Advanced Data Analysis – code 27084

8 September 2018

Course overview and course website

Relational databases

Single-table databases: features of a database table, fields and records.
Multiple-table databases: information redundancy, space waste, relations, relational databases.
Relations: one-to-many, many-to-one, one-to-one, many-to-many, junction table, primary key and ID, referential integrity, foreign key, foreign key connected to two tables. Temporal versus static databases.
Designing a new database: think at your problem, build a paper diagram, start from external tables, check that relations are properly oriented, add junction tables, fields names and types with options, description of the not obvious ones and non-standard decisions.

Microsoft Access

What is Access, Access' competitors.
MyFarm database example, Northwind database example, database file locking and security, Northwind database overview.
Automatic data save by Access, undo, non-automatic object save for Access, how to backup database.
Tables: datasheet view, design view, field types (text, memo, number, currency, date/time, yes/no, OLE object, index, hyperlink), field properties (size, format, decimal digits, required, allow zero length, validation rule, validation text), table validation rule, building Access tables, importing data from Excel files.
Relationships: relationships' diagram, create/modify/delete relations, enforcing referential integrity. Create relations using Lookup Wizard, create predetermined lists using Lookup Wizard and using values taken from another table.
Forms: what is a form, create using wizard, view/add/modify/delete records, subform, form locking.
Query: what is a select query, create using wizard, design view, no-show option, sorting, criteria, criteria in And and Or, adding tables, adding fields, creating virtual fields, Expression Builder (DATE, DATEDIFF, YEAR, LOG, ABS, SQR, EXP, INT, LIKE, BETWEEN). How to save objects inside database.
Summary query: group by option, count/sum/average/max/min options, using where option by for conditions.
Report: what is a report, create from tables or queries, create using wizard, grouping levels, sorting, export in RTF.
Non-selection queries: make table, delete, append, update. Right/left joins. Database documentation.

Cryptocurrencies and blockchain technology

Cryptography and document signing
Bitcoin transactions, public and private key, blockchain, wallet, transaction example, privacy issues, the double-spending problem, comparison with traditional transactions.
Nakamoto's article, block confirmation, proof-of-work, mining, competing chains, mining reward and total supply, fees, scaling problem
Ethereum, Monero, Bitcoin Cash, exchanges
Social advantages, smart contracts, initial coin offering ICO, personal advantages, other blockchain applications

Personal disadvantages, technical threats, social criticism.

Statistical analysis with R

Questionnaires. Variables: scale, nominal, ordinal, Likert scale. Missing values: NA, NaN.

R overview: portable version, installing packages, loading packages, R commander, saving script, output, workspace, loading workspace.

Descriptive statistics for one nominal variable, for one scale variable. Graphs for one nominal variable: column plot, pie chart, radar graph, bar plot, line plot, area plot, 3D. R: color palette, bar plot, pie chart. Graphs for one scale variable: histogram, box plot, plot case by case. R: histogram, box plot, index plot.

Descriptive statistics for two variables: contingency table, row and column percentage, statistics by groups, Pearson correlation, Spearman correlation. Graphs for two variables: clustered column plot, stacked column plot, 3D column plot, boxplots by groups, histograms by groups, scatterplot, mathematical graph. Three variables: surface plot, bubble chart, scatterplot by groups.

Restrict data set, drop unused factor levels, remove cases with missing values, binning, recode variables, massive recoding, compute new variables. Basic vector operations.

Statistical tests: sample and population, hypotheses, significance.

Student t test for one variable. Prerequisite and normality checking. Sign test. One-sided test versu two.sided test. Chi-square test for one dimensional contingency table. Chi-square test for a two-dimensional contingency table, Student's t test for two populations, prerequisite, Mann-Whitney test, ANOVA, Kruskal-Wallis test, correlations' tests, when testing difference of two scale variables, Student's t test for two paired variables, Wilcoxon signed-rank test.

Normality: histogram with normal curve, QQ-plot, skewness and excess Kurtosis, Shapiro Wilk test.