

Univerzita Pardubice
Fakulta elektrotechniky a informatiky

PROGRAM PRO TESTOVÁNÍ NORMALITY ČÍSELNÝCH DAT

Lukáš Vladyka

Diplomová práce

2017

Univerzita Pardubice
Fakulta elektrotechniky a informatiky
Akademický rok: 2015/2016

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Lukáš Vladyka**
Osobní číslo: **I14219**
Studijní program: **N2612 Elektrotechnika a informatika**
Studijní obor: **Řízení procesů**
Název tématu: **Program pro testování normality číselných dat**
Zadávací katedra: **Katedra řízení procesů**

Z á s a d y p r o v y p r a c o v á n í :

1. Přehled metod testování normality
2. Sestavení programu na testování normality/nenormality jednorozměrných číselných dat, rozsah bodů do 1000. Program bude využívat grafické (EDA) i numerické nástroje. U grafů je třeba vytvořit možnost zadání ano/ne po prohlédnutí a pak pro všechny testy vytvořit závěrečnou tabulku, shrnující výsledky testů, včetně celkového vyhodnocení (tj. soubor má/nemá normální rozdělení).
3. Program musí mít grafické rozhraní, které umožňuje ruční i dávkové načítání dat, taktéž výstup na obrazovku či do souboru (ev. přímo na tiskárnu).
4. Otestovat program vůči renomovaným programům (Adstat, Statistica).

Rozsah grafických prací:

Rozsah pracovní zprávy:

Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam odborné literatury:

MELOUN, M., MILITKÝ, J. Statistická analýza experimentálních dat. Academia, Praha 2004.

HENDL, J. Přehled statistických metod. Praha: Portál, 2012. ISBN 978-80-262-0200-4. 734 str.

HEBÁK, P. a kol. Statistické myšlení a nástroje analýzy dat. Praha: Informatorium, 2013. ISBN 978-80-7333-105-4. 876 str.

Vedoucí diplomové práce:

doc. Ing. Milan Javůrek, CSc.

Katedra řízení procesů

Datum zadání diplomové práce:

30. října 2015

Termín odevzdání diplomové práce:

13. května 2016



prof. Ing. Simeon Karamazov, Dr.
děkan



L.S.



Ing. Daniel Honc, Ph.D.
vedoucí katedry

V Pardubicích dne 14. listopadu 2015

Prohlášení

Prohlašuji:

Tuto práci jsem vypracoval samostatně. Veškeré literární prameny a informace, které jsem v práci využil, jsou uvedeny v seznamu použité literatury.

Byl jsem seznámen s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Beru na vědomí, že v souladu s § 47b zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších předpisů, a směrnicí Univerzity Pardubice č. 9/2012, bude práce zveřejněna v Univerzitní knihovně a prostřednictvím Digitální knihovny Univerzity Pardubice.

V Pardubicích dne 5. 4. 2017

Lukáš Vladyka

Poděkování

Chci poděkovat docentovi Milanu Javůrkovi za vstřícnou výpomoc při hledání správných vzorců pro numerické výpočty a tabulek kritických hodnot.

V Pardubicích dne 5. 4. 2017

Lukáš Vladyka

ANOTACE

Diplomová práce se zaměřuje na vyhodnocování normality / nenormality jednorozměrných číselných dat o rozsahu do 1 000 bodů. Pro vyhodnocení je vytvořen program, který disponuje grafickým rozhraním. V tomto grafickém rozhraní se provádí testování jednorozměrných číselných dat, která se porovnávají s daty z renomovaných programů a zjišťují se tak odchylky. Závěrem diplomové práce je vyhodnocení odchylek při testování normality / nenormality jednorozměrných číselných dat.

KLÍČOVÁ SLOVA

Normalita dat, jednorozměrná číselná data

TITLE

PROGRAM FOR TESTING NORMAL OF NUMERICAL DATA

ANNOTATION

The thesis focuses for evaluation normality / abnormality of univariate numerical data in a range from 1 000 points. Program are created for evaluation, that this program has graphical interface. This graphical interface is conducted testing of univariate numerical data, that this is compared with data from renowned program and find the variances. In the end this thesis is evaluation variances when tested normality / abnormality of univariate numerical data.

KEYWORDS

Data normality, Univariate numerical data

OBSAH

Seznam zkratk a značek	9
Seznam symbolů proměnných veličin a funkcí	10
Seznam obrázků	11
Seznam tabulek	14
ÚVOD	15
1 TEORETICKÁ ČÁST	16
1.1 Grafická identifikace statistických zvláštností dat.....	18
1.1.1 Kvantilový graf	19
1.1.2 Diagram rozptýlení	19
1.1.3 Rozmítnutý diagram rozptýlení	19
1.1.4 Krabicový graf	19
1.1.5 Vrubový krabicový graf	20
1.1.6 Graf symetrie.....	21
1.1.7 Graf polosum.....	21
1.1.8 Graf rozptýlení s kvantily	21
1.1.9 Graf hustoty pravděpodobnosti	22
1.1.10 Kvantil-kvantilový graf (graf Q-Q)	23
1.2 Numerické testy	24
1.2.1 Pearsonův Chí-kvadrát pro test dobré shody	24
1.2.2 Test normality podle Shapira a Wilka.....	25
1.2.3 Testy koeficientů šikmosti a špičatosti	26
1.2.4 Andersonův – Darlingův test	26
1.2.5 Wilcoxonův pořadový test	27
1.2.6 Test podle Kolmogorova-Smirnova.....	27
1.3 JAVA	28
1.4 GUI.....	28
1.5 AWT.....	29
1.6 SWING.....	29
1.7 SWT	29
1.8 Jak se kreslí	29
2 PROGRAM PRO TESTOVÁNÍ NORMALITY DAT	30
2.1 Hlavní menu.....	30

2.2	Ruční zadávání dat.....	31
2.3	Import dat ze souboru	32
2.4	Tabulka s importovanými daty	34
2.5	Výpis výsledků exploratorní analýzy.....	35
2.6	Zobrazení grafů.....	36
2.7	Ukládání dat do souboru	37
3	POUŽÍVÁNÍ PROGRAMU	38
4	OTESTOVÁNÍ PROGRAMU.....	39
4.1	Ověřování Pearsonova Chí-kvadrát pro test dobré shody.....	40
4.2	Ověřování normality podle Shapira a Wilka	40
4.3	Ověřování koeficientů šikmosti a špičatosti	40
4.4	Ověřování Andersonova-Darlingova testu	41
4.5	Ověřování Wilcoxonova pořadového testu.....	42
4.6	Ověřování testu podle Kolmogorova-Smirnova	43
4.7	Testování programu na předem stanovených úlohách.....	44
4.7.1	Úloha B2.02	44
4.7.2	Úloha B2.12	52
4.7.3	Úloha B2.22	59
5	ZÁVĚR	66
	POUŽITÁ LITERATURA	67
	PŘÍLOHY	68

SEZNAM ZKRATEK A ZNAČEK

GUI	Graphical User Interface
AWT	Abstract Windowing Toolkit
JFC	Java Foundation Classes
SWT	Standard Widget Toolkit
IDE	Integrated Development Environment
JDK	Java Development Kit
API	Application Programming Interface

SEZNAM SYMBOLŮ PROMĚNNÝCH VELIČIN A FUNKCÍ

α	zvolená hladina významnosti v intervalu $\langle 0;1 \rangle$
$A_{n,i}$	tabelované hodnoty pro četnost n a pořadí i
D	sedecil
E	oktil
F	kvartil
Φ	hodnota distribuční funkce normálního rozdělení
h	šířka pásu
i	pořadová hodnota pořádkové statistiky
K	jádrová funkce symetrická kolem nuly
k	počet parametrů distribuční funkce
L	vybírá se z náhodně generovaných čísel $\langle 1; 6 \rangle$
M	medián
n	počet prvků ve výběrovém souboru
R_F	rozpětí kvartilů
σ	směrodatná odchylka
s	směrodatná odchylka
s^2	odhad rozptylu
T	menší hodnota ze součtů kladných a záporných pořadí
μ	střední hodnota výběru
\bar{x}	aritmetický průměr
x_{\max}	největší prvek výběru
x_{\min}	nejmenší prvek výběru
Z_i	tabulkové hodnoty normovaného normálního rozdělení.

