

# Statistika na webu

Prof. RNDr. Milan Meloun, DrSc.

Katedra analytické chemie, Univerzita Pardubice, 532 10 Pardubice

<http://meloun.upce.cz>,

*milan.meloun@upce.cz*



# Licenční studium

**UNIVERZITA PARDUBICE,  
Fakulta chemicko-technologická, Katedra analytické chemie  
Projekt dvouletého licenčního studia chemometrie na téma**

# Statistické zpracování dat

Vedoucí licenčního studia a odborný garant:

**Prof. RNDr. Milan Meloun, DrSc.**

Katedra analytické chemie, Fakulta chemicko-technologická,  
Univerzita Pardubice, nám. Čs. Legií 565, 53210 Pardubice,

**IČO: 216 275, DIČ: 248-00216275**

**Telefon: 466037026, Fax: 466037068**

**E-mail: milan.meloun@upce.cz, <http://meloun.upce.cz>**

**Datum otevření posledního 10. licenčního studia:**

**20. září 2004**

**Přehled základních informací k licenčnímu studiu:**

<u>Charakter studia</u>	<u>Organizace studia</u>	<u>Vyučující (foto)</u>	<u>Plán studia</u>	<u>Sylaby předmětů</u>	<u>Přihláška do adresáře</u>	<u>Absolventi, diplomky</u>
<u>Návody úloh</u>	<u>Otázky ke zkoušce</u>	<u>Vzory semestr. prací</u>	<u>Potřebná literatura</u>	<u>Potřebný software</u>	<u>Zákony Murphyho</u>	<u>Promoce licen. studia</u>

# Intenzivní kurzy

# **Statistické zpracování dat**

## **1. Inovovaný kurz počítačové analýzy dat**

Intenzivní výuka nepředpokládá předběžné znalosti a probíhá stylem "škola hrou" tj. po hodině teorie následuje vždy hodina praktických úloh na počítači.

**Kurz je s novou učebnicí Kompendium .....+ CD.!**

**Termíny:**

Každoročně vždy v měsících únoru, červnu a září (dotažte se na termín nejbližšího kurzu)

**Místo:**

Elektronická učebna LE1, UNIVERZITA PARDUBICE, nám. Čs. Legií 565, 532 10 Pardubice

**Organizátor:**

Katedra analytické chemie, Fakulta chemicko-technologická, Univerzita Pardubice, nám. Čs. Legií 565, 53210 Pardubice, **Telefon:** 466037026, **Fax:** 466037068, **E-mail:** milan.meloun@upce.cz

**Cena:**

**10.990,- Kč** zahrnuje Kompendium úloh za 495,- Kč, ale neobsahuje ubytování a stravu

1. **Základy metrologie:** klasifikace chyb měření, přesnost přístrojů, mezní hodnoty chyb, třída přesnosti přístroje, kvantilové a momentové chyby, chyby analytických výsledků, nejistoty výsledků, propagace chyb absolutních a relativních, propagování nejistot u analytických stanovení. Hromadění nejistot u instrumentálního stanovení.
2. **Interaktivní statistická analýza jednorozměrných dat:** diagnostiky exploratorní analýzy dat (grafy: krabicové, kvantilový, polosum, symetrie, Q-Q, rankitový, hustoty pravděpodobnosti, kruhový), ověření předpokladů o datech, transformace dat. Druhy rozdělení a určení rozdělení výběru. Základní statistiky polohy a rozptýlení při analýze výběru. Robustní statistiky ve stopové analýze. Hornův postup analýzy malých výběrů. Statistické testování, testy správnosti a testy přesnosti. Porovnávání výsledků s normou a validace analytických postupů. Párový test.
3. **Analýza rozptylu:** Základní pojmy. Jednofaktorová analýza rozptylu. Technika vícenásobného porovnávání. Dvoufaktorová analýza rozptylu vyvážená a nevyvážená.
4. **Regresní modely:** Lineární a nelineární regresní modely. Předpoklady metody nejmenších čtverců. Regresní diagnostika při výstavbě modelu. "Regresní triplet" (DATA, MODEL, METODA). Validizace analytické metody, porovnání s normou. Párový test dvou metod. Heteroskedasticita, autokorelace a multikolinearita v experimentálních závislostech. Věrohodnost stanoveného modelu. Výstavba vícerozměrného lineárního a nelineárního modelu. Analýza polynomů a odstranění multikolinearity.
5. **Kalibrace (lineární a nelineární):** Intervalový odhad neznámé koncentrace (obsahu). Limity přesnosti kalibrace: kritická úroveň, mez detekce a mez stanovení. Kritická hladina a šum metody. Určení kalibračního modelu pomocí regresní diagnostiky. Analýza vlivných bodů. Vliv přesnosti přístroje, metody a experimentu na limitu detekce.

# Hledání na internetu

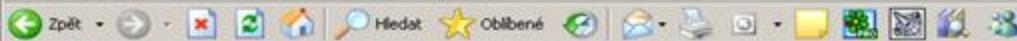
Knihovny dat

Knihovny řešených úloh

Manuály software

Demonstrační software

<http:// .....adresa na webu.....>



ISSN 1069-1089



An International Journal on the Teaching and Learning of Statistics

Advertise in JSE!  
advertise@amstat.org[Volume 12 \(2004\)](#)[Archive \(1993-2003\)](#)[Index](#)[Data Archive](#)[Information Service](#)[Editorial Board](#)[Authors](#)[Data Contributors](#)[Home Page](#)[Contact JSE](#)[ASA Publications](#)[Search JSE](#)

Go

## JSE Data Archive

[Datasets from \*The Basic Practice of Statistics\*, by David S. Moore](#)[Datasets from \*Applied Regression Analysis: A Research Tool\*, by John O. Rawlings](#)[4c.dat](#) (the basic data file)[4c1.dat](#) (includes indicator or "dummy" variables)[4c.txt](#) (the documentation file)

NAME: Pricing the C's of Diamond Stones

TYPE: Observational Regression Analysis Data

SIZE: 308 observations, 5 variables

The [article associated with this dataset](#) appears in the *Journal of Statistics Education*, Volume 9, Number 2 (July 2001).

