

Softwarové řešení agnostiky

Ing. Lubomír Pavliska
ZÚ Ostrava



Gnostický SW

- Předpoklady realizace gnostických funkcí
- Zdrojové a vývojové prostředí
- Náklady
- Alternativní řešení
- Příklad použití
- Závěr

Původní realizace

- Nutnost použití masivního statistického SW
- S-PLUS® 6.2 Insightful Corporation
- Statistický programovací jazyk
- Dostatečné množství elementárních funkcí
- Podpora tvorby grafického rozhraní
- Současná verze
- Velmi vysoké náklady na provoz a vývoj

Podmínky alternativního řešení

- Výkonný statistický jazyk
- Vývojové prostředí
- Zajištění dalšího vývoje
- Tvorba uživatelských aplikací
- Dostupnost



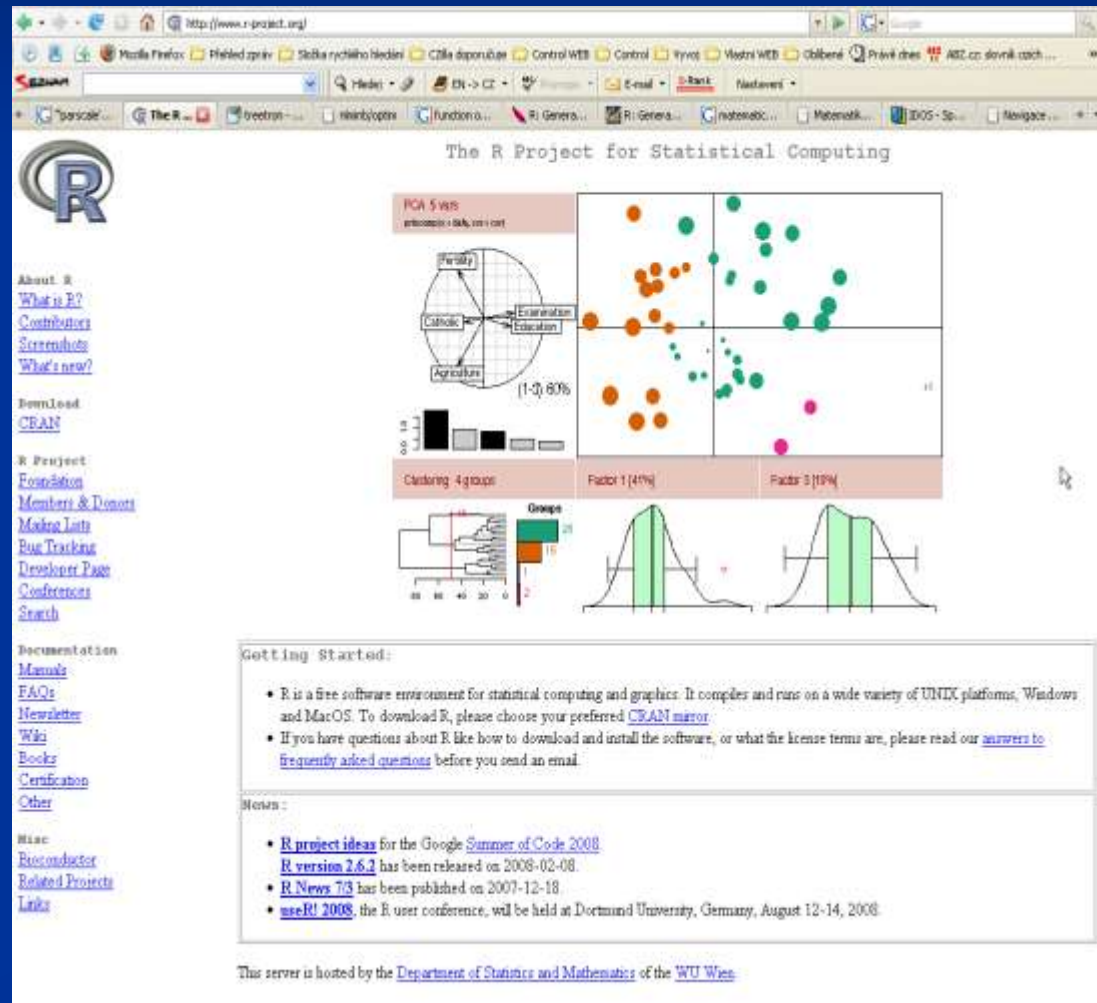
R-project

- Licence GNU (free-ware)
- R jazyk
 - Programovací jazyk pro statistickou analýzu
 - Interpret
 - Vývojové a běhové prostředí
 - Množství podpůrných knihoven
 - Platformová nezávislost (Windows, Mac OS, Linux)
 - Kvalitní vývoj (verze s periodou 1/4 roku)
 - Podpora grafického výstupu
 - Podpora tvorby grafického uživatelského prostředí
 - Dostupnost informací a dokumentace

Prostředí R

■ R „GNU S“

<http://www.r-project.org/>



The R Project for Statistical Computing

PCA 5 vars