SEZNAM OBRÁZKŮ

Obr. 1.1 – Konstrukce a) krabicového grafu a b) vrubového krabicového grafu	20
Obr. 1.2 – Graf rozptýlení s kvantily	22
Obr. 2.1 – Položka soubor hlavního menu.....	30
Obr. 2.2 – Položka data hlavního menu.....	31
Obr. 2.3 – Položka grafy hlavního menu	31
Obr. 2.4 – Dialogové okno pro vytvoření tab.	31
Obr. 2.5 – Dialogové okno pro import dat ze souboru	32
Obr. 2.6 – Tabulka zadaných dat s ovládacími prvky	34
Obr. 2.7 – Výpis výsledků exploratorní analýzy	35
Obr. 2.8 – Zobrazení kvantilového grafu.....	36
Obr. 2.9 – Dialogové okno pro ukládání dat do souboru.....	38
Obr. 4.1 – Hodnoty výběru pro testování Andersonova-Darlingova testu	41
Obr. 4.2 – Tabulka s hodnotami pro příklad provádění Wilcoxonova testu.....	42
Obr. 4.3 – Tabulka s hodnotami pro příklad postupu testu podle Kolmogorova-Smirnova.....	43
Obr. 4.4 – Zadání úlohy B2.02	44
Obr. 4.5 – Kvantilový graf úlohy B2.02 Adstat.....	47
Obr. 4.6 – Kvantilový graf úlohy B2.02 QCExpert.....	47
Obr. 4.7 – Kvantilový graf úlohy B2.02 testovaného programu	47
Obr. 4.8 – Bodový a krabicový graf úlohy B2.02 Adstat	47
Obr. 4.9 – Bodový a krabicový graf úlohy B2.02 QCExpert	47
Obr. 4.10 – Bodový a krabicový graf úlohy B2.02 testovaného programu	48
Obr. 4.11 – Graf polosum úlohy B2.02 Adstat	48
Obr. 4.12 – Graf polosum úlohy B2.02 QCExpert	48
Obr. 4.13 – Graf polosum úlohy B2.02 testovaného programu.....	48
Obr. 4.14 – Graf symetrie úlohy B2.02 Adstat	49
Obr. 4.15 – Graf symetrie úlohy B2.02 QCExpert	49
Obr. 4.16 – Graf symetrie úlohy B2.02 testovaného programu.....	49
Obr. 4.17 – Q-Q graf úlohy B2.02 Adstat	49
Obr. 4.18 – Q-Q graf úlohy B2.02 QCExpert.....	49
Obr. 4.19 – Q-Q graf úlohy B2.02 testovaného programu	50
Obr. 4.20 – Graf hustoty pravděpodobnosti úlohy B2.02 Adstat	50
Obr. 4.21 – Graf hustoty pravděpodobnosti úlohy B2.02 QCExpert	50

Obr. 4.22 – Graf hustoty pravděpodobnosti úlohy B2.02 testovaného programu.....	50
Obr. 4.23 – Q-Q graf úlohy B2.02 Adstat	51
Obr. 4.24 – Q-Q graf úlohy B2.02 QCExpert.....	51
Obr. 4.25 – Kvantilový graf rozptýlení úlohy B2.02 testovaného programu.....	51
Obr. 4.26 – Zadání úlohy B2.12	52
Obr. 4.27 – Kvantilový graf úlohy B2.12 Adstat	54
Obr. 4.28 – Kvantilový graf úlohy B2.12 QCExpert.....	54
Obr. 4.29 – Kvantilový graf úlohy B2.12 testovaného programu	54
Obr. 4.30 – Bodový a krabicový graf úlohy B2.12 Adstat	55
Obr. 4.31 – Bodový a krabicový graf úlohy B2.12 QCExpert	55
Obr. 4.32 – Bodový a krabicový graf úlohy B2.12 testovaného programu.....	55
Obr. 4.33 – Graf polosum úlohy B2.12 Adstat.....	55
Obr. 4.34 – Graf polosum úlohy B2.12 QCExpert	55
Obr. 4.35 – Graf polosum úlohy B2.12 testovaného programu.....	56
Obr. 4.36 – Graf symetrie úlohy B2.12 Adstat.....	56
Obr. 4.37 – Graf symetrie úlohy B2.12 QCExpert	56
Obr. 4.38 – Graf symetrie úlohy B2.12 testovaného programu.....	56
Obr. 4.39 – Q-Q graf úlohy B2.12 Adstat	57
Obr. 4.40 – Q-Q graf úlohy B2.12 QCExpert.....	57
Obr. 4.41 – Q-Q graf úlohy B2.12 testovaného programu	57
Obr. 4.42 – Graf hustoty pravděpodobnosti úlohy B2.12 Adstat	57
Obr. 4.43 – Graf hustoty pravděpodobnosti úlohy B2.12 QCExpert	57
Obr. 4.44 – Graf hustoty pravděpodobnosti úlohy B2.12 testovaného programu.....	58
Obr. 4.45 – Q-Q graf úlohy B2.12 Adstat	58
Obr. 4.46 – Q-Q graf úlohy B2.12 QCExpert.....	58
Obr. 4.47 – Kvantilový graf rozptýlení úlohy B2.12 testovaného programu	58
Obr. 4.48 – Zadání úlohy B2.22	59
Obr. 4.49 – Kvantilový graf úlohy B2.22 Adstat	61
Obr. 4.50 – Kvantilový graf úlohy B2.22 QCExpert.....	61
Obr. 4.51 – Kvantilový graf úlohy B2.22 testovaného programu	61
Obr. 4.52 – Bodový a krabicový graf úlohy B2.22 Adstat	61
Obr. 4.53 – Bodový a krabicový graf úlohy B2.22 QCExpert	61
Obr. 4.54 – Bodový a krabicový graf úlohy B2.22 testovaného programu.....	62
Obr. 4.55 – Graf polosum úlohy B2.22 Adstat.....	62

Obr. 4.56 – Graf polosum úlohy B2.22 QCExpert	62
Obr. 4.57 – Graf polosum úlohy B2.22 testovaného programu.....	62
Obr. 4.58 – Graf symetrie úlohy B2.22 Adstat	63
Obr. 4.59 – Graf symetrie úlohy B2.22 QCExpert	63
Obr. 4.60 – Graf symetrie úlohy B2.22 testovaného programu.....	63
Obr. 4.61 – Q-Q graf úlohy B2.22 Adstat	63
Obr. 4.62 – Q-Q graf úlohy B2.22 QCExpert.....	63
Obr. 4.63 – Q-Q graf úlohy B2.22 testovaného programu	64
Obr. 4.64 – Graf hustoty pravděpodobnosti úlohy B2.22 Adstat	64
Obr. 4.65 – Graf hustoty pravděpodobnosti úlohy B2.22 QCExpert	64
Obr. 4.66 – Graf hustoty pravděpodobnosti úlohy B2.22 testovaného programu	64
Obr. 4.67 – Q-Q graf úlohy B2.22 Adstat	65
Obr. 4.68 – Q-Q graf úlohy B2.22 QCExpert.....	65
Obr. 4.69 – Kvantilový graf rozptýlení úlohy B2.22 testovaného programu	65

SEZNAM TABULEK

Tab. 1.1 – Výpis písmenových hodnot	17
Tab. 1.2 – Charakteristiky symetrie, rozptýlení a špičatosti.....	18
Tab. 4.1 - Hodnoty testování Pearsonova Chí-kvadrát testu dobré shody.....	40
Tab. 4.2 - Hodnoty testování testu podle Shapira a Wilka	40
Tab. 4.3 - Hodnoty testování koeficientů šikmosti a špičatosti	41
Tab. 4.4 - Hodnoty testování Andersonova-Darlingova testu	42
Tab. 4.5 - Hodnoty testování Wilcoxonova testu	42
Tab. 4.6 - Hodnoty testování testu podle Kolmogorova-Smirnova.....	43
Tab. 4.7 – Klasické odhady parametrů úlohy B2.02	45
Tab. 4.8 – Písmenové hodnoty úlohy B2.02.....	45
Tab. 4.9 – Kvantilové míry úlohy B2.02	46
Tab. 4.10 – Klasické odhady parametrů úlohy B2.12	52
Tab. 4.11 – Písmenové hodnoty úlohy B2.12.....	53
Tab. 4.12 – Kvantilové míry úlohy B2.12	53
Tab. 4.13 – Klasické odhady parametrů úlohy B2.22	59
Tab. 4.14 – Písmenové hodnoty úlohy B2.22.....	60
Tab. 4.15 – Kvantilové míry úlohy B2.22	60

ÚVOD

Úkolem diplomové práce bylo vytvořit program pro testování normality dat. K tomu slouží průzkumová analýza dat, která se využívá jak v grafické podobě, tak i v podobě numerických nástrojů. Jedná se především o jednorozměrné výběry číselných dat v rozsahu do 1000 bodů.

Pro hodnocení jednorozměrného výběru byl vytvořen program na úrovni grafického rozhraní umožňující nahrávání dat ze souborů, ruční zadávání dat, vyhodnocení dat pomocí numerických i grafických nástrojů, možnost tisku výsledků a také ukládání výsledků do souboru.

Program má sloužit pro vyhodnocování jednorozměrných výběrů číselných dat, které se testuje proti renomovaným programům. Tím se zaručí jistá správnost řešení programu a při testování jsou odhaleny nepřesnosti, které jsou následně vysvětleny. Nadále program může sloužit jako výukový nástroj při vyhodnocování normality/nenormality jednorozměrných výběrů číselných dat.

1 TEORETICKÁ ČÁST

Účelem průzkumové analýzy dat je odhalit jejich zvláštnosti a ověřit předpoklady pro následné statistické zpracování. Při průzkumové analýze složitějších, nákladných nebo unikátních měření je účelem posoudit zvláštnosti chování dat ještě před vlastní statistickou analýzou. Tak lze předcházet provádění numerických výpočtů bez hlubších statistických souvislostí (Meloun, 2004).

Při analýze všech typů měření je nezbytné uvažovat náhodný charakter měřených hodnot x_i , $i = 1, \dots, n$. Při zpracování jednorozměrných výběrů je sledován pouze jeden znak. Počet prvků přitom může být omezený a soubor je pak konečný, nebo neomezený a soubor je nekonečný. Vymezení souboru je velmi důležité, protože ovlivňuje způsob vybírání prvků ze souboru do výběru a interpretaci výsledků. Cílem statistického zpracování je z chování výběru usuzovat na chování celého souboru.