SUBMITTED BY:

Singfat Chu

Faculty of Business Administration

National University of Singapore

10 Kent Ridge Crescent

Singapore 119260

fbachucl@nus.sg

[93cars.dat](#) (the basic data file)[93cars.txt](#) (the documentation file)

NAME: 1993 New Car Data

TYPE: Sample

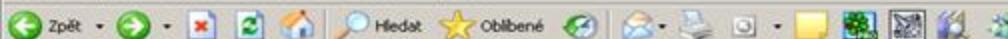
SIZE: 93 observations, 26 variables

The [article associated with this dataset](#) appears in the *Journal of Statistics Education*, Volume 1, Number 1 (July 1993).

SUBMITTED BY:

Robin H. Lock

Mathematics Department



# Statistics on the Web



*Links!Go*  
Probability and  
Statistics

©1994-2004 Clay Helberg. All rights reserved

This is a list of statistics resources I have discovered on the World Wide Web (WWW). I hope you will find them useful. If you know of (or maintain) a statistics resource on the Web that you don't see here, please [drop me a line](#) to let me know about it, and I'll take a look at it. Enjoy!

Note: this site, and my expertise, have to do with **statistical methods**. You won't find any information here about:

- Which web sites are most popular
- Demographic profiles of computer owners or net surfers
- How many web pages there are

Or other related questions. For answers, you might try [Media Metrix](#) or [DoubleClick](#). I also can't tell you how many ophthalmologists there are in Montana, or how many people visited Disneyland in 1997, or what the most popular brand of fish sticks is. For those, try [LEXIS-NEXIS](#) or [DIALOG](#). Also see Paula Berenstein's nice book, [Finding Statistics Online](#), which deals with such questions much better than I could.

- [Professional Organizations](#)
- [Institutes and Consulting Groups](#)
- [Educational Resources](#)
  - [Web courses](#)
  - [Online textbooks](#)
- [Publications and Publishers](#)
- [Statistics Book List](#)
- [Software-oriented Pages](#)
- Mailing Lists and Discussion Groups
  - [List of Statistics Lists](#) courtesy of Mike Fuller
  - [Statistical List Subscription Service](#) complements of [Stat/Transfer](#)
- [Other lists of links](#)
- [Statisticians and Other Statistical People](#)

Multivariate Data Analysis Software and References - Microsoft Internet Explorer

Soubor Úpravy Zobrazit Oblibené Nástroje Nápověda

Zpět Hledat Obliběný Přejít Odkazy

Adresa http://astro.u-strasbg.fr/~fmurtagh/mda-sw/

# Fionn Murtagh's Multivariate Data Analysis Software and Resources Page

## Contents:

1. MDA-J: Multivariate Data Analysis - Java
2. Multivariate Data Analysis Software as Java Applications
3. Multivariate Data Analysis Software in Fortran (and C)
4. Multivariate Data Analysis Software in S-Plus/R (and Fortran),  
including hierarchical clustering with weights on cases (in R and C)  
and including hierarchical Haar wavelet transform (in R) **New 2005-01-19 (Mac OS X information).**
5. Gaussian Mixture Modeling with Bayes Factors in C
6. Resources and Links, including JP Benzécri's Pascal code for Correspondence Analysis; Multidimensional Scaling; Point Pattern Matching; and Reading FITS Files in R.

Legal notice: all software here can be freely used and incorporated into any system, for any purpose. It is required that the origin and authorship of any code taken from here is acknowledged in code and other documentation.

## 1. MDA-J: Multivariate Data Analysis - Java

A new area has been set up for this code, which has its own address: see [www.correspondances.info](http://www.correspondances.info).

## 2. Multivariate Data Analysis Software as Java Applications

- [Doc:] [Documentation](#) on all programs which follow.
- [Sample data:] [Fisher iris data](#), iris.dat.  
Format: row and column numbers (integer), followed by sequence of matrix data values (floating), read row-wise. The output obtained is available in each case below.
- [PCA:] [Principal Components Analysis on correlations](#), PCAcorr.java,  
which requires the JAMA class library. (See below for a link to JAMA.)  
To run:
  - (i) Assume JAMA class library is in subdirectory Jama in current directory.
  - (ii) Compile: javac PCAcorr.javaGeneral note: if your classpath variable is not set, you can run "javac" or "java" as: "javac -classpath ." and "java -classpath .".
  - (iii) Get syntax to standard output: java PCAcorr
  - (iv) Run on iris.dat: java PCAcorr iris.dat > [pcaoutput.txt](#)
- Hierarchical clustering - background reading: A small text, but one with everything necessary on the most effective hierarchical clustering algorithms (i.e., using nearest neighbor chains, and/or reciprocal or mutual nearest neighbors) is F. Murtagh, *Multidimensional Clustering Algorithms*, Physica-Verlag, 1985. See in particular the table on page 68.  
[HCT-1 Hierarchical Clustering HCT.java](#)

Examples - Microsoft Internet Explorer

Soubor Úpravy Zobrazit Oblibené Nástroje Nápověda

Zpět Hledat Oblibené Náš návod

Adresa <http://www.uc.edu/sashtml/stat/chap30/sect49.htm> Přejít Odkazy

[Chapter Contents](#)  [Previous](#)  [Next](#)  [Top](#)

The GLM Procedure

## Examples

---

[Example 30.1: Balanced Data from Randomized Complete Block with Means Comparisons and Contrasts](#)

[Example 30.2: Regression with Mileage Data](#)

[Example 30.3: Unbalanced ANOVA for Two-Way Design with Interaction](#)

[Example 30.4: Analysis of Covariance](#)

[Example 30.5: Three-Way Analysis of Variance with Contrasts](#)

[Example 30.6: Multivariate Analysis of Variance](#)

[Example 30.7: Repeated Measures Analysis of Variance](#)

[Example 30.8: Mixed Model Analysis of Variance Using the RANDOM Statement](#)

[Example 30.9: Analyzing a Doubly-multivariate Repeated Measures Design](#)

[Example 30.10: Testing for Equal Group Variances](#)

[Example 30.11: Analysis of a Screening Design](#)

[Chapter Contents](#)  [Previous](#)  [Next](#)  [Top](#)

Copyright © 1999 by SAS Institute Inc., Cary, NC, USA. All rights reserved.