articulos + dia, emi cat

Family
Catholic
Agriculture
Emigration
Education
(1-2) 80%

Clustering: 4 groups
Groups
11
2

Factor 1 (41%)
Factor 3 (19%)

Getting Started:

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To download R, please choose your preferred [CRAN mirror](#).
- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News:

- [R project ideas](#) for the Google Summer of Code 2008
- [R version 2.6.2](#) has been released on 2008-02-08.
- [R News 7/3](#) has been published on 2007-12-18.
- [useR! 2008](#), the R user conference, will be held at Dortmund University, Germany, August 12-14, 2008.

This server is hosted by the [Department of Statistics and Mathematics](#) of the [WU Wien](#).

Kompatibilita

- S-PLUS a R mají společný základ (S jazyk)
- R je Open Source implementace S jazyka
- Kompatibilita kódu s S-plus
- Rozdíly dané odlišným vývojem
- Přenesení gnostických funkcí z S-PLUS
- Rozdíly v knihovnách elementárních funkcí
- Výkonnost
- Běhové a spouštěcí podmínky

Studie

- **Marginální (jednorozměrná) gnostická analýza**
 - R je schopno nahradit prostředí S-PLUS v potřebném rozsahu
 - Tvorba aplikace
 - Vstupy a výstupy dat
 - Grafické uživatelské prostředí

