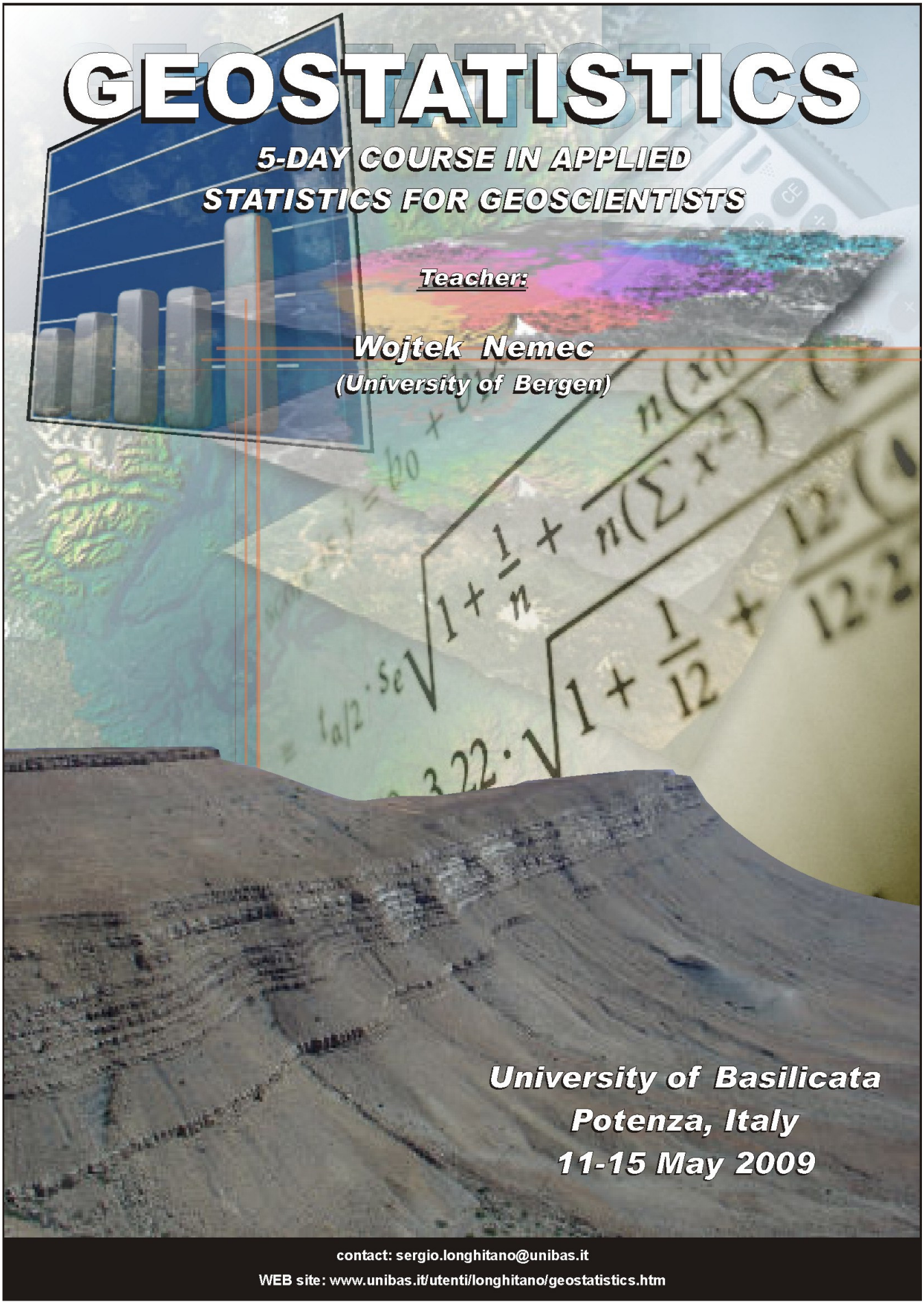
GEOSTATISTICS
5-DAY COURSE IN APPLIED
STATISTICS FOR GEOSCIENTISTS
Teacher:
Wojtek Nemec
(University of Bergen)
University of Basilicata
Potenza, Italy
11-15 May 2009
contact: sergio.longhitano@unibas.it
WEB site: www.unibas.it/utenti/longhitano/geostatistics.htm

Università degli Studi della Basilicata
Facoltà di Scienze MM. FF. e NN.
Dottorato di Ricerca in Scienze della Terra
Dipartimento di Scienze Geologiche
GeoSed – Associazione Italiana di Geologia del Sedimentario

G E O S T A T I S T I C S

5-day course in applied statistics for geoscientists

Aula 39 Edificio 3D
Campus universitario di Macchia Romana, V.le dell'Ateneo lucano 10, Potenza.

