparison with asymptotic inference.
Monte Carlo principle and confidence intervals. Specific examples.

Literature:
Course notes for NMST434, Chapter 5, Available at
http://msekce.karlin.mff.cuni.cz/~omelka/Soubory/nmst434/nmst434_course-notes.pdf

10. Permutation tests (Matěj Kovář)
Two sample permutation tests (comparision with t-test and Wilcoxon test)
K-sample permutation tests (comparision with ANOVA and Kruskall-Wallis test)
Test of independence

Literature:
Course notes for NMST434, Chapter 5, Available at
http://msekce.karlin.mff.cuni.cz/~omelka/Soubory/nmst434/nmst434_course-notes.pdf
Davison, A. C. and Hinkley, D. V. (1997). Bootstrap Methods and their Application. Cambridge University Press, New York. Chapter 4.3.
Efron, B. and Tibshirani, R. J. (1993). An Introduction to the Bootstrap. Chapman & Hall. Chapter 15.