Internet



Licenční studium



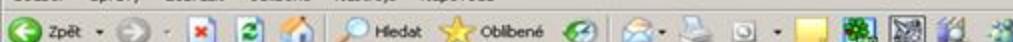
Literatura



Intenzivní kurzy



Nejlepší úlohy na internetu



## Stories by Method

[ANCOVA](#)  
[ANOVA](#)  
[Assumptions, regression](#)  
[Boxplot](#)  
[Causation](#)  
[Chi square test](#)  
[Cluster analysis](#)  
[Collinearity](#)  
[Confidence interval](#)  
[Contingency table](#)  
[Correlation](#)  
[Diagnostics](#)  
[Discriminant analysis](#)  
[Distribution](#)  
[Dotplot](#)  
[Dummy variable](#)  
[Efficiency](#)  
[Experimental design](#)  
[General linear model](#)  
[Histogram](#)  
[Independence](#)  
[Interaction](#)  
[Lurking variable](#)  
[Mann Whitney U test](#)  
[MANOVA](#)  
[Mean](#)  
[Median](#)  
[Multicollinearity](#)  
[Multivariate normality](#)  
[Multivariate regression](#)  
[Nested model](#)  
[Nonlinear regression](#)  
[Oregar, regression thru](#)  
[Outlier](#)  
[Paired t-test](#)  
[Partial regression plot](#)  
[Plot matrix](#)  
[Polynomial regression](#)  
[Pooled t-test](#)  
[Post Hoc test](#)  
[Principal components](#)  
[Probability](#)  
[Durbin-Watson](#)

# The Data and Story Library



DASL (pronounced "dazzle") is an online library of [datafiles](#) and [stories](#) that illustrate the use of basic statistics methods. We hope to provide data from a wide variety of topics so that statistics teachers can find real-world examples that will be interesting to their students. Use DASL's powerful search engine to locate the story or datafile of interest.

[Power Search](#)

[List all topics](#)

[List all methods](#)

[Data subjects](#)

[Help](#)

[Submit a story](#)

[Copyright](#)

## Overview

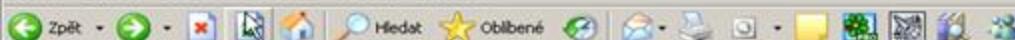
Teachers use examples to illustrate statistics concepts. A good example can make a lesson on a particular statistics method vivid and relevant. DASL is designed to help teachers locate and identify datafiles for teaching. We hope that DASL will also serve as an archive for datasets from the statistics literature.

The archive contains two types of files, stories and datafiles. Each story applies a particular statistical method to a set of data. Each datafile has one or more associated stories. The data can be downloaded as a space- or tab-delimited table of text, easily read by most statistics programs.

Stories are classified according to statistical methods and major topics of interest. Power search through DASL's stories and datafiles in five different ways.

1. **Title Search:** Searches through all of the story titles.
2. **Method Search:** Statistical methods such as regression or ANOVA.
3. **Topic Search:** Topics such as psychology or health.
4. **Datafile Subject Search:** Data subjects such as finance or astronomy.
5. **Full-text Search:** Searches through all of the stories and datafiles.

The first four specialized searches are slightly faster than the full-text search. Use these searches if you know what you want. The full-text search is helpful if you're interested in something more



[Book Contents](#) [Previous](#) [Next](#)

## SAS/STAT User's Guide

# Changes and Enhancements to SAS/STAT Software in V7 and V8

[Changes and Enhancements to SAS/STAT Software]

---

[Overview](#)

[Output Delivery System](#)

[ANOVA Procedure](#)

[BOXPLOT Procedure](#)

[CATMOD Procedure](#)

[CORRESP Procedure](#)

[FACTOR Procedure](#)

[FASTCLUS Procedure](#)

[FREQ Procedure](#)

[GENMOD Procedure](#)

[GLM Procedure](#)

[GLIMMIX Procedure](#)

[KDE Procedure](#)

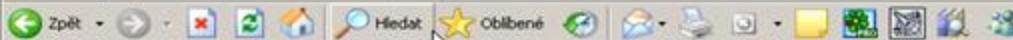
[KRIGE2D Procedure](#)

[LIFETEST Procedure](#)

[LOGISTIC Procedure](#)

[LOESS Procedure](#)

[MIXED Procedure](#)



THE UNIVERSITY of York

Department of Mathematics

Home For Students Admissions Staff Research Events Info

## PORTRAITS OF STATISTICIANS

[Click here for information about sources](#)[Click here for information about the life and work of statisticians](#)**A B C D E F G H I J K L M N O P Q R S T U V W X Y Z**Complementary information can be found at <http://members.aol.com/jayKplanr/images.htm>

1. [ABBE, Ernst](#) 1840-1905
2. [ADAMS, John Couch](#) 1819-1892
3. [ADRAN, Robert](#) 1775-1843
4. [AIRY, Sir George Biddell](#) 1801-1892
5. [ATTCHISON, John](#)
6. [AITKEN, Alexander Craig](#) 1895-1967
7. [ANDERSON, Oskar Nikolayevich](#) 1887-1960 = ANDERSON, Oskar Johann Viktor
8. [ANDERSON, Roy Malcolm](#)
9. [ANDERSON, Theodore Wilbur](#) 1915- with TUKEY, John Wilder
10. [ANDERSON, Virgil](#) 1915-
11. [ANSCOMBE, Frank J](#) 1918-2001
12. [ARBUTHNOT, John](#) 1667-1735
13. [ARMITAGE, Peter](#) 1924-
14. [ARNAULD, Antoine](#) 1560-1619

Statistics on the Web - Microsoft Internet Explorer

Soubor Úpravy Zobrazit Oblibené Nástroje Nápověda

Zpět Hledat Oblibené Přejít Odkazy

Adresa <http://my.execpc.com/~helberg/statistics.html#books>

## Statistics Textbooks on the web

- [Engineering Statistics Handbook](#) at NIST
- [Hypertext Intro Stat Textbook](#), edited by Jan de Leeuw
- [Introductory Statistics: Concepts, Models, and Applications](#), by David Stockburger
- [SurfStat Australia \(an intro-level resource\)](#), by the University of Newcastle (Australia) Statistics Department
- [HyperStat \(an intro-level resource\)](#), by David Lane
- [SticiGui](#), short for "Statistical Tools for Internet and Classroom Instruction with a Graphical User Interface" (whew!), by Philip B. Stark of UC Berkeley
- [Machine Learning, Neural and Statistical Classification](#) by Michie, Spiegelhalter, and Taylor.