The Department of Geological Science of the University of Basilicata, in collaboration with the Italian Association of Sedimentary Geology (GeoSed), organizes an advanced course in **Geostatistics**, to be given by Professor **Wojtek Nemec** (University of Bergen, Norway). The course will take place in Potenza (Basilicata, Italy), from **May 11th** till **May 15th**, **2009**, in the building 3D of the University Campus, kindly made available by the Faculty of Science of the Basilicata University.

This is a methodological course in applied statistics for Master/PhD students and professional researchers in all fields of Earth Sciences. The programme consists of lectures and numerical exercises, with the presentation of each method followed directly by practical geological examples (which will possibly be selected to meet the personal research interests of the participants). No special prior knowledge in mathematics, statistics or informatics is required.

The course gives a practical introduction to the use of statistical methods for the analysis of quantitative and qualitative geological data – as a basis for drawing objective, justified and verifiable scientific conclusions. Both parametric and non-parametric statistical methods will be demonstrated, with reference to a wide range of geological problems. The emphasis will be on the various possible ways in which particular data-sets can be analysed, on the corresponding calculation procedures (using a pocket calculator, rather than any 'black-box' software) and on the geological interpretation of the numerical results.

The course is opened to a limited number (ca. 50) of participants. If you wish to attend the course, please fill in the attached **registration form** and return it by e-mail to the organizers at the address given below. (In the form, the questions about your field of research interest/activity and the type of data-sets you are dealing with in your research are meant to help the teacher to select possible relevant examples for the exercises.) The deadline for the registration is **March 15th**, **2009**. After this date, a second circular with further details (payment procedure etc.) and logistic information will be send to the registered participants.

The registration fee depends upon the participant's professional status and is set as follows:

- Oil company/industry researchers ..... € 500
- University researchers (permanent position) ..... € 250
- Master/PhD students and researchers in non-permanent position ..... € 150

The fee includes a copy of the lecture notes book and service for all coffee breaks. Accommodation can be booked in hotels near the University Campus in Potenza. The telephone numbers, e-mail addresses and room prices of the hotels will be given in the second circular.

Send your registration form by e-mail to Sergio Longhitano: sergio.longhitano@unibas.it

GEOSTATISTICS

5-day course in applied statistics for geoscientists

by Wojtek Nemec

LECTURE PROGRAMME

1. Introduction
- 1.1. Why do we need statistics in geology?
- 1.2. Geostatistics and the scope of the present course

2. Basic terminology
- 2.1. Variables – their types and scales
- 2.2. Sample population and general population; population size and dimension

3. What is a "representative" sample population?

4. Estimation of an optimal size of sample population

5. Frequency distribution
- 5.1. Non-grouped and grouped data; frequency plots
- 5.2. Some properties of normal distribution
- 5.3. Common distribution types
- 5.4. Exceedence frequency plots
- 5.5. The significance of power-law distribution (fractal statistics)
- 5.6. Strategy for using parametric and non-parametric methods

6. Basic parameters and their estimators (statistics)
- 6.1. Arithmetic mean, variance and standard deviation
- 6.2. Other useful basic statistics: median, mode, coefficient of skewness, geometric mean, harmonic mean, standard deviation of the mean and coefficient of variation
- 6.3. Simple solutions for simple problems: examples

7. Confidence intervals for basic parameters

8. Confidence interval for mean proportion

9. Analysis of univariate populations
- 9.1. Parametric comparison of two populations
  - 9.1.1. Snedecor F-test for the difference between two variances
  - 9.1.2. Student t-test for the difference between two means
  - 9.1.3. Student t-test for the conformity between population mean and assumed mean
  - 9.1.4. Chi-square test for the difference between population variance and assumed variance
  - 9.1.5. Confidence interval for the difference between two means
  - 9.1.6. The Tukey test for the difference between two means
  - 9.1.7. Confidence interval for the difference between two variances
  - 9.1.8. Confidence interval for the difference between two proportions
- 9.2. Normality tests
  - 9.2.1. The Cramer-von Mises test for non-grouped data
  - 9.2.2. The Pearson test for grouped data
- 9.3. Nonparametric comparison of two populations
  - 9.3.1. Runs test (U-test)
  - 9.3.2. The Mann-Whitney test (rank sum U-test)
  - 9.3.3. The Kolmogorov-Smirnov test (D-test) for grouped data
  - 9.3.4. Contingency tables with chi-square test for qualitative variable
- 9.4. Simultaneous comparison of several populations
  - 9.4.1. The Bartlett chi-square test for homogeneity of variances
  - 9.4.2. The Tukey-Cramer test for homogeneity of means
  - 9.4.3. One-way ANOVA as a test for homogeneity of means
  - 9.4.4. Two-way and three-way ANOVA
  - 9.4.5. The Kruskal-Wallis test (nonparametric rank sum H-test)