V průzkumové analýze se ze všech různých typů výběrů nejvíce uplatňuje náhodný výběr, jehož prvky, tj. naměřené hodnoty, jsou chápány jako realizace jisté náhodné proměnné. Pro testování normality dat musí výběr prvků splňovat následující předpoklady:

- Jednotlivé prvky výběru x_i jsou vzájemně nezávislé.
- Výběr je homogenní, tj. všechna x_i pocházejí ze stejného rozdělení pravděpodobnosti s konstantním rozptylem.
- Všechny prvky souboru mají stejnou pravděpodobnost, že budou zařazeny do výběru (Meloun, 2004).

Uvedené předpoklady jsou základem statických metod vyhodnocení výsledků měření. Pokud výběr prvků ze souboru nespĺňuje uvedené předpoklady, je jeho statistická analýza daleko komplikovanější a nelze použít klasické postupy průzkumové analýzy dat.

Pro průzkumovou analýzu se především používají grafické metody, které umožňují uživateli posoudit statické zvláštnosti v datech. Grafické metody jsou také vhodné pro zjednodušený popis dat, identifikaci typu rozdělení výběru a zlepšení rozdělení dat.

Dále se využívají charakteristiky exploratorní analýzy dat, které umožňují sledování chování dat výběru. Vychází se z pořádkových statistik, které lze chápat jako vzestupně seříděné prvky výběru $x_1 < x_2 < \dots < x_n$. Dále se z pořádkové statistiky vypočítá pořadová pravděpodobnost označená jako

$$P_i = \frac{i}{n+1}, \quad (1.1)$$

kde i – pořadová hodnota pořádkové statistiky,

n – počet prvků ve výběrovém souboru.

Optimální hodnoty P_i závisí na předpokládaném rozdělení výběru. Pro normální rozdělení se volí

$$P_i = \frac{i - \frac{3}{8}}{n + \frac{1}{4}} \quad (1.2)$$

a pro průzkumovou analýzu se volí

$$P_i = \frac{i - \frac{1}{3}}{n + \frac{1}{3}} \quad (1.3)$$

Vynesením závislosti hodnot x_i na P_i , kde $i = 1, \dots, n$, se získá hrubý odhad kvantilové funkce. Ta je inverzní k distribuční funkci a charakterizuje rozdělení výběru. Pro libovolnou hodnotu α v intervalu $[0, 1]$ se vyčíslí 100α % kvantil za pomoci lineární interpolace

$$\tilde{x}_\alpha = (n+1) \left(\alpha - \frac{i}{n+1} \right) (x_{i+1} - x_i) + x_i, \quad (1.4)$$

kde pro pořadovou hodnotu i pořádkové statistiky musí být splněna nerovnost

$$\frac{i}{n+1} \leq \alpha \leq \frac{i+1}{n+1}. \quad (1.5)$$

V průzkumové analýze se používá kvantilů označené písmenem L pro pořadové pravděpodobnosti $P_i = 2^{-i}$, kde $i = 1, 2, \dots$, které se nazývají písmenové hodnoty. Užívají se pro značení, které je uvedené v tab. 1.1.

Tab. 1.1 – Výpis písmenových hodnot (Meloun, 2004)

i	kvantil	Pořadová pravděpodobnost P_i	Symbol písmenové hodnoty L	Hodnota kvantilu u_{P_i}
1	medián	$2^{-1} = 1/2$	M	0
2	kvartily	$2^{-2} = 1/4$	F	-0,674
3	oktily	$2^{-3} = 1/8$	E	-1,150
4	sedecily	$2^{-4} = 1/16$	D	-1,530

Symbol u_{P_i} označuje kvantil normovaného normálního rozdělení $N(0,1)$. Kromě mediánu existují pro každé $i > 1$ dvojice kvantilů, a to dolní i horní písmenová hodnota L_D a

L_H . Dolní písmenová hodnota je pro pořadovou pravděpodobnost $P_i = 2^{-i}$, ale naopak horní písmenová hodnota je pro $P_i = 1 - 2^{-i}$.

Mezi základní statistické zvláštnosti rozdělení dat patří symetrie výběrového rozdělení a jeho relativní délky konců. K vyjádření symetrie a špičatosti se využívají jednoduché funkční charakteristiky vycházející z písmenových hodnot. Definice jsou uvedeny v tab. 1.2.

Tab. 1.2 – Charakteristiky symetrie, rozptýlení a špičatosti (Meloun, 2004)

Název	Definice	Charakteristika	Platí pro L
Polosuma Z_L	$0,5(L_D + L_H)$	symetrie při $Z_L = 0$	F, E, D
Rozpětí R_L	$(L_H - L_D)$	rozptýlení	F, E, D
Šikmost S_L	$(M - Z_L)/R_L$	symetrie při $S_L = 0$	F, E, D
Pseudosigma* G_L	$R_L / (-2u_{P_i})$	špičatost (Gaussovo $G_L = \text{konst.}$)	F, E, D
Délky konců T_L	$\ln(R_L/R_F)$	špičatost	E, D

* u_{P_i} – kvantil standardizovaného normálního rozdělení pro $P_i = 2^{-i}$

Hodnoty šikmosti S_L jsou záporné, pokud je rozdělení zešikmené k vyšším hodnotám a při zešikmení k nižším hodnotám naopak kladné. Rozdělení s delšími konci, než má normální rozdělení, rostou hodnoty pseudosigmy G_L se vzdáleností od mediánu. Naopak kratší konce, než normální rozdělení je rozdělení, jejichž hodnoty G_L klesají s rostoucí vzdáleností od mediánu.

Další hodnotou vyjádřenou pro statistiku je centrální moment. Pro přirozené číslo je j -tý centrální moment jisté reálné číslo charakterizující rozdělení náhodné veličiny. Centrální moment je definován vztahem

$$M_j = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^j, \quad (1.6)$$

kde \bar{x} – aritmetický průměr.

1.1 GRAFICKÁ IDENTIFIKACE STATISTICKÝCH ZVLÁŠTNOSTÍ DAT

Mezi základní statistické zvláštnosti především patří symetrie a špičatost rozdělení výběru, lokální koncentrace dat a přítomnost vybočujících dat. Jednotlivé grafy dávají informaci o několika, případně pouze jedné statistické zvláštnosti.

1.1.1 Kvantilový graf

Umožňuje přehledně znázornit data a umožnit uživateli rozlišit tvar rozdělení. Rozdělení může být symetrické, zešikmené k vyšším nebo nižším hodnotám, mohou být lokálně koncentrovaná data nebo vybočující data. U kvantilového grafu se na osu x vynáší pořadová pravděpodobnost P_i a na osu y pořadková statistika x_i . Pro snadnější porovnání s normálním rozdělením se do grafu zakreslují i kvantilové funkce normálního rozdělení $N_{P_i} = \hat{\mu} + \hat{\sigma}u_{P_i}$, pro $0 \leq P_i \leq 1$. Používají se:

1. klasické odhady parametry polohy a rozptýlení $\hat{\mu} = \bar{x}$ a $\hat{\sigma} = s$,
2. robustní odhady $\hat{\mu} = \tilde{x}_{0,5}$ a $\hat{\sigma} = R_F / 1,349$,

kde R_F – rozpětí kvartilů
 s – směrodatná odchylka.

1.1.2 Diagram rozptýlení

Představuje jednorozměrný kvantilový graf do osy x . U diagramu rozptýlení se na osu x vynáší hodnoty x_i výběrového souboru a na osu y je zvolena libovolná úroveň, obvykle se volí $y = 0$. I pro svou jednoduchost tento diagram ukazuje lokální koncentraci dat a indikuje podezřelá a vybočující data.

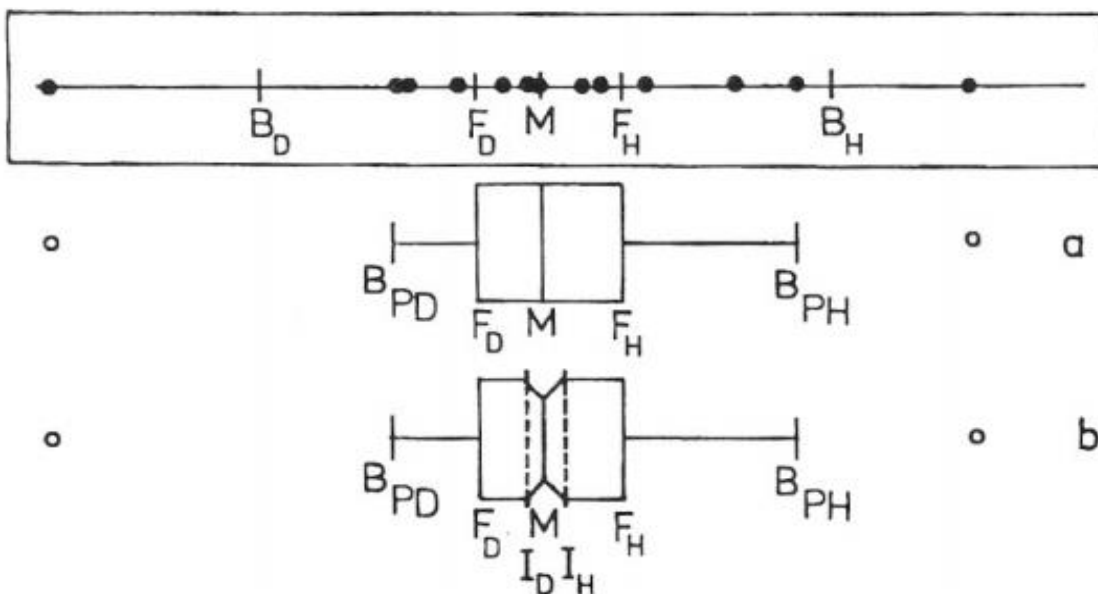
1.1.3 Rozmítnutý diagram rozptýlení

Diagram představuje rovněž projekci kvantilového grafu. Body jsou však rozmítnuté ve směru osy y tak, aby nedocházelo ke splývání bodů a byly tak jednotlivé body zřetelné. Další vylepšení schopností je rozmítání dat s variabilní výškou závislou na hustotě pravděpodobnosti $f(x)$, konstruované na základě počtu hodnot nalézajících se v okolí místa x .