---

## Publications

[Chapman & Hall](#)  
Chapman and Hall was recently acquired by [CRC Press](#). You can find information about Chapman & Hall books and journals at their site.

[Current Index to Statistics](#)  
This page gives information on how to order the main bibliographic index for statistics. Based at University of Chicago

[Duxbury Press](#)  
Home page of Duxbury Press, which specializes in industrial statistics and statistics education. Site features an [online catalog](#).

[Intelligent Data Analysis](#)  
This is a very nice online journal, published by [Elsevier Science](#).

[InterStat](#)  
An online statistics journal, published quarterly and offering abstracts and complete articles in PDF or PostScript format.

[Journal of Statistics Education](#)  
Contains articles from the Journal of Statistics Education (as you might expect), as well as EDSTAT-L (AKA newsgroup [sci.stat.edu](#)) and STAT-L (AKA newsgroup [sci.stat.consult](#)) archives. Also includes pointers to other discussion groups, software, etc. There is also a [Gopher server](#), which you can try if you have trouble getting through on the Web.

[Journal of Statistical Software](#)  
The UCLA Statistics department publishes this online journal. Content is self evident from the title.

[Marcel Dekker](#)  
Dekker publishes lots of fine technical books.

[McGraw-Hill](#)  
McGraw-Hill publishes numerous statistics textbooks, under their [Irwin](#) and [WCB](#) banners.

[Oxford University Press](#)  
Lots of good statistics books here. Note that the above link is to their North American web site. Their official UK website is [here](#).

[Sage Publications](#)  
This company publishes lots of quantitative books, including those handy little green books, the [Quantitative Applications in the Social Sciences](#) series.

[Springer-Verlag](#)  
This company specializes in scientific and technical publishing. [Samples](#) of some of their chapters are available in [Adobe Acrobat](#) format.

[John Wiley & Sons](#)  
Well known in the statistics community for their *Wiley Series in Probability and Mathematical Statistics*, they have a web page featuring information on their publications (including journals), as well as some online offerings. Information on specific titles is available from their [online catalog](#).

CHAPTER 2 - Microsoft Internet Explorer

Soubor Úpravy Zobrazit Obílené Nástroje Nápověda

Zpět Hledat Obílené Adresa <http://web1.iop.kcl.ac.uk/iop/Departments/BioComp/data%20sets.htm> Přejít Odkazy

## Data Sets Used in the Book

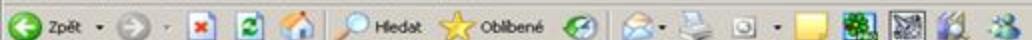
Load down all data sets as a zip-file: [data.zip](#)

For model answers click on the exercise number

Data Set	Table	Description	Book cross-reference	Exercise number
Data21.sav	2.1	Lifespan of rats	2.3.1 and 3.3.2	<a href="#">2.4.2</a>
Data22.sav	2.2	Ages at marriage in the U.S.	2.3.2 and 3.3.1	
Data23.sav	2.3	Guessing the width of a lecture hall		<a href="#">2.4.1</a>
Data24.sav	2.4	Motor vehicle theft in the U.S.		<a href="#">2.4.3</a>
Data25.sav	2.5	Anorexia Nervosa therapy		<a href="#">2.4.4</a>
Data31.sav	3.1	Belief in the afterlife by gender	3.3.3	
Data32.sav	3.2	Incidence of suicidal feelings in a sample of psychotic and neurotic patients	3.3.4	
Data33.sav	3.3	Oral contraceptive use and blood clots	3.3.5	
Data34.sav	3.4	Alcohol and infant malformation	3.3.6	
Data35.sav	3.5	Death penalty verdict by defendant's race and victim's race	3.3.7	<a href="#">9.4.3</a>
Data36.sav	3.6	Depersonalization and recovery from depression		<a href="#">3.4.1</a>
Data37.sav	3.7	Drug treatment of psychiatric patients		<a href="#">3.4.2</a>
Data38.sav	3.8	Tics and gender		<a href="#">3.4.3</a>
Data39.sav	3.9	Hair color and eye color		<a href="#">3.4.4</a>
Data41.sav	4.1	Cleaning cars	4.3.1	<a href="#">4.4.3 and 6.4.3</a>
Data42.sav	4.2	Temperatures in America	4.3.2	
Data43.sav	4.4	Air pollution in the U.S.		<a href="#">4.4.1 and 11.4.1</a>
Data44.sav	4.5	Body fat		<a href="#">4.4.2</a>
Data51.sav	5.1	Fecundity of fruit flies	5.3.1	
Data52.sav	5.2	Finger tapping and caffeine consumption	5.3.2	
Data53.sav	5.3	Female social skills	5.3.3	<a href="#">5.4.3</a>
Data54.sav	5.4	Cortisol levels in psychotics		<a href="#">5.4.1</a>
Data55.sav	5.5	Cycling and knee-joint angles		<a href="#">5.4.2</a>
Data57.sav	6.1	Effects of marijuana use	6.2.1	

Hotovo Internet

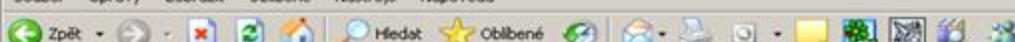
<http://web1.iop.kcl.ac.uk/iop/Departments/BioComp/data%20sets.htm>



## Datasets from the book: **A Handbook of Small Data Sets**

edited by D.J. Hand, F. Daly, A.D. Lunn, K.J. McConway and E. Ostrowski  
from Chapman and Hall, ISBN 0 412 39920 2.