Aplikace GNDF

- GNDF gnostická distribuční funkce



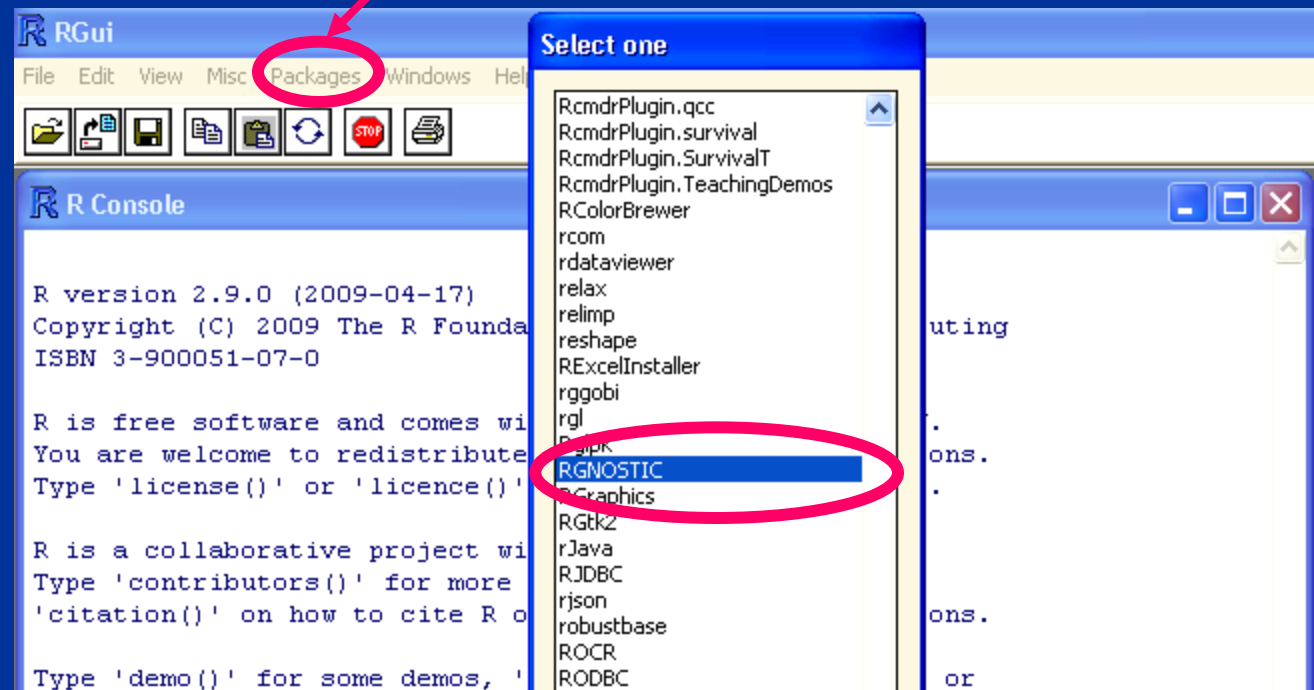
Vstupní data

- Typy souborů
 - Textové
 - Excel
- Struktury R
 - Nutnost ošetřit nedefinované hodnoty
 - Data frame
 - Numerické matice

Prostředí R

■ RGNOSTIC

Load packages



Prostředí R

- R konsola

- Příkazový řádek

- `Recogn(Sample)`

- `defDX <- GNDF(Sample)`

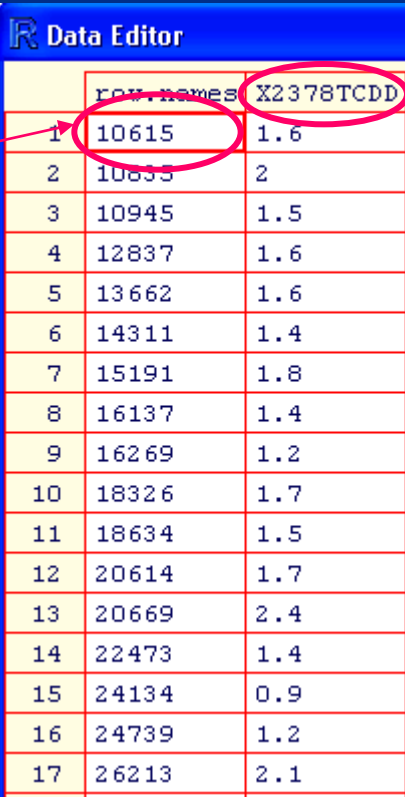
- `defDX <- GNDF(Sample, Aprweights, Censoring, Bcensup, Additive = T, Type = "E", Kind= "G", Fast=T, Lowbound, Upbound, HardLD = F, HardUD = F, Scaleparm, Scstart, Resolution = 400, Compress = T, varS = F, Graph = T, Filter = T)`