1.1.4 Krabicový graf

Krabicový graf lze využít z části pro sumarizaci dat, která umožňuje

1. zobrazit robustní odhady polohy, medián,
2. posoudit symetrii v okolí kvantilů,
3. posoudit symetrii u konců rozdělení,
4. identifikovat odlehlé body (Meloun, 2004).



Obr. 1.1 – Konstrukce a) krabicového grafu a b) vrubového krabicového grafu

V nejjednodušší variantě má krabicový graf tvar obdélníku o délce $R_F = F_H - F_D = \tilde{x}_{0,75} - \tilde{x}_{0,25}$ s vhodně zvolenou šířkou, která je úměrná hodnotě \sqrt{n} . V místě mediánu se zobrazuje vertikální čára. Od protilehlých stran obdélníku pokračují úsečky, které jsou ukončeny přílehlými hodnotami B_{PH} a B_{PD} ležícími uvnitř vnitřních hradeb nejbližše k jejich hranicím B_H a B_D . Vše je vyobrazeno výše na obr. 1.1. Vnitřní hradby lze vyjádřit jako

$$B_H = F_H + 1,5R_F, \quad (1.7)$$

$$B_D = F_D - 1,5R_F, \quad (1.8)$$

kde F_H – horní kvartil,

F_D – dolní kvartil.

Krabicové grafy se pro svoji jednoduchost a přehlednost používají k porovnání několika výběrů.

1.1.5 Vrubový krabicový graf

Tento graf je obdobou krabicového grafu, který umožňuje posoudit variabilitu mediánu. Variabilita je vyjádřena robustním intervalem spolehlivosti $I_D \leq M \leq I_H$. Tyto meze jsou dány vztahy

$$I_D = M - \frac{1,57R_F}{\sqrt{n}}, \quad (1.9)$$

$$I_H = M + \frac{1,57R_F}{\sqrt{n}}. \quad (1.10)$$

1.1.6 Graf symetrie

Symetrická rozdělení jsou charakterizována horizontální přímkou $y = \tilde{x}_{0,5} = M$. V případě, že přímka nemá nenulovou směrnici, je směrnice odhadem parametru šikmosti.

U grafu symetrie se na osu x vynáší $\frac{u_{P_i}^2}{2}$ pro vztah (1.1) a na osu y se vynáší

$$Z_i = 0,5(x_{(n+1-i)} + x_i). \quad (1.11)$$

1.1.7 Graf polosum

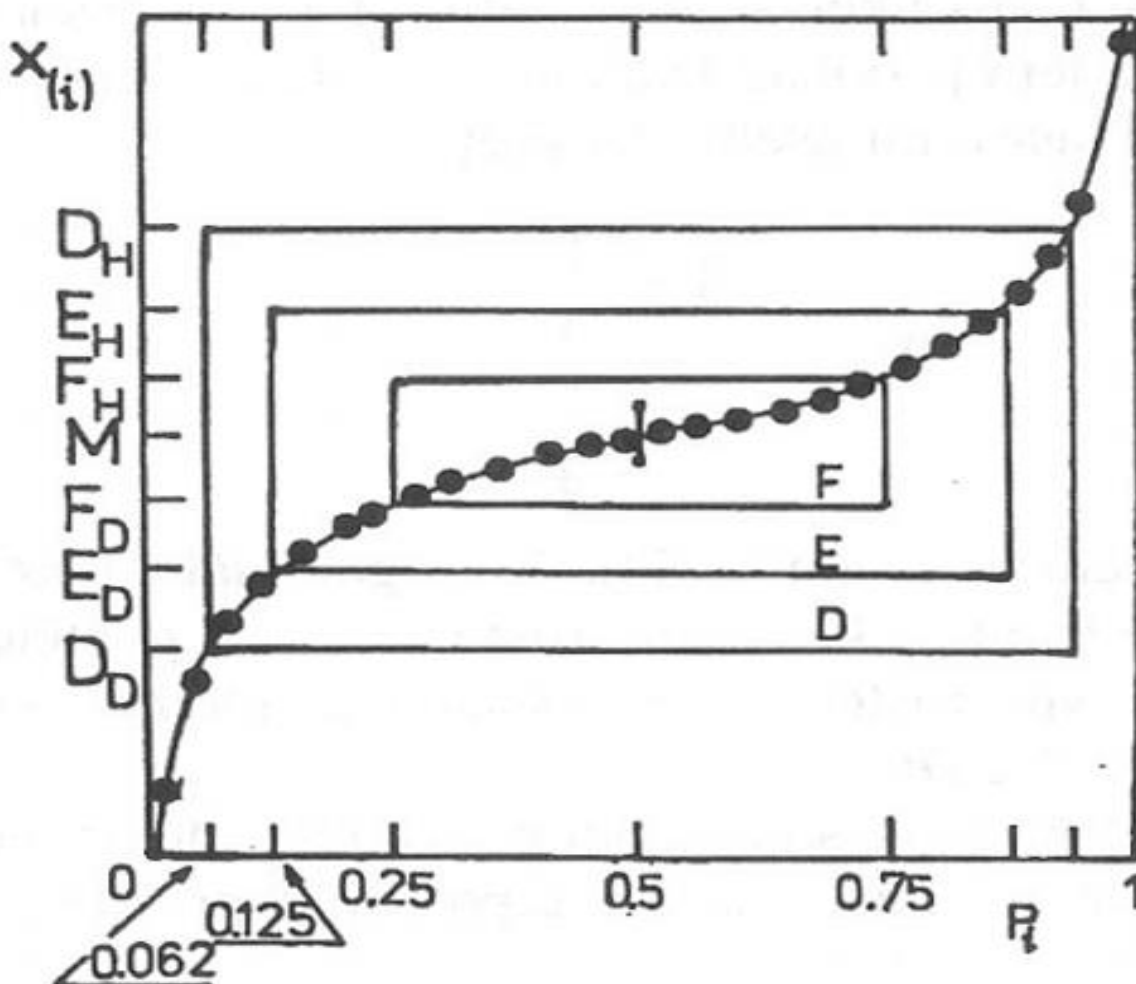
Graf má vyneseno na osu x pořadové statistiky hodnot x_i výběrového souboru. Symetrická rozdělení jsou charakterizována horizontální přímkou hodnoty mediánu $y = \tilde{x}_{0,5} = M$. Na grafu se dobře ověřuje symetrie výběru s tím, že je nutné sledovat rozmezí hodnota na ose y . Na osu y se vynáší hodnoty vztahu (1.10).

1.1.8 Graf rozptýlení s kvantily

U grafu rozptýlení s kvantily je kvantilová funkce tvořena z pořadové pravděpodobnosti P_i vyneseno na osu x a pořádkové statistiky x_i vyneseno na osu y . Především vyniká jednoduchou konstrukcí a univerzálností k posouzení statistických zvláštností dat. Pro pořadovou pravděpodobnost se obvykle volí vztah (1.3). V případě symetrického rozdělení má kvantilová funkce sigmoidální tvar. Pro rozdělení zešikmené k vyšším hodnotám je konvexně rostoucí a pro rozdělení zešikmené k nižším hodnotám konkávně rostoucí. V grafu se vynáší s kvantilovou funkcí také tři obdélníky F, E a D pro snadnější interpretaci:

1. Kvantilový obdélník F má na ose y vrcholy dané hodnotami kvartilů F_D a F_H . Na osu x se vynáší hodnoty pořadové pravděpodobnosti $P_2 = 2^{-2} = 0,25$ a $1 - 2^{-2} = 0,75$ (Meloun, 2004).
2. Oktilový obdélník E má na ose y vrcholy dané hodnotami oktilů E_D a E_H . Na osu x se vynáší hodnoty pořadové pravděpodobnosti $P_3 = 2^{-3} = 0,125$ a $1 - 2^{-3} = 0,875$ (Meloun, 2004).

3. Sedecilový obdélník D má na osy y vrcholy dané hodnotami sedecily D_D a D_H . Na osu x se vynáší hodnoty pořadové pravděpodobnosti $P_4 = 2^{-4} = 0,0625$ a $1 - 2^{-4} = 0,9375$ (Meloun, 2004).



Obr. 1.2 – Graf rozptýlení s kvantily

1.1.9 Graf hustoty pravděpodobnosti

Pro střední výběry je možné konstruovat grafy hustoty pravděpodobnosti podle vztahu

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left[\frac{(x-x_i)}{h}\right], \quad (1.12)$$

kde h – šířka pásu, který určuje stupeň vyhlazení,

K – jádrová funkce symetrická kolem nuly a má všechny vlastnosti hustoty pravděpodobnosti.