1. [Germinating seeds](#)
2. [Guessing lengths](#)
3. [Darwin's cross-fertilized and self-fertilized plants](#)
4. [Intervals between cars on the M1 motorway](#)
5. [Tearing factor for paper](#)
6. [Abrasion loss](#)
7. [Mortality and water hardness](#)
8. [Tensile strength of cement](#)
9. [Weight gain in rats](#)
10. [Weight of chickens](#)
11. [Flicker frequency](#)
12. [Effect of ammonium chloride on yield](#)
13. [Comparing dishwashing detergents](#)
14. [Software system failures](#)
15. [Piston-ring failures](#)
16. [Strength of chemical pastes](#)
17. [Human age and fatness](#)
18. [Motion sickness](#)
19. [Plum root cuttings](#)
20. [Spectacle wearing and delinquency](#)
21. [Yield of isatin derivative](#)
22. [Fecundity of fruitflies](#)
23. [Butterfat](#)
24. [Snoring and heart disease](#)
25. [Trees' nearest neighbors](#)
26. [Air pollution in US cities](#)
27. [Acacia ants](#)
28. [Vaccination](#)
29. [Peppers in glasshouses](#)
30. [A clinical trial in lymphoma](#)
31. [Danish do-it-yourself](#)
32. [Testing cement](#)
33. [Irises](#)
34. [Water voles](#)
35. [Facilities in East Jerusalem](#)
36. [Yields of winter wheat](#)
37. [WISC blocks](#)
38. [Byssinosis](#)

NIST  
SEMATECH**HANDBOOK CHAPTERS**

- 1. Explore
- 2. Measure
- 3. Characterize
- 4. Model
- 5. Improve
- 6. Monitor
- 7. Compare
- 8. Reliability

**HOW TO USE HANDBOOK****TOOLS & AIDS****SEARCH HANDBOOK****DETAILED CONTENTS****ACKNOWLEDGMENTS**

# ENGINEERING STATISTICS

**H A N D B O O K**

**Welcome!** The goal of this handbook is to help scientists and engineers incorporate statistical methods in their work as efficiently as possible.

To reference the Handbook please use a citation of the form:

*NIST/SEMATECH e-Handbook of Statistical Methods*, <http://www.itl.nist.gov/div898/handbook/>, date.  
(Links to specific pages can also be referenced this way, if suitable.)

Printer friendly versions of each chapter in the e-Handbook can be found [here](#).

Information on Getting the e-Handbook on CD can be found [here](#).

Feedback on the Handbook sent to [handbook@nist.gov](mailto:handbook@nist.gov) is also much appreciated.

Soubor Úpravy Zobrazit Obíbene Nástroje Nápověda

Zpět Hledat Oblíbené Přejít Odkazy

Adresa <http://solon.cma.univie.ac.at/~neum/statdat.html>

# Statistical Data Sets

[UCI Machine Learning Repository](#)  
A very extensive archive with over hundred data collections from applications; get the README file ([local copy](#)) first

[UCI Knowledge Discovery in Databases Archive](#) for large data sets  
"The primary role of this repository is to enable researchers in knowledge discovery and data mining to scale existing and future data analysis algorithms to very large and complex data sets."

[StatLog datasets](#) from [Machine Learning, Neural and Statistical Classification](#) (online copy of the book by Michie, Spiegelhalter and Taylor)

[Natural Stimuli Collection](#) (van Hateren natural image database);

[Image Understanding Datasets](#) (includes among others stereo images, motion sequences)

[MeasTex Image Texture Database](#)

[Computer Vision Test Images](#)

[The Yale Face Database B](#)

[Benchmarking of Learning Algorithms](#) (also includes papers on methodology and assessment of the status quo)

[Unipen handwriting data](#) (not free)

[Peterson-Barney vowel data](#) (file [PetersonBarney.tar.Z](#))

[Audio File Format FAQ](#)

[Switchboard Transcription System](#), several hundred informal speech dialogs recorded over the telephone

[Speech databases](#) from the comp.speech FAQ

[Speech data](#)

[More speech data](#) from the Center for Spoken Language Understanding (\$30 per corpus)

[Linguistic Data Consortium LDC Catalog](#) of 210 corpora of language data (the cheapest, TIMIT and CTIMIT cost \$100 each)

[EconData](#), economic time series

[Time Series Data Library](#) (very extensive)

[University of Colorado at Boulder Time Series Repository](#)  
(containing a.o. the Santa Fe Institute Time Series Competition Data)

StatLib---Datasets Archive - Microsoft Internet Explorer

Soubor Úpravy Zobrazit Obíbene Nástroje Nápověda

Zpět Hledat Oblíbené Přejít Odkazy

Adresa <http://lib.stat.cmu.edu/datasets/>

# StatLib---Datasets Archive

If you have an interesting dataset, or collection of data from a book, please consider submitting the data.

To submit a dataset, please see the submissions guidelines, via

[send submissions from general](#)

Some of the entries are shar archives. If you don't know how to deal with a shar archive, send the message

[send shar from general](#)

for instructions.

---

## The datasets archive currently contains:

[NIST Statistical Reference Datasets \(StRD\)](#)  
A pointer to a NIST site that contains reference datasets for the objective evaluation of the computational accuracy of statistical software. Both users and developers of statistical software can use these datasets to ensure and improve software accuracy.

[agresti](#)  
Contains data from "An Introduction to Categorical Data Analysis," by Alan Agresti, John Wiley, 1996, plus SAS code for various analyses. (aa@stat.ufl.edu) [28/Feb/96] (12k)

[alr](#)  
This file contains data from Applied Linear Regression, 2nd Edition, by Sanford Weisberg, John Wiley, 1985 (sandy@umnstat.stat.umn.edu) (36808 bytes)

[analcatdata](#)  
A collection of the data sets used in the book "Analyzing Categorical Data," by Jeffrey S. Simonoff, Springer-Verlag, New York, 2003. Submitted by Jeff Simonoff (jsimonof@stern.nyu.edu). [6/Jul/03] (1.2M)

[Andrews](#)  
This data for the book DATA by Andrews and Herzberg. Available by FTP, gopher, WWW, but not e-mail.