- `Hdf <- homogenizeE(defDX)`

- Export výstupů a grafy

Vzorek dat

■ Vstupní data



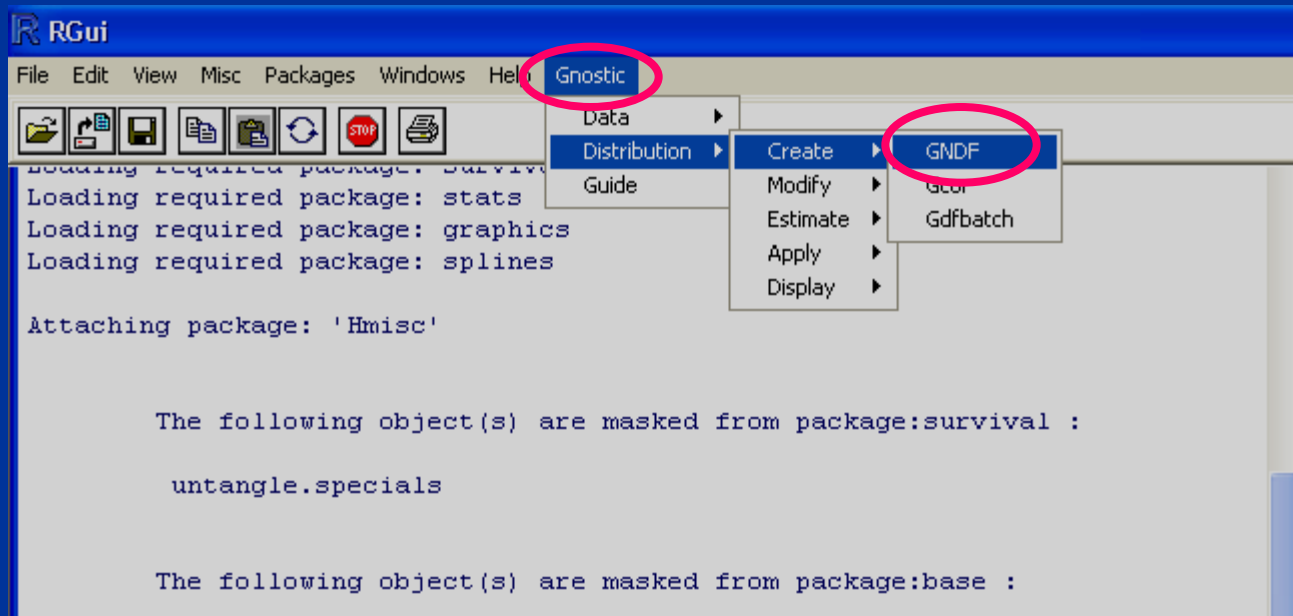
The image shows a screenshot of the R Data Editor window. The window title is "R Data Editor". The data is presented in a table with 17 rows and 2 columns. The first column is labeled "row.names" and the second column is labeled "X2378TCDD". The data values are as follows:

row.names	X2378TCDD
1	1.6
2	2
3	1.5
4	1.6
5	1.6
6	1.4
7	1.8
8	1.4
9	1.2
10	1.7
11	1.5
12	1.7
13	2.4
14	1.4
15	0.9
16	1.2
17	2.1

Annotations in the image include a red arrow pointing to the "row.names" header with the text "Row names", and another red arrow pointing to the "X2378TCDD" header with the text "Column name". Both headers are circled in red.