Vlastní výběr tvaru jádrové funkce není pro odhad $\hat{f}(x)$ rozhodující. Omezíme se proto na jednoduché bikvadratické jádro

$$K(x) = 0,9375(1 - x^2)^2 \text{ pro } -1 \leq x \leq 1, \quad (1.13)$$

$$K(x) = 0 \text{ pro } x < -1 \text{ nebo } x > 1. \quad (1.14)$$

Volba parametru h rozhoduje o kvalitě odhadu $\hat{f}(x)$. Při průzkumové analýze se doporučuje obecnější postup. Nejdříve se podle vztahu (1.12) provede předběžný odhad hustoty $\hat{f}_0(x)$ s šířkou pásu

$$h_0 = 0,75 \left(\frac{n}{100} \right)^{-0,2} \min_i (x_{i+\text{int}(n/2)} - x_i), \quad (1.15)$$

a jádrovou funkcí podle vztahů (1.14) a (1.15). Nakonec se sestrojí vlastní odhad hustoty pravděpodobnosti $\hat{f}_K(x)$ s využitím jádrové funkce podle vztahů (1.13) a (1.14), ale s nekonztantní šířkou pásu

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_i} K \left[\frac{(x - x_i)}{h_i} \right]. \quad (1.16)$$

Uvedenou konstantní šířku pásu lze vyjádřit vztahem

$$h_i = h_0 \left(\frac{\hat{f}_0(x_i)}{\max_i \hat{f}_0(x_i)} \right)^{-\alpha}, \quad (1.17)$$

kde α – zvolená hladina významnosti v intervalu $\langle 0; 1 \rangle$ a ovlivňuje hladkost odhadu $\hat{f}(x)$.

U grafu hustoty pravděpodobnosti se na osu x vynáší pořádková statistika x_i a na osu y odhad hustoty pravděpodobnosti $\hat{f}(x)$. K základnímu empirickému odhadu hustoty pravděpodobnosti se pro porovnání vynáší do grafu také normální (Gaussův) odhad hustoty pravděpodobnosti ve vztahu

$$\hat{f}_G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (1.18)$$

kde σ – směrodatná odchylka,

μ – střední hodnota výběru.

1.1.10 Kvantil-quantilový graf (graf Q-Q)

Z grafických technik pro porovnání výběrového rozdělení s rozděleními teoretickými jsou nejpoužívanější grafy Q-Q. Pomocí těchto grafů je možné posoudit shodu výběrového

rozdělení, jež je charakterizováno kvantilovou funkcí $Q_E(P)$, s kvantilovou funkcí zvoleného teoretického rozdělení $Q_T(P)$. Odhadem kvantilové funkce výběru je pořádková statistika x_i . U Q-Q grafu se na osu x vynáší skutečné rozdělení $Q_S(P_i)$ a na osu y se vynáší x_i . Skutečné rozdělení $Q_S(P_i)$ je podstatě rovna inverzi distribuční funkce

$$F^{-1}(P_i) = \frac{-9,4 \ln\left(\frac{1}{P_i} - 1\right)}{\text{abs}\left[\ln\left(\frac{1}{P_i} - 1\right)\right] + 14}. \quad (1.19)$$

K posouzení linearity se provádí lineární regrese a vypočtená regresní přímka se zakresluje přímo do Q-Q grafu.

1.2 NUMERICKÉ TESTY

Kromě řady výše uvedených grafických nástrojů se používá řada klasických numerických testů. Vyhodnocení jsou jednoznačná, ale při zpracování dat dochází ke ztrátě informací. Například dochází k numerické kompenzaci různých odchylek, které vyčíslují určité statické charakteristiky, jež jsou následně porovnávány s tabulkami tzv. kritických hodnot pro určitou statistickou míru jistoty a určité rozdělení pravděpodobnosti. U numerických testů je rozhodování jednoznačné, protože experimentální hodnoty výpočtů jsou porovnávány s tabelovanými kritických hodnot numerických testů. Testy využívající výše popsany postup označujeme jako parametrické testy a testy neparametrické, které využívají spíše empirické porovnávací postupy.

1.2.1 Pearsonův Chí-kvadrát pro test dobré shody

Tento test slouží k objektivnímu zhodnocení, do jaké míry se rozdělení náhodného výběru shoduje s předpokládaným rozdělením. Nejlépe se testují zejména výběry většího rozsahu, a to pro spojitá i nespojitá rozdělení. Je založen na porovnání celkové neshody mezi skutečnými četnostmi výskytu náhodné veličiny v určitých třídách a četnostmi, které lze očekávat, má-li náhodná veličina dané rozdělení.

Náhodný výběr ze souboru se roztřídí do m tříd, kde četnost náhodného výběru n vychází ze vztahu

$$m = 3,322 \log n + L, \quad (1.20)$$

kde L – vybírá se z náhodně generovaných čísel $\langle 1; 6 \rangle$.

Šířka tříd je dána vztahem

$$h = 1,001 \frac{x_{\max} - x_{\min}}{m}, \quad (1.21)$$

kde x_{\max} – největší prvek výběru,

x_{\min} – nejmenší prvek výběru.

Hranice tříd jsou dány vztahem

$$r_0 = x_{\min} - 0,0005(x_{\max} - x_{\min}), \quad (1.22)$$

$$r_j = r_{j-1} + h \quad (1.23)$$

a středy tříd jsou dány vztahem

$$s_j = 0,5(r_j - r_{j-1}). \quad (1.24)$$

Náhodná veličina může nabývat hodnot v rozsahu $(-\infty; \infty)$, a proto se rozšíří první interval od $-\infty$ a poslední interval do ∞ . Pravděpodobnosti všech intervalů se počítají jako rozdíl distribuční funkce normálního rozdělení v horním a dolním intervalu, kde se používají normované hodnoty. Normování provedeme pomocí dosazení vypočtených hodnot aritmetického průměru a směrodatné odchylky

$$p_j = \Phi \left[\frac{(r_j - \bar{x})}{s} \right] - \Phi \left[\frac{(r_{j-1} - \bar{x})}{s} \right], \quad (1.25)$$

kde Φ – hodnota distribuční funkce normálního rozdělení.

Testovací charakteristika má χ^2 rozdělení podle vzorce

$$\chi_{\text{exp}}^2 = \sum_{j=1}^m \frac{(n_j - np_j)^2}{np_j}, \quad (1.26)$$

která se porovná s tabelovanou kritickou hodnotou $\chi^2(\alpha, m - k - 1)$,

kde k – počet parametrů distribuční funkce, tj. 2 hodnoty.

1.2.2 Test normality podle Shapira a Wilka

Analyzovaný výběrový soubor dat se setřídí vzestupně podle velikosti hodnot. Vypočte se koeficient podle vztahu

$$b = \sum_{i=1}^j A_{n,i} (x_{n-i+1} - x_i), \quad (1.27)$$

kde $A_{n,i}$ – tabelované hodnoty pro četnost n a pořadí i ,

j – hodnota $n/2$ pro sudé n a $(n-1)/2$ pro liché n .

Testovací charakteristika se vypočítává podle vztahu

$$W_{\text{exp}} = \frac{b^2}{(n-1)s^2}, \quad (1.28)$$

kde s^2 – odhad rozptylu.

Výše uvedená testovací charakteristika (1.28) se porovnává se tabelovanými kritickými hodnotami $W_{\text{krit}}(\alpha, n)$.

1.2.3 Testy koeficientů šikmosti a špičatosti

Podobnost aktuálního rozdělení pravděpodobnosti s normálním vyjadřují také třetí a čtvrtý centrální moment, tj. koeficienty šikmosti a špičatosti podle vztahu

$$m_{x,3,\text{exp}} = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^3, \quad (1.29)$$

$$m_{x,4,\text{exp}} = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}. \quad (1.30)$$

Pro normální rozdělení je koeficient šikmosti nulový a koeficient špičatosti je roven hodnotě tři. Kritické hodnoty lze spočítat pomocí vztahů

$$m_{x,3,\text{krit}}(n) = 4,8009 \sqrt{\frac{(n-2)}{((n+1)(n+3))}}, \quad (1.31)$$

$$m_{x,4,\text{krit}}(n) = 3 - \frac{6}{(n+1)} \pm \frac{9,6018}{(n+1)} \sqrt{\frac{n(n-2)(n-3)}{((n+3)(n+5))}}. \quad (1.32)$$

1.2.4 Andersonův – Darlingův test

Testovací kritérium je dáno vztahem

$$AD = - \frac{\sum_{i=1}^n (2i-1)(\ln Z_i + \ln(1-Z_{n-i+1}))}{n} - n, \quad (1.33)$$

kde Z_i – tabulkové hodnoty normovaného normálního rozdělení.