10. Analysis of vectorial data populations (*optional topic*)
- 10.1. The construction of rose diagram
- 10.2. Analysis of two-dimensional (azimuthal) vectorial populations
  - 10.2.1. The Kuiper test of distribution uniformity
  - 10.2.3. The Watson test of distribution uniformity
  - 10.2.4. Chi-square test of distribution uniformity for grouped data
  - 10.2.5. Parameters and statistics of population with circular-normal distribution: the mean vector direction and length, vectorial concentration, the Batschelet cricular variance and standard deviation
  - 10.2.6. The Rayleigh test of distribution uniformity (parametric)
  - 10.2.7. Evaluation of sample population and analytical strategy
  - 10.2.8. Confidence interval (sector) for the mean vector direction
- 10.3. Comparison of two vectorial populations

10.3.1. Test t for the difference between two mean vectors
10.3.2. Test F for the difference between two mean vectors
10.3.3. Test for the difference between vectorial concentrations
10.3.4. Nonparametric test for the difference between two vectorial populations
10.3.5. Simultaneous comparison of several vectorial populations
10.4. Optimization of sampling of a vectorial population
10.5. Analysis of three-dimensional (spherical) vectorial populations
10.5.1. Numerical representation of geological measurements
10.5.2. Chi-square test for spherical distribution uniformity
10.5.3. Parameters and statistics: mean vector direction and length, vectorial concentration
10.5.4. Confidence interval (cone) for the mean vector
10.6. Comparison of two spherical vectorial populations
10.6.1. Test for the differences between two mean vectors
10.6.2. Test for the difference between two vectorial concentrations

11. Analysis of bivariate populations
11.1. Relationship between two quantitative variables: linear correlation and regression
11.1.1. Covariance and Pearson's correlation coefficient
11.1.2. The Fisher test for the significance of linear correlation
11.1.3. The problem of data closure
11.1.4. Straight-line regression
11.1.5. The Fisher test for the significance of regression
11.1.6. The Fisher test for the difference between two regression coefficients
11.1.7. The regression line's goodness-of-fit
11.1.6. Standard error of estimation for a regression line
11.1.7. Confidence interval and confidence belt for regression line
11.1.8. Curvilinear regression: polynomial, exponential, logarithmic and power models
11.1.9. Nonparametric alternative for linear correlation and regression
11.2. Relationship between a quantitative and one qualitative variable
11.3. Relationship between two qualitative variables: association analysis
11.3.1. Association of two ordinal variables
11.3.2. Association of two nominal variables

12. Analysis of multivariate populations
12.1. Relationship between a quantitative variable ad several other quantitative variables: multicorrelation and multiregression
12.2. Classification of samples (R-mode analysis) or variables (Q-mode analysis)
12.2.1. Cluster analysis and dendrograms
12.2.2. Principal component (factor) analysis
12.3. Discriminatory analysis of two populations

13. Analysis of sequential (serial) data populations
13.1. Tests for randomness of data series
13.1.1. Runs test for binomial data series
13.1.2. Up-and-down runs test for binarized quantitative or ordinal data series
13.1.3. Kendall's turning-point test for binarized data series
13.1.4. Fisz's median-crossing test
13.1.5. Gold's length-of-runs test
13.1.6. Meacham's rank-difference test
13.1.7. Hurst's statistic (rescaled range analysis)
13.2. Tests for asymmetric trends in data series
13.2.1. Spearman's rank correlation test
13.2.2. Kendall's rank correlation test
13.2.3. The Moore-Wallis sign test
13.2.4. Mean-square successive difference test
13.2.5. The Wald-Bertram test for two equal-length data series
13.3. Statistical filtering of quantitative data series
13.4. Autocorrelation and cross-correlation methods
13.5. Autoassociation and cross-association methods
13.6. Semivariogram
13.7. Markov-chain analysis
13.7.1. Embedded and regular chain models
13.7.2. Test for the significance of Markovian difference matrix
13.7.3. Test for the significance of the individual elements of difference matrix
13.7.4. Test for stationarity of Markov chain
13.8. Fourier (spectral) analysis

14. Analysis of continuous 'regionalized' variables (data with map location coordinates)
14.1. Trend surface analysis
14.2. Other surface-fitting methods

GEOSTATISTICS

University of Basilicata, Potenza
11-15 May 2009

REGISTRATION FORM

Your first name:	
Your second name:	
Your position*:	

*PhD students and academic researchers in non-permanent position must attach a written statement about their status/position certified by the scientific advisor or the institute/department head.

Your affiliation (company/university):	
Postal address:	
You phone number:	
Your E-mail:	

Your field of research interest: (e.g., sedimentology, geochemistry, structural geology, palaeontology, etc.)	
What type(-s) of data-sets are you dealing with in your research? (i.e., what are you systematically measuring or counting or recording?)	

Registration fee	☐ € 500.00 (oil company/industrial researcher) ☐ € 250.00 (university researcher) ☐ € 150.00 (Master/PhD student or academic researcher in non-permanent position)

Please fill in, scan and return as a PDF file
to <sergio.longhitano@unibas.it> by 15 March 2009.