Diagnostics of data

Lecturer: Karol Dziedziul (GUT); 15 h

All lectures take place at a computer lab according to following schedule.
A lecturer explains the problems. Using real data he shows how to solve problems applying software SAS.
Next the participants following the instruction try to do it by themselves. Finally they solved independently posed exercises.

Program includes code SAS mainly prepared by UCLA and from Agresti's book
http://www.ats.ucla.edu/stat/sas/examples/icda/default.htm

1. Categorical variables.
Binomial distribution, example: proportion in a society willing to protect the environment and accepting cuts in standard living,
Parameter estimation: MLE maximum likelihood estimation, MVUE minimum variance unbiased estimation. Testing hypothesis: significance test, Wald test, likelihood-ration test, small-sample influence using the ordinary p-value. Confidence intervals for binomial proportion (three methods) (2h)

2. Contingency tables.
Example: bread cancer and outcome of test(positive, negative). Joint distribution, margin distribution, conditional distribution, sensitivity and specificity in diagnostic tests.
Example: MI(myocardial infraction and treatment(placebo, aspirin). Difference of proportion, relative risk, odds ratio and log odds ratio, estimation of odds ratio.
Example: Party identification(democrat, independent, republican) by gender. Independence, test of independence(chi-square (Pearson's test), likelihood ration chi-square statistics,
residuals for cells in a contingency table.
Example: infant malformation and mother's alcohol consumption. Test of independency for ordinal data. (Ordinal and nominal variables).
Fisher's exact test. SAS procedure freq and genmod.(2h)

3. Association in three-way tables.
Simpson's paradox. conditional independence -- exercises. (2h)

Projekt „Rozwój interdyscyplinarnych studiów doktoranckich na Politechnice Gdańskiej w zakresie nowoczesnych technologii" jest współfinansowany przez Unię Europejską w ramach Europejskiego Funduszu Społecznego.

4. GLE: Generalized linear models.

Example snoring and heart disease. Linear probability model, logistic regression model, probit regression model.

Poisson distribution. (recalling some examples from lectures stochastic modeling). Example: number of crabs satellites by female's color, spine condition, width and weight.

SAS proc genmod, proc logistic. Figures are created using proc gplot. Exercises (3h)

5. Multicategory Logit Models.

Example alligator food choice. Log linear Models for Contingency Tables. Example alcohol, cigarette, marijuana use. SAS proc logistic, The option link=glogit . The option aggregate. (3h)

6. Continuous variables. Parametric and nonparametric tests: ANOVA, Kruskal Wallice. Remarks on Nonparametric estimation of density and regression function: Wavelets basis. Approximation. AMISE. Multivariate analysis: analysis of discrimination, data mining, principal components. SAS enterprise guide(3h)

TERMINY WYKŁADÓW			
Data	**Dzień tygodnia**	**Godzina**	**Sala**
25.02.2013	Poniedziałek	17-18/ 18-20	463 GG/ 13 GG (Lab.)
04.03.2013	Poniedziałek	17-18/ 18-20	463 GG/ 13 GG (Lab.)
11.03.2013	Poniedziałek	17-18/ 18-20	463 GG/ 13 GG (Lab.)
18.03.2013	Poniedziałek	17-18/ 18-20	463 GG/ 13 GG (Lab.)
25.03.2013	Poniedziałek	17-18/ 18-20	463 GG/ 13 GG (Lab.)

Projekt „Rozwój interdyscyplinarnych studiów doktoranckich na Politechnice Gdańskiej w zakresie nowoczesnych technologii" jest współfinansowany przez Unię Europejską w ramach Europejskiego Funduszu Społecznego.