Normalita se zamítá v případě $AD > D_{1-\alpha}$. Hodnota kvantilu $D_{1-\alpha}$ lze vypočítat například pro hladinu významnosti 5 % vztahem

$$D_{0,95} = 1,0348 \left(1 - \frac{1,013}{n} - \frac{0,93}{n^2} \right). \quad (1.34)$$

1.2.5 Wilcoxonův pořadový test

Tento test pracuje se seříděnými daty výběru podle velikosti. Jedná se o silnější test než znaménkový test, ze kterého vychází princip Wilcoxonova pořadového testu. Provádí se porovnání absolutních hodnot vůči předpokládané hodnotě mediánu, přičemž se předpokládá, že součet pořadí hodnot menších T^- a větších T^+ než medián by mělo být zhruba shodné. V případě shodných skupin absolutních hodnot rozdílů měření dosadíme místo jejich pořadí průměrnou hodnotu spočtenou z pořadí shodných skupin. Test nulové hypotézy $H_0 : \tilde{\mu} = \tilde{\mu}_0$ proti alternativě $H_1 : \tilde{\mu} \neq \tilde{\mu}_0$ provedeme srovnáním menší z obou hodnot T^- a T^+ s kritickými hodnotami Wilcoxonova testu. Pokud je T menší než kritická mez, hypotéza H_0 se zamítá na příslušné hladině významnosti.

Při větším rozsahu výběru se využije aproximace pomocí normálního rozdělení. Testovací charakteristika je poté dána vztahem

$$u_{T,\text{exp}} = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}, \quad (1.35)$$

kde T – menší hodnota ze součtů kladných a záporných pořadí.

Testovací charakteristika lze využít pouze při větším počtu dat výběru větší než 25. Výše uvedená modifikace Wilcoxonova testu se označuje jako jednovýběrová. Testovací charakteristika se porovnává s kritickou hodnotou jednovýběrové distribuční funkce.

1.2.6 Test podle Kolmogorova-Smirnova

Test je zcela obecný pro jakýkoli typ rozdělení a zkoumá se pomocí něj průběh celé distribuční funkce. Hodnotí, zda data pocházejí z populace s určitou distribuční funkcí $F_0(x)$. Testuje se hypotéza $H_0 : F(x) = F_0(x)$ proti alternativě $H_A : F(x) \neq F_0(x)$, kde $F(x)$ je distribuční funkce, která generuje data výběru. Při provedení testu se porovná přirozeným

způsobem empirická distribuční funkce $F(x)$ s teoretickou distribuční funkcí $F_0(x)$ pro všechna x . Testovací statistika se počítá jako maximum jejich absolutního rozdílu

$$D = \max |\hat{F}(x) - F_0(x)|, \quad (1.36)$$

kteřá se porovnává s tabelovanými kritickými hodnotami.

1.3 JAVA

V překladu znamená „kafé“ a jedná se o programovací jazyk pocházející od firmy Sun Microsystems, později koupené firmou Oracle. Objektově orientovaný jazyk vycházející z C++, ke kterému má také syntakticky nejbliž, nepočítáme-li C#, který vznikl až po příchodu programovacího jazyku JAVA.

Velkým přínosem jazyku JAVA je plná přenositelnost programů na libovolnou platformu bez nutnosti jejich rekompile. Programy se nepřekládají do strojového kódu konkrétního procesoru, ale do nezávislé podoby, tzv. bytového kódu (byte code uvedený výše v bodech). Tento kód může být interpretován na jakémkoliv počítači nebo průmyslovém zařízení. Kompatibilita je zajištěna na binární úrovni.

Hlavním praktickým nedostatkem jazyku JAVA je malá rychlost interpretovaných programů. Zejména se to týká internetových prohlížečů, které mají mnohdy problémy se stabilitou appletů, ovšem nikoliv vinou jazyku JAVA. Tento nedostatek však vyřešily JIT kompilátory. Jde vlastně o myšlenku dynamické (re-) kompilace kódu. Dochází tak k překladu byte code v okamžiku, kdy se vykonává, a tak se neukládá do externí cache.

Druhým problémem jsou zvýšené nároky na paměť, vznikající v důsledku automatické správy paměťových prostředků. Kapacita pamětí počítačů však neustále roste, a tudíž se jedná o poněkud méně závažný nedostatek.

1.4 GUI

Původní návrh byl hodně nedokonalý, neumožňoval snadné a systematické využití. Z toho důvodu se již s JDK 1.0 vytvořila u vývojářů jistá averze vůči javovské grafice. V dalších verzích docházelo k neustálému zlepšování, a tím se JAVA brzy stala vhodnou platformou pro tvorbu grafických aplikací.

1.5 AWT

První implementací grafiky byla knihovna zvaná AWT. Objevila se hned na začátku v JDK 1.0 a její nepříjemné a nesystematické API přetrvává v Javě dodnes. Něco se stále používá, ale jsou to spíš věci související s obecnou grafikou, a nikoli s GUI. V dalších verzích bylo sice API přepracováno a podstatně rozšířeno, ale jiné nevýhodné vlastnosti přetrvaly (Kysilka, 2006).

Filosofie AWT byla taková, že každá komponenta v Javě měla svůj nativní protějšek v systému. Byl to tedy podobný model, jako dnes používají některé GUI platformy v C/C++. I když se jednalo o abstraktní na platformě nezávislé rozhraní, přenositelnost byla problematická vzhledem k rozdílným vlastnostem nativní úrovně GUI (Kysilka, 2006).

1.6 SWING

Od verze 1.2 se nachází v jazyce JAVA nová grafická knihovna zvaná JFC, známá spíše pod názvem Swing, dále v textu bakalářské práce je již uváděno toto označení. Původní koncept byl zavržen, Swing byl vytvořen prakticky kompletně na zelené louce, i když v sobě obsahuje AWT API.

Swing má takové vlastnosti, které jsou pro programátory velice příjemné, ale na druhou stranu mají relativně vysoké nároky na rychlost procesoru, a hlavně na paměť. V reakci na to vznikly grafické knihovny, které se snaží tento problém odstraňovat. Nejvýznamnější z nich je SWT (Kysilka, 2006).

1.7 SWT

GUI knihovna s podobnými vlastnostmi, jako má Swing, ovšem s nativní implementací grafických komponent. Původní cíl vzniku, tedy snížení procesorové a paměťové náročnosti, příliš naplněn nebyl a srovnatelné aplikace postavené na SWT jsou zhruba stejně náročné jako ty co používají Swing (Kysilka, 2006).

1.8 JAK SE KRESLÍ

Ke kreslení se používá tzv. grafický kontext, což je implementace abstraktního kreslicího rozhraní. Při kreslení se využívá široká škála nástrojů, která jsou k dispozici ve standardní knihovně. Kreslení nebo spíš skládání grafického rozhraní je díky předpřipraveným komponentám snadnou záležitostí. Jednoduše pomocí tažení myši se komponenty skládají na

grafický panel a každá komponenta, která je využita, se automaticky generuje do kódu. Takto založeným komponentám může uživatel libovolně nastavovat vlastnosti a funkčnosti jak kódově, tak i v nastavení parametrů v grafického rozhraní (Kysilka, 2006).

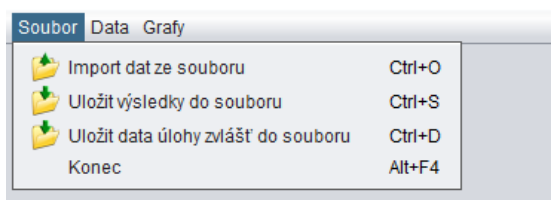
2 PROGRAM PRO TESTOVÁNÍ NORMALITY DAT

Všechny uvedené teoretické poznatky exploratorní analýzy jsou používány v samotném těle celého programu, který na výstupu dává relevantní výsledky, které jsou potvrzeny níže v provedených pokusech. Program disponuje možností nahrávat data dávkově ze souboru nebo pomocí ručního zadávání dat přímo v programu. Umožňuje přehledně zobrazit všechny položky exploratorní analýzy včetně zobrazení hodnocení grafů, zda mají normální rozdělení. Grafy jsou zobrazeny v samostatných dialogových oknech s možností okamžitého hodnocení grafů. Všechny vypočtené hodnoty včetně tabulkového hodnocení grafů je možné ukládat do PDF souboru. Každý graf lze ukládat samostatně nebo rovnou poslat do tiskárny pro tisk. Detailnější popis programu pro testování normality dat je uveden níže v jednotlivých oddílech.

2.1 HLAVNÍ MENU

Program má všechny části rozděleny do logických jednotlivých bloků tak, aby bylo uživateli nabídnuto komfortní ovládání a nemusel vše složitě hledat. Rozdělení je patrné z obrázků níže, kde hlavní menu má tři základní položky jako jsou Soubor, Data a Grafy.

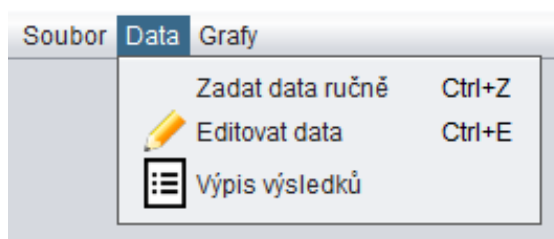
Položka souboru v menu nabízí možnost práce s vkládáním a ukládáním dat programu nebo ukončení celého programu. Na položce Uložit data do souboru je nastavena funkce deaktivace odkazu, což znamená, že v případě programu, který nemá nahraná data, tak nelze logicky nic uložit do souboru. Tím je zajištěno, že položka není spuštěna, dokud se v programu neprovede první výpočet s výpisem výsledků exploratorní analýzy. Vše je vyobrazeno na obr. 2.1 níže, kde jsou všechny položky aktivní.