[Arsenic](#)  
This datafile contains measurements of drinking water and toenail levels of arsenic, as well as related covariates, for 21 individuals with private wells in N.H. Source: Karagas MR, Morris JS, Weiss JE, Spate V, Baskett C, Greenberg ER. Toenail Samples as an Indicator of Drinking Water Arsenic Exposure. Cancer Epidemiology, Biomarkers and Prevention 1996;5:849-852. (Therese.A.Stukel\_AT\_Dartmouth.EDU) (MS Word format) [21/Jul/98] (5 kbytes)

[arsenic.zip](#)  
This is a zip file containing the arsenic data from Southwestern Taiwan that was used in the analysis reported by the National Academy of Sciences reports on arsenic (NRC 1999 and 2001), analyzed by Morales et al. (2000) and also discussed by Ryan (2004). See the following for a good summary, plus links to the NAS reports: <http://www4.nas.edu/news.nsf/isbn/0309076293?OpenDocument> Submitted by Louise Ryan (lryan@hsph.harvard.edu). [19/Dec/03] (35 kbytes)

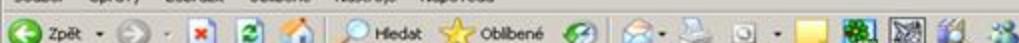
[backache](#)  
This file contains the 'backache in pregnancy' data analysed in Exercise D.2 of Problem-Solving: A Statistician's Guide, 2nd edn., by C. Chatfield, Chapman and Hall, 1995. (cc@maths.bath.ac.uk) [2/Oct/95] (16 kbytes)

[balloon](#)  
A data set consisting of 2001 observations of radiation, taken from a balloon. The data contain a trend and outliers. Source: Laurie Davies (mata00@de0hrz1a.BITNET) (43k) [5/Feb/93]

[bankresearch](#)  
A pre-classified dataset containing 11,000 web pages from 11 different categories. Although this dataset was designed for unsupervised clustering experiments it can be used for any type web page machine learning techniques. For more information see Paul-Soren Distant-Diver Subreddits by matalab00@de0hrz1a.BITNET [11/Feb/02] (64k)

Hotovo Internet

# Literatura



## StatLib---Andrews & Herzberg Archive

This area contains files of data for the book DATA by Andrews and Herzberg. The following is a list of the files available. Each entry contains a description of the tables in the book and the first and last 4 lines of the data.

Table 1.1 Sepal Length and Width and Petal Length and Width in Centimetres of Iris setosa, Iris versicolor and Iris virginica (3200 bytes)

```

1   1      1    5.1 3.5 1.4 0.2 7.0 3.2 4.7 1.4 6.3 3.3 6.0 2.5
1   1      2    4.9 3.0 1.4 0.2 6.4 3.2 4.5 1.5 5.8 2.7 5.1 1.9
1   1      3    4.7 3.2 1.3 0.2 6.9 3.1 4.9 1.5 7.1 3.0 5.9 2.1
1   1      4    4.6 3.1 1.5 0.2 5.5 2.3 4.0 1.3 6.3 2.9 5.6 1.8
1   1     47    5.1 3.8 1.6 0.2 5.7 2.9 4.2 1.3 6.3 2.5 5.0 1.9
1   1     48    4.6 3.2 1.4 0.2 6.2 2.9 4.3 1.3 6.5 3.0 5.2 2.0
1   1     49    5.3 3.7 1.5 0.2 5.1 2.5 3.0 1.1 6.2 3.4 5.4 2.3
1   1     50    5.0 3.3 1.4 0.2 5.7 2.8 4.1 1.3 5.9 3.0 5.1 1.8

```

Table 2.1 Heights of Zea Mays (424 bytes)

```

2   1      1    1 23.5 17.375
2   1      2    1 12 20.375
2   1      3    1 21 20
2   1      4    2 22 20
2   1     12    4 21 18
2   1     13    4 22.125 12.75
2   1     14    4 23 15.5
2   1     15    4 12 18

```

Table 3.1 Annual Number of Lynx Trappings in the MacKenzie River District for the Period 1821 to 1934 (1678 bytes)

```

3   1      1    1821 269 1859 684 1897 587
3   1      2    1822 321 1860 299 1898 105
3   1      3    1823 585 1861 236 1899 153
3   1      4    1824 871 1862 245 1900 387
3   1     35    1855 1638 1893 377 1931 1000
3   1     36    1856 2725 1894 1292 1932 1590
3   1     37    1857 2871 1895 4031 1933 2657
3   1     38    1858 2119 1896 3495 1934 3396

```

Table 3.2 Number of Lynx Pelts and Unit Price Paid to the Hudson's Bay Company for the Years 1857-1911 (1303 bytes)

```

3   2      1    1821 269 1859 684 1897 587
3   2      2    1822 321 1860 299 1898 105
3   2      3    1823 585 1861 236 1899 153
3   2      4    1824 871 1862 245 1900 387
3   2     63    1881 15386 12/7 1909 9704 87/4
3   2     64    1882 9443 14/8 1910 3410 123/1
3   2     65    1883 7599 16/9 1911 3774 102/4
3   2     66    1884 8061 19/2

```

## Datasets

All public domain datasets listed below are public dataset. If you click at the appropriate entry, this will take you to a subdirectory from where you can do the download. This subdirectory contains more information (in a .doc file) with a brief description of the dataset and previous test results on this dataset. Some larger datasets have been split into train and test set (as used in the Statlog project).