Gnostická (GUI) aplikace

■ Menu gnostických funkcí



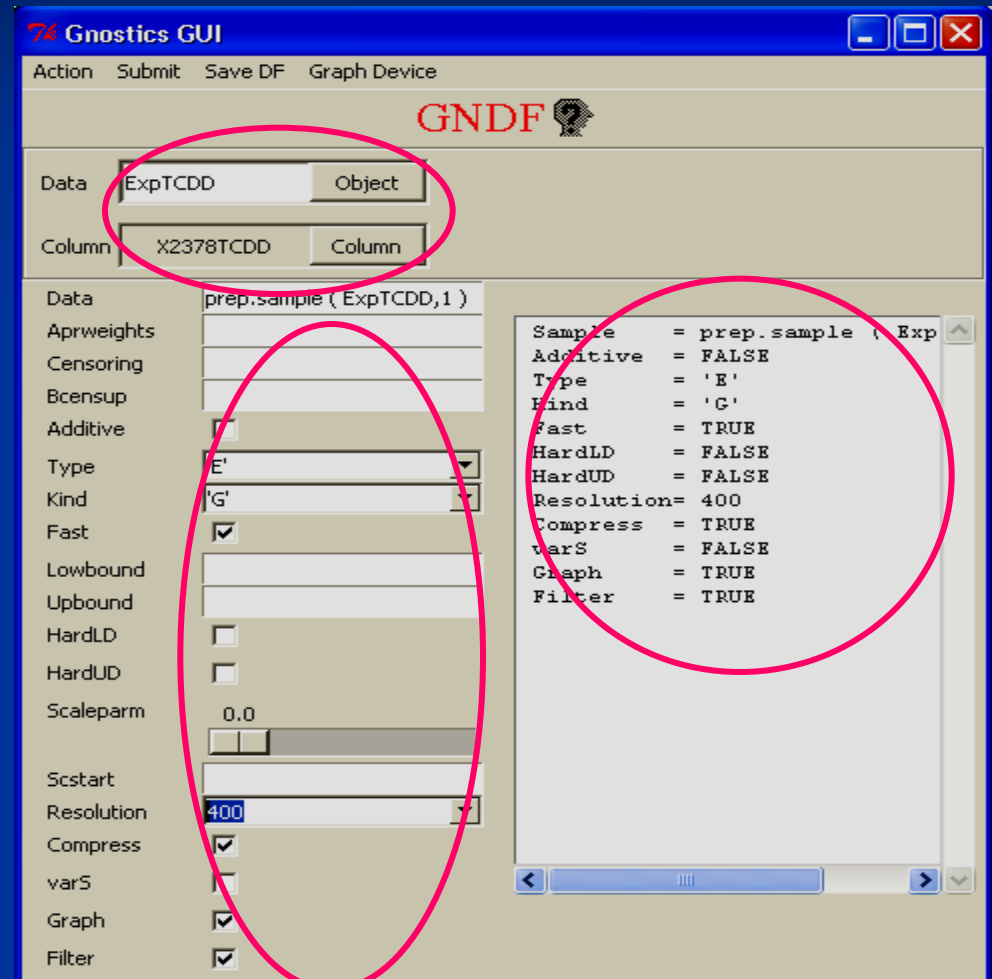
GNDF aplikace

■ GNDF

Data

Parametry

Výstup

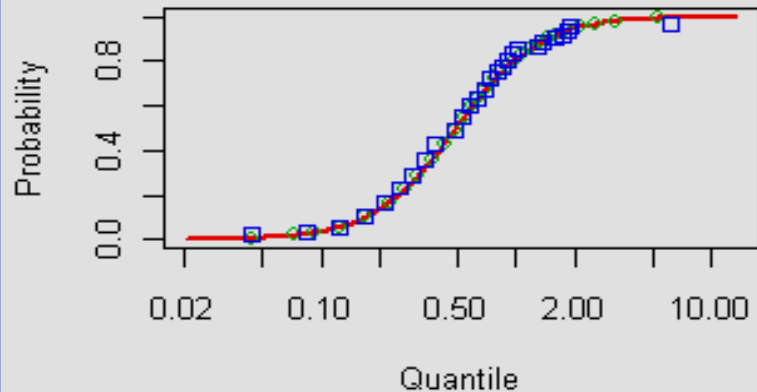


Grafický výstup

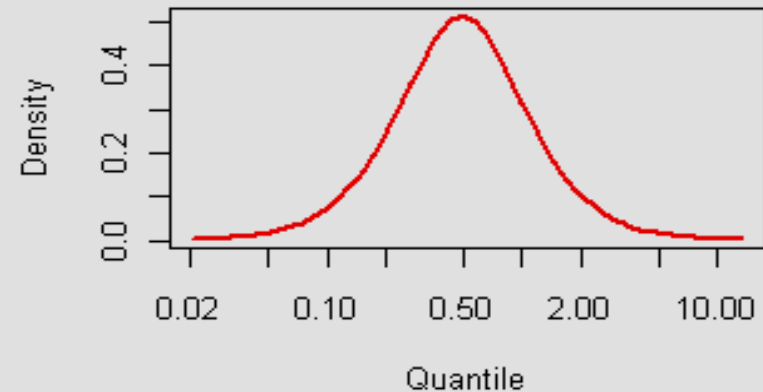
R Graphics: Device 2 (ACTIVE)