Obr. 2.1 – Položka soubor hlavního menu

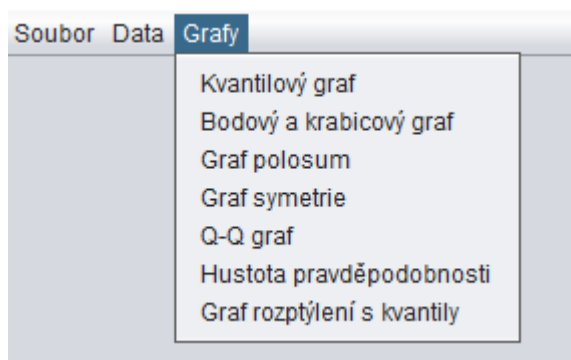
Další položkou menu jsou Data, která v sobě ukrývají možnost Zadání dat ručně, Editace dat a Výpis výsledků. Při prvním spuštění programu je aktivní pouze položka Zadání

dat ručně, protože opět nelze editovat data a vypsat výsledky, když nebylo provedeno žádné zadání dat do programu a jejich výpočet. Vše je vyobrazeno na obr. 2.2 níže, kde jsou všechny položky aktivní.



Obr. 2.2 – Položka data hlavního menu

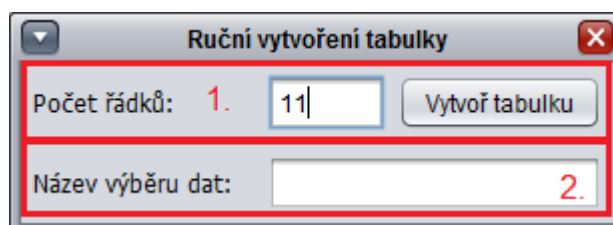
Poslední položkou menu jsou Grafy. V listu položky jsou odkazy na grafy exploratorní analýzy s tím, že všechny jsou neaktivní. Opět souvisí s nahráním a výpočtem dat, kde nelze zobrazit graf, který nemá žádné data. Vše je vyobrazeno na obr. 2.3 níže, kde jsou všechny položky aktivní.



Obr. 2.3 – Položka grafy hlavního menu

2.2 RUČNÍ ZADÁVÁNÍ DAT

Uživatel programu může zadávat data ručně. Při kliku na položku Zadat data ručně viz obr. 2.2 výše. Zobrazí se dialogové okno pro zadání počtu řádku a názvu tab. dat viz obr. 2.4 níže.



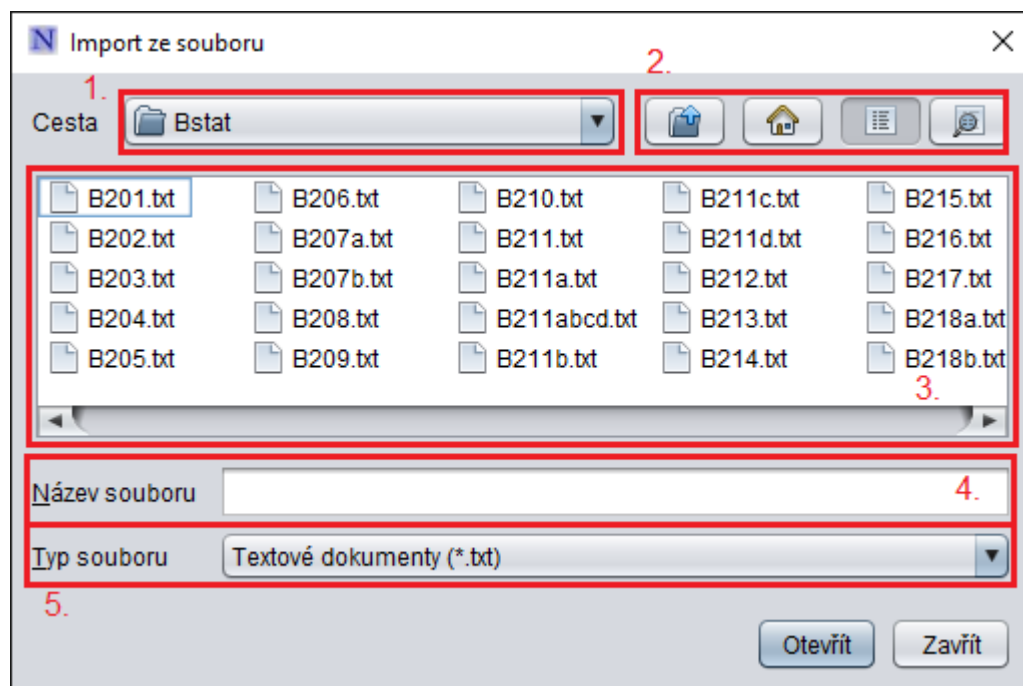
Obr. 2.4 – Dialogové okno pro vytvoření tab.

Dialogové okno pro vytvoření tab. obsahuje části vyobrazené a označené na obr. 2.4 podle bodů:

1. Obsahuje pole pro zadávání počtu řádků tab. a tlačítko na vytvoření tab.. Pole má v základu nastaveno na hodnotu 11, protože data menších výběrů, většinou menších 10, nejsou statisticky významná pro exploratorní analýzu. Při kliku na tlačítko se vytvoří tabulka s počtem řádků zadaných v poli a případně lze potvrdit pomocí klávesy ENTER při zadávání počtu řádků. Kontroluje se, zda je počet řádků větší jak deset, a také název z bodu 2.
2. Pole pro zadávání názvu výběrů dat, které se využívá jako název tab. i celé exploratorní analýzy. Uživatel si tím identifikuje, o jakou úlohu se jedná s tím, že se název také ukládá do souboru. Tím pádem má při pozdější kontrole jasnou identifikaci, o jaká data se jedná.

2.3 IMPORT DAT ZE SOUBORU

Import souboru lze provést podle položky viz obr. 2.1 nebo stiskem klávesové zkratky CTRL+O a otevře se dialogové okno pro výběr souboru, viz obr. 2.5.



Obr. 2.5 – Dialogové okno pro import dat ze souboru

Dialogové okno pro import dat ze souboru obsahuje části vyobrazené a označené na obr. 2.5 podle bodů:

1. Pomocí výběrového okna, které se klikem zobrazí, lze vybrat konkrétní složku se souborem, který se má importovat do programu. V programu je nastaven jako výchozí disk C, který by měl mít každý počítač.

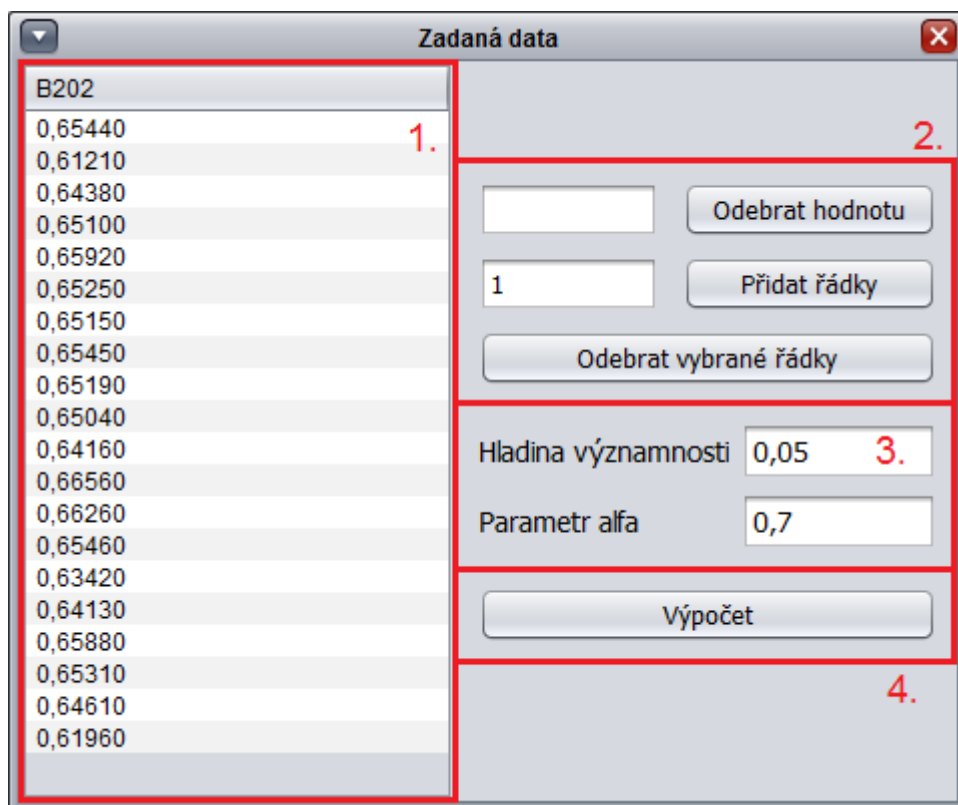
2. Ikona se složkou a šipkou obsahuje funkčnost vrácení se zpět o jednu úroveň ve struktuře složek. Na obr. 2.5 by se uživatel vrátil do složky, která by obsahovala podsložku Bstat. Ikona domečku obsahuje domovskou složku, tedy na každém počítači to je plocha. Poslední dvě ikony slouží jako přepínače s tím, že je vybrána jedna z položek. Ikona listu, která je aktuálně vybraná ovlivní výpis bodu 3. který zobrazí pouze názvy souborů a složek. V případě výběru ikony listu s lupou se zobrazí také informace o souborech nebo složkách jako jsou velikosti, datum změny nebo typ složky.
3. Zobrazuje aktuální výpis složky se soubory a soubory. Aktuálně na obr. 2.5 jsou vyfiltrovány soubory typu txt.
4. Pole pro zadávání názvu souboru, který vyhledá příslušný soubor ze složky. Název souboru je také spjat s výběrem typu souboru v bodu 5. Pokud daný soubor neexistuje, je uživatel o této skutečnosti informován. V opačném případě je soubor otevřen a data importována do programu.
5. Výběrové okno pro filtrování typů souborů v aktuální složce. Na obr. 2.5 je aktuálně vybrán typ souboru txt. Protože program pracuje pouze s daty z txt souboru, tak je v základu předvybrán tento typ. V opačném případě, pokud uživatel vybere jiný typ, tak bude informován, že nelze importovat data z jiného souboru než typu txt.