- [Australian \(Australian credit\)](#)
- [Diabetes \(Diabetes of pima-indians\)](#)
- [Dna \(DNA sequence\)](#)
- [German \(German credit\)](#)
- [Heart \(Heart disease \)](#)
- [Letter \(Letter recognition\)](#)
- [Segment \(Image segmentation\)](#)
- [Shuttle \(Shuttle control\)](#)
- [Satimage \(Landsat satellite image\)](#)
- [Vehicle \(Vehicle recognition using silhouettes\)](#)

Some datasets came originally from the *UCI Repository of Machine Learning Databases and Domain Theories* which is managed by D.W.Aha. The datasets were often processed and the repository mentioned above contains both the unprocessed and processed versions. The datasets available from here are *processed* datasets.

The following datasets were used in StatLog, but are *not public* (hence *not available* for download):

- |                          |                            |
|--------------------------|----------------------------|
| - Credit Management,     | - Handwritten digits,      |
| - Karhunen-Loeve digits, | - Chromosome,              |
| - Cut,                   | - Head injury,             |
| - Technical,             | - Belgian power,           |
| - Machine Faults,        | - Tsetse fly distribution. |

Maintained by P.Brazdil  
Last update March 1999.



**Děkuji za pozornost!**

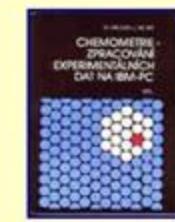
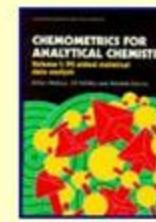
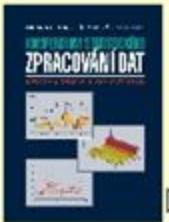
<http://meloun.upce.cz>

## CHEMOMETRIE/CHEMOMETRICS

### A. Monografie - Počítačové metody ke statistickému zpracování experimentálních dat, (Computer-Assisted Methods to the Statistical Treatment of Experimental Data)

Má se čist mnoho, nikoliv však mnohé.

Plinius: Epistulae VII, 9 (15)



Academia

1. M. Meloun, J. Militký : **KOMPENDIUM STATISTICKÉHO ZPRACOVÁNÍ DAT**, Metody a řešené úlohy včetně CD, Academia Praha 2002, 766 stran, ISBN 80-200-1008-4.(zde je celá kniha dostupná jako hypertext v ADOBE ACROBAT READER).
2. M. Meloun, J. Militký : **STATISTICKÉ ZPRACOVÁNÍ EXPERIMENTÁLNÍCH DAT** - v chemometrii, biometrii, ekonometrii a v dalších oborech přírodních, technických a společenských věd, EAST PUBLISHING Praha, 1998, 840 stran, ISBN 80-7219-003-2.
3. M. Meloun, J. Militký : **STATISTICKÉ ZPRACOVÁNÍ EXPERIMENTÁLNÍCH DAT** - v chemometrii, biometrii, ekonometrii a v dalších oborech přírodních, technických a společenských věd, PLUS Praha 1994, 1995 (dotisk), 840 stran, ISBN 80-85297-56-6.
4. M. Meloun, J. Militký : **STATISTICKÉ ZPRACOVÁNÍ EXPERIMENTÁLNÍCH DAT - Sbírka úloh**, Univerzita Pardubice 1996 (1. vydání 1000 ks), 308 stran, ISBN 80-7194-075-5.
5. M. Meloun, J. Militký , M. Forina: **CHEMOMETRICS FOR ANALYTICAL CHEMISTRY: Volume 1: PC-Aided Statistical Data Analysis**. Ellis Horwood, Chichester 1992, 330 stran, ISBN 0-13-126376-5.
6. M. Meloun, J. Militký , M. Forina: **CHEMOMETRICS FOR ANALYTICAL CHEMISTRY. Volume 2: PC-Aided Regression and Related Methods**, Ellis Horwood, Chichester 1994, ISBN 0-13-123788-7.
7. M. Meloun, J. Militký : **CHEMOMETRIE - Zpracování experimentálních dat na IBM PC**, SNTL, Praha 1991, 384 stran, ISBN 80-03-00462-4.



8. M. Meloun, M. Javůrek: **CHEMOMETRIE - I. díl Aplikace počítačů v analytické a fyzikální chemii**, (skriptum pro základní studium), Vys. Šk. Chemickotechnol., Pardubice 1986, 290 stran, číslo vydání 55-706-86.
9. M. Meloun, J. Militký , M. Javůrek: **CHEMOMETRIE - II. díl Aplikace počítačů v analytické a fyzikální chemii**, (skriptum pro postgraduální studium), Vys. Šk. Chemickotechnol., Pardubice 1988, 322 stran
10. M. Meloun, J. Militký , M. Javůrek: **CHEMOMETRIE - II. díl Aplikace počítačů v analytické a fyzikální chemii**, (skriptum pro postgraduální studium), Vys. Šk. Chemickotechnol., Pardubice 1991, 322 stran.



# Statistická analýza experimentálních dat

Milan Meloun  
Jiří Militký

ACADEMIA

Nová vysokoškolská učebnice statistického zpracování dat obsahuje celkem 980 stran a je proti předešlým vydáním rozšířena o nových 350 stran analýzy vícerozměrných dat.

Součástí učebnice je CD s textem jednak doprovodné sbírky 900 úloh s manuály metod ve formátu PDF a daty od stejných autorů zvané *Kompendium statistického zpracování dat*, Academia 2002 a jednak statistický software QC-Expert (na 30 dní) a konečně také programy v systému Matlab s daty a a podrobným manuálem.

Čtenáři se dostává do rukou ojedinělá publikace, v níž se obráží současná doba charakteristická tím, že vznikají, archivují se a zpracovávají velké objemy dat. Ukazuje, že k získávání informaci z takovýchto databází je nezbytné volit statistické metody a postupy založené na zkoumání statistických zvláštností dat a adaptivní tvorbě datově orientovaných modelů. Aplikace statistické analýzy dat na rozsáhlé databáze vedla v minulosti ke vzniku nových oborů jako chemometrie, biometrie, psychometrie, ekonometrie, technometrie aj. Pozornost autorů je nejprve zaměřena na teorii chyb instrumentálních měření i s analýzou nejistoty výsledků měření, a dále na vyšetřování statistických zvláštností dat průzkumovou analýzou. Potom autoři podrobně popisují vícerozměrné statistické metody (analýzu faktorovou, diskriminační, shlukovou a klasifikaci dat) vhodné nejen pro zpracování výsledků měření, ale zejména pro realizaci dobývání znalostí postupy použitelnými i pro neexperimentální data z průmyslu, obchodu, bankovnictví, státní správy atd. V další části jsou vysvětleny lineární a nelineární regresní modely, analýza rozptylu, korelace a korelační modely, a též interpolace a approximace. Jádrem výkladu látky je vždy počítacem podporovaný rozbor příkladu je zde dokumentováno více než 400 řešených příkladů a úloh z praxe laboratoří technické kontroly kvality a z ostatních technických laboratoří průmyslu. Kniha je určena studentům vysokých škol technického, přírodovědného, ale i humanitního a ekonomického směru, dále pracovníkům výzkumných ústavů, a také podnikových a kontrolních laboratoří a též zkušeben, jež se zabývají analýzou dat.

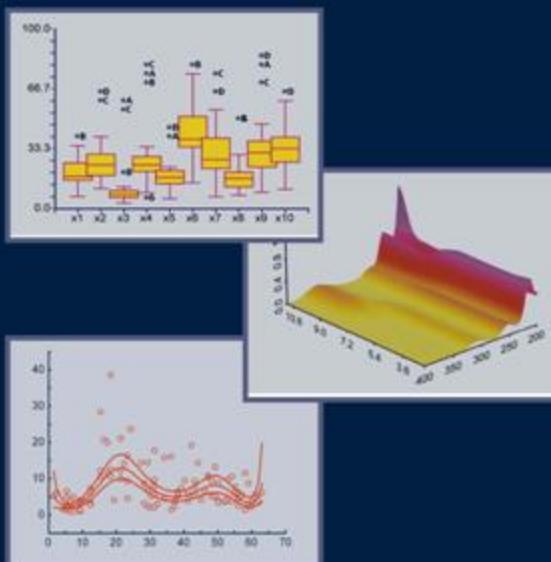
Knihu vydalo nakladatelství ACADEMIA Praha, prosinec 2004 s doporučenou cenou 690,- Kč.

**CD přiloženo.**

MILAN MELOUN • JIŘÍ MILITKÝ • ACADEMIA

# KOMPENDIUM STATISTICKÉHO ZPRACOVÁNÍ DAT

METODY A ŘEŠENÉ ÚLOHY VČETNĚ CD



**Kniha:** je učebnicí i sbírkou úloh s CD, 10 kapitol, úplná statistická metodologie a počítačové postupy, 1000 řešených úloh, desítky vzorových příkladů s výkladem. Vhodné pro studenty i pro inženýry a manažery jakosti, pro OHS, KHS, OKŘJ, pro analytické, lékařské, biochemické, hutnické, potravinářské, zemědělské laboratoře.

**CD obsahuje:** statistický software ADSTAT 2.5 (Windows) na 30 dní gratis, hypertextovou formu celé knihy, manuály k software, Adobe Acrobat gratis, databáze ke všem 1000 úlohám ve formátech \*.xls, \*.txt, \*.vts.



5  
0 10 20 30 40 50 60 70

Vyšlo v září 2005

**Milan Meloun**

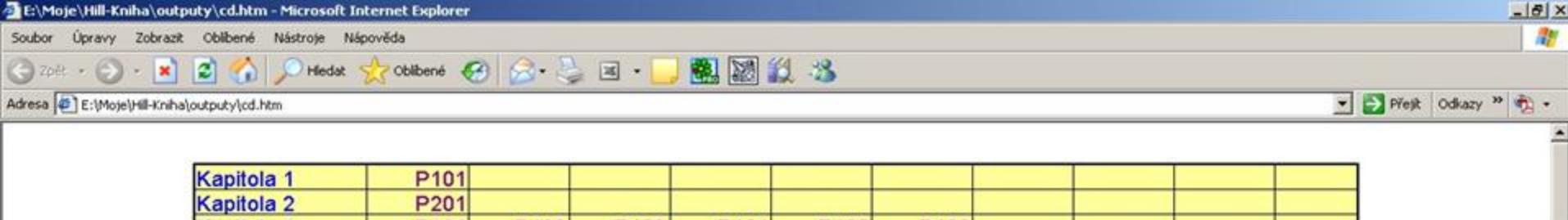
**Jiří Militký**

**Martin Hill**

# **Počítačová analýza vícerozměrných dat v příkladech**

**v oborech přírodních, technických  
i společenských věd**

**Učebnice s operativním CD:  
Každý příklad vyřešen 5 programy  
a outputy jsou k dispozici v \*.PDF formátu.**



Kapitola 1	P101									
Kapitola 2	P201									
Kapitola 4	P401	P402	P403	P404	P405	P406				
Kapitola 5	P501	P502	P503							
Kapitola 6	P601	P602	P603							
Kapitola 7	P701	P702	P703							
Kapitola 8	P801	P802	P803	P804	P805	P806	P807	P808		
Kapitola 9	P901	P902	P903	P904	P905	P906	P907	P908	P909	P910
	P911	P912	P913	P914	P915	P916	P917	P918		
Kapitola 10	P1001	P1002	P1003	P1004	P1005	P1006	P1007			
Kapitola 11	P1101	P1102	P1103	P1104	P1105	P1106	P1107	P1108	P1109	

Kliknutím na číslo příkladu se objeví output programu ....

PŘÍKLAD 1.1 Popisné statistiky jednorozměrné analýzy zdrojové matice dat Hrách				
	STATISTICA	NCSS2000	STATGRAPHICS	MATLAB
	Data	Data	Data	Data
	PCA	PCA	PCA	PCA
	CORR		FA	FA
	CLU	CLU	CLU	CLU
	FDA		DA	DA
			MDS	MDS
				nahoru
PŘÍKLAD 2.1 Grafy různých forem škálování, Skaly				
	STATISTICA	NCSS2000	STATGRAPHICS	MATLAB
	Data	Data	Data	Data
	PCA	PCA	PCA	PCA