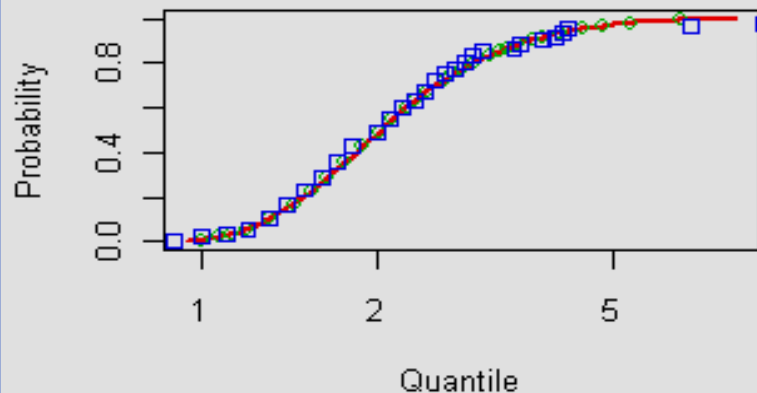
EGDF OF X2378TCDD (INF.)



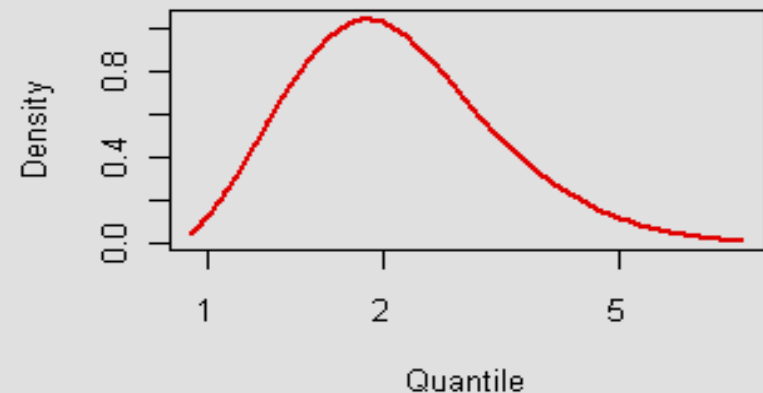
EGDF's DENSITY (INF.)



EGDF OF X2378TCDD (FIN.)



EGDF's DENSITY (FIN.)



Numerický výstup

■ GNDF

Menu výstupů

Okno hodnot

The screenshot shows the Gnostics GUI interface. The 'Run' menu is open, with 'Parameters' highlighted. The 'Output Values' window is also open, displaying a list of parameters and their values. The 'Parameters' menu item and the 'Output Values' window are circled in red.

Run Menu:

- Run
- Parameters
- Data
- Data DF
- Graph DF
- Graph DF(Fin.)
- Devices:off
- Exit

Output Values:

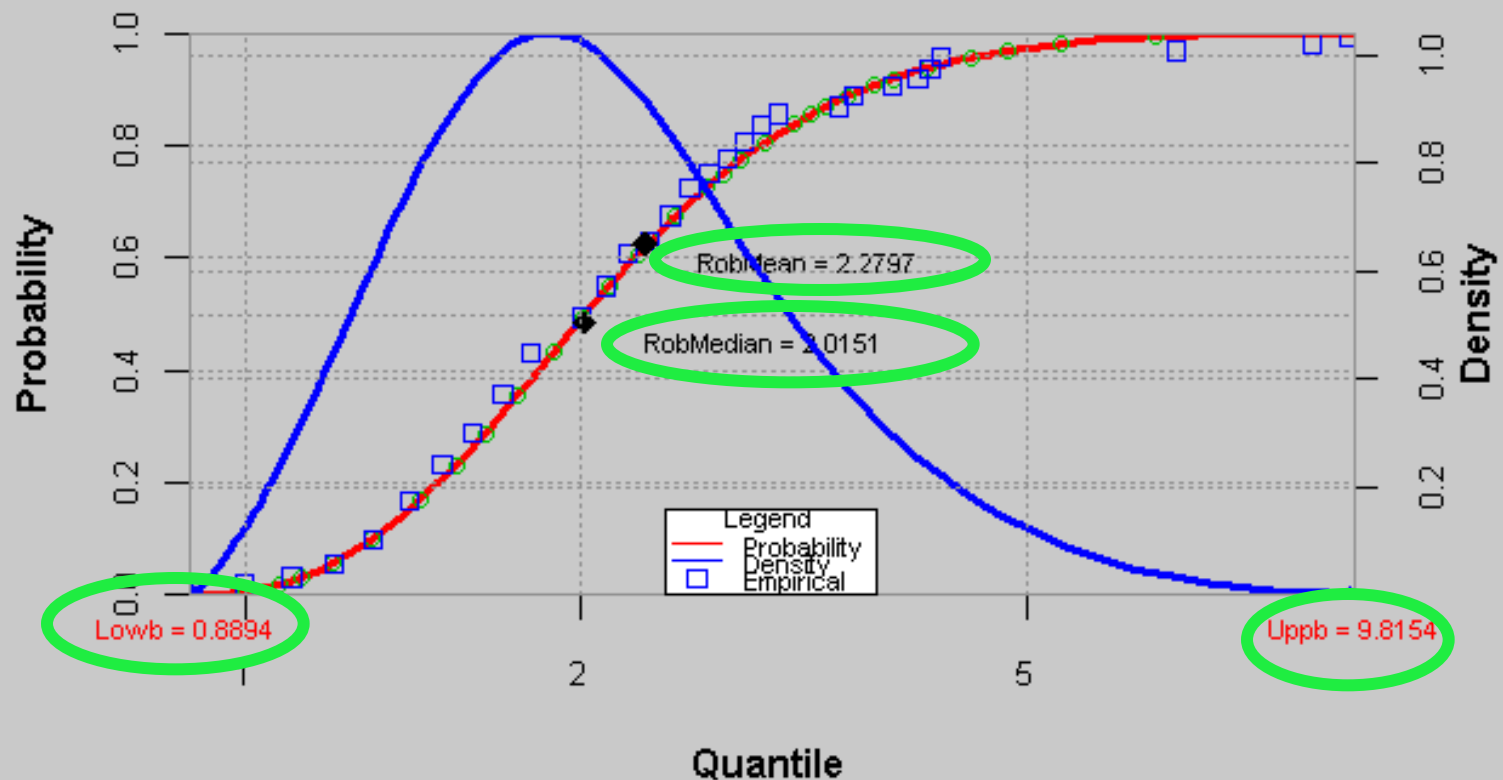
As.add	= FALSE
Type	= E
Scaleopt	= 1.62196452873372
Lowb	= 0.88941704468145
Uppb	= 9.81541530577504
RobMed	= 2.01509577189571
LocParm	= 1.97239206035252
InfQ	= 0.999566238546232
MAFE	= 0.164223222302138
MXFE	= 3.63238129805716
MAPE	= 0.018237491355281
MXPE	= 0.048289878433083
Kind	= G
varS	= FALSE
HardLD	= FALSE
HardUD	= FALSE
RobMean	= 2.27969285366252
STD.DF	= 1.03009285071933

Podrobný graf

R Graphics: Device 3 (ACTIVE)



THE EGDF PROBABILITY AND DENSITY OF X2378TCDD (FIN.)



Závěr

- Výsledky studie gnostické marginální analýzy zpracované v R prostředí potvrdily oprávněnost použití Open Source prostředků pro vytváření SW prostředí gnostických funkcí.
- Tak jako je netradiční způsob analýzy jednorozměrných dat, tak existují vícerozměrné robustní regresní modely.
- Jejich používání zatím omezovaly vysoké náklady na pořízení základního SW a neochota přijmout nové netradiční alternativní postupy.
- Vývoj probíhající v ZÚ Ostrava umožňuje užívání nových metod.

Děkuji za pozornost

Gnostic +