Import souboru je dále ošetřen také na formy chyb, jako jsou nevhodně zvolené znaky v souboru, kde soubor smí obsahovat pouze text, mezery, čísla a desetinnou tečku nebo čárku. Ostatní znaky jsou automaticky odmazány. Importovat lze pouze soubor, který má jeden sloupec dat, které jsou v txt souboru zadány pod sebou. Jakmile jsou importována vícerozměrná data oddělená pomocí mezer, program tento typ rozezná a informuje uživatele o této skutečnosti. Na prvním řádku souboru lze mít zadaný název sloupce, který bude importován do programu jako název sloupce. Pokud bude chtít uživatel nazvat soubor pouze číslem, tak program tuto skutečnost nezaznamená a první řádek naimportuje jako standardní číslo. V opačném případě lze v prvním řádku uvést textový řetězec i s číslem a na další řádky se již smí vypsát pouze čísla s desetinnou tečkou nebo čárkou. Pokud nebude uveden název prvního řádku, program zobrazí dialogové okno s polem pro vyplnění názvu importovaných dat, který musí být vždy vyplněn, kvůli identifikaci výběru dat.

Pokud uživatel bude chtít naimportovat prázdný soubor, tak program je schopný tuto skutečnost rozeznat a uživatele informuje. Program také dokáže rozeznat, že v souboru je uveden pouze jeden řádek s názvem sloupce, ale bez dalších řádků s číselnými daty.

2.4 TABULKA S IMPORTOVANÝMI DATY

Pro uživatele je důležité si zobrazit data, která si může importovat ze souboru nebo zadat ručně. K tomu slouží základní grafický prvek tab. s ovládacími prvky zobrazené na obr. 2.6 níže. Tabulku s importovanými daty lze zobrazit pomocí položky Editovat data v menu viz na obr. 2.2.



Obr. 2.6 – Tabulka zadaných dat s ovládacími prvky

Dialogové okno pro zadaná data obsahuje části vyobrazené a označené na obr. 2.6 podle bodů:

3. Tabulka s daty pro jednorozměrnou exploratorní analýzu dat. V hlavičce je název sloupce, který je aktuálně převzat ze souboru. Klikem nebo dvojklikem do buňky s číslem lze editovat dané číslo.
4. Panel pro editaci tab.. První nevyplněné pole je pro zadávání hodnot, které se mají smazat z tab. popsané v bodě 1. Slouží například pro odebrání odlehlých hodnoty výběru. Druhé pole s hodnotou udává, kolik řádků tab. se má přidat. Řádky lze přidat pomocí zadání hodnoty a stiskem tlačítka Přidat řádky nebo stiskem klávesy Enter se přidá další řádek do tab. v bodě 1. Pole je ošetřené na chybné zadání hodnoty. To znamená, že uživatel nesmí zadat hodnoty rovny nebo menší nule.

Tlačítkem Odebrat vybrané řádky se mažou řádky tab. v bodě 1., které funguje pouze při vybraných řádcích v tab.. Uživatel musí v tab. dat označit řádek nebo více řádků, které chce mazat a poté klikne na tlačítko Odebrat vybrané řádky. Označování buněk v tab. lze standardní způsobem, a to pomocí stisku klávesy CTRL nebo SHIFT a označováním buněk pomocí myši.

5. Pole pro zadávání Hladiny významnosti a Parametru alfa. Obě pole jsou ošetřena na zadávání dat v intervalu $\langle 0;1 \rangle$.
6. Tlačítko Výpočet provede výpočet exploratorní analýzy. Při stisku na tlačítko se také provádí kontrola, zda je zadané správné množství dat, které má nastaveno minimální počet na hodnotu 10. Menší počet dat nejsou statistiky významné a v takovém případě nelze provést exploratorní analýzu dat.

2.5 VÝPIS VÝSLEDKŮ EXPLORATORNÍ ANALÝZY

Jakmile uživatel klikne na tlačítko Vypočítat v předchozím oddíle 2.4, tak je proveden výpočet exploratorní analýzy, čímž se po výpočtu ihned zobrazí okno s výsledky celé analýzy. Tabulku s výsledky exploratorní analýzy lze zobrazit pomocí položky Výpis výsledků v menu viz na obr. 2.2.

Numerické výpočty

Ověření normality dat pomocí Exploratorní analýzy
Název úlohy: B202

Vstupní parametry a podmínky:
Počet dat: 20
Hladina významnosti: 0,05

Vstupní data:

1. 6,5440E-01	2. 6,1210E-01	3. 6,4380E-01	4. 6,5100E-01	5. 6,5920E-01
6. 6,5250E-01	7. 6,5150E-01	8. 6,5450E-01	9. 6,5190E-01	10. 6,5040E-01
11. 6,4160E-01	12. 6,6560E-01	13. 6,6260E-01	14. 6,5460E-01	15. 6,3420E-01
16. 6,4130E-01	17. 6,5880E-01	18. 6,5310E-01	19. 6,4610E-01	20. 6,1960E-01

(1) Klasické odhady parametrů

Medián:	6,5170E-01	Průměr:	6,4794E-01	Rozptyl:	1,7753E-04	3. Cent. moment:	-2,9710E-06
4. Cent. moment:	1,2576E-07	Šikmost:	-1,3564E+00	Špičatost:	4,4212E+00	8. odchylka:	1,3324E-02

(2) Kvantilové a písmenové hodnoty

Kvantilové míry:

Procento	Kvantil	Procento	Kvantil
5	6,1923E-01	10	6,3274E-01

Celkové vyhodnocení souboru: Data nemají normální rozdělení

Vyhodnocení testů

Chi-kvadrát test	nemá normální rozdělení	Test koefi. šikmosti a špičatosti	nemá normální rozdělení	Shapir a Wilk	má normální rozdělení
Anderson-Darling	nemá normální rozdělení	Kolmogorov-Smirnov	má normální rozdělení	Wilcoxon	nemá normální rozdělení

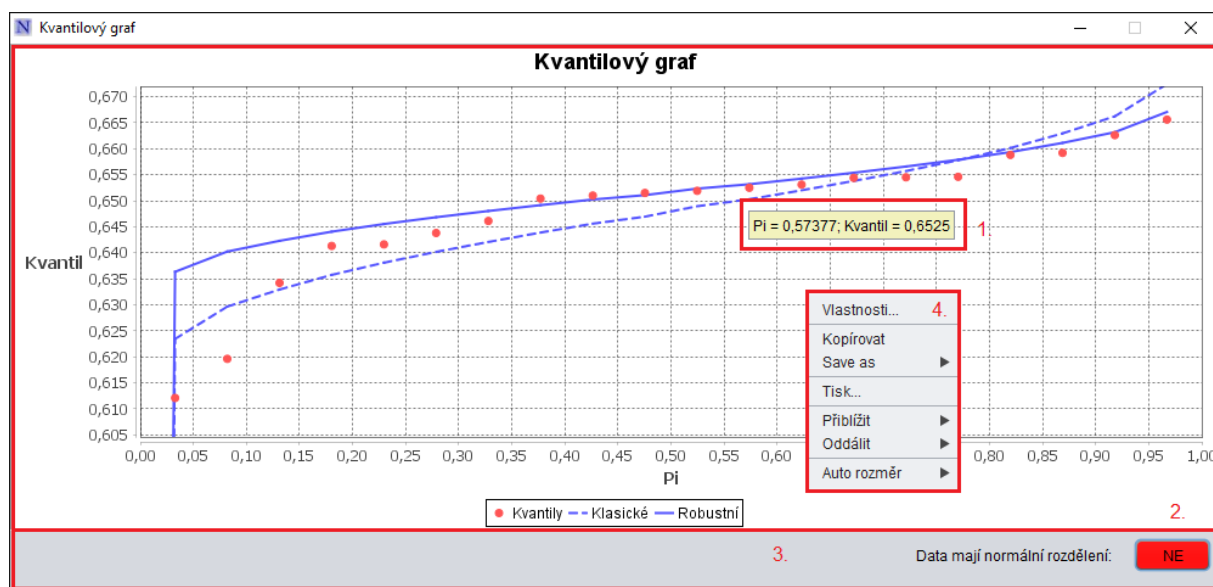
Obr. 2.7 – Výpis výsledků exploratorní analýzy

Dialogové okno pro výpis dat obsahuje části vyobrazené a označené na obr. 2.7 podle bodů:

1. Zobrazuje veškeré numerické výpočty exploratorní analýzy od mediánu, průměru až po parametrické a neparametrické testy. Vše je vyobrazeno přehledně v tabulkách.
2. Výpis hodnocení grafů, