

## Vícerozměrné metody

### Analýza hlavních komponent

1. A census provide information, by tract, on five socioeconomic variables for the Madison, Wisconsin, area. The data from 61 tracts are available (variables: pop – total population [thousands], pd – professional degree [percent], ea – employed age over 16 [percent], ge – government employment [percent], mhw – median home value [\$100 000]). Determine principal components. [Dataset: socioecon.txt]
2. In study of size and shape relation for painted turtles, Joliccou and Mosimann measured carapace length, width, and height. Data suggest an analysis in terms of logarithms. Determine principal components. [Dataset: turtles.txt]
3. The weekly rates of return for five stocks (JPMorgan, Citibank, WellsFargo, RoyalDutchShell, Exxon-Mobil) listed on the New York Stock Exchange were determined for the period January 2004 through December 2005. The weekly rates of return are defined

$$\frac{\text{current week closing price} - \text{previous week closing price}}{\text{previous week closing price}},$$

adjusted for stock splits and dividends. The observations in 103 successive weeks appear to be independently distributed, but the rates of return across stock are correlated, because as one expects, stocks tend to move together in response to general economic conditions. Determine principal components.

[Dataset: stock\_rates.txt]

### Faktorová analýza

1. Lawley and Maxwell (Factor Analysis as a Statistical Method, 1971) present the sample correlation matrix of examination scores in  $p = 6$  subjects (Gaelic, English, History, Arithmetic, Algebra, Geometry) for  $n = 220$  male students. The correlation matrix is

$$R = \begin{pmatrix} 1.000 & 0.439 & 0.410 & 0.288 & 0.329 & 0.248 \\ 0.439 & 1.000 & 0.351 & 0.354 & 0.320 & 0.329 \\ 0.410 & 0.351 & 1.000 & 0.164 & 0.190 & 0.181 \\ 0.288 & 0.354 & 0.164 & 1.000 & 0.595 & 0.470 \\ 0.329 & 0.320 & 0.190 & 0.595 & 1.000 & 0.464 \\ 0.248 & 0.329 & 0.181 & 0.470 & 0.464 & 1.000 \end{pmatrix}$$

Find maximum likelihood solution of a factor model, determine a number of factors, communalities, specific variances, proportion of variance explained by each factor, draw a graph of estimated loadings.

[Dataset: exam\_scores.txt]

2. Hirschey and Wichern (Accounting and Market-Value Measures of Profitability: Consistency, Determinants and Uses, Journal of Business and Economic Statistics, 2(4),1984, 375–383) investigate the consistency, determinants and uses of accounting and market-value measures of profitability. As a part of their study, a factor analysis of accounting profit measures and market estimates of economic profits was conducted. The correlation matrix of accounting historical, accounting replacement, and market-value

measures of profitability for a sample of firms operating in 1977 is as

$$R = \begin{pmatrix} 1.000 & 0.738 & 0.731 & 0.828 & 0.681 & 0.712 & 0.625 & 0.604 \\ 0.738 & 1.000 & 0.520 & 0.688 & 0.831 & 0.543 & 0.322 & 0.303 \\ 0.731 & 0.520 & 1.000 & 0.652 & 0.513 & 0.826 & 0.579 & 0.617 \\ 0.828 & 0.688 & 0.652 & 1.000 & 0.887 & 0.867 & 0.639 & 0.563 \\ 0.681 & 0.831 & 0.513 & 0.887 & 1.000 & 0.692 & 0.419 & 0.352 \\ 0.712 & 0.543 & 0.826 & 0.867 & 0.692 & 1.000 & 0.608 & 0.610 \\ 0.625 & 0.322 & 0.579 & 0.639 & 0.419 & 0.608 & 1.000 & 0.937 \\ 0.604 & 0.303 & 0.617 & 0.563 & 0.352 & 0.610 & 0.937 & 1.000 \end{pmatrix}$$

Variables: Historical return on assets (HRA), historical return on equity (HRE), historical return on sales (HRS), replacement return on assets (RRA), replacement return on equity (RRE), replacement return on sales (RRS), market Q ratio (Q), Market relative excess value (REV). Perform a factor analysis, determine a number of factors, communalities, specific variances, proportion of variance explained by each factor, draw a graph of estimated loadings. [Dataset: profitability.txt]

- The dataset contains a hypothetical sample of 300 responses on 6 items from a survey of college students' favorite subject matter. The items range in value from 1 to 5, which represent a scale from Strongly Dislike to Strongly Like. Our 6 items asked students to rate their liking of different college subject matter areas, including biology (BIO), geology (GEO), chemistry (CHEM), algebra (ALG), calculus (CALC), and statistics (STAT). Perform a factor analysis, determine a number of factors, communalities, specific variances, proportion of variance explained by each factor, draw a graph of estimated loadings. [Dataset: subject\_rates.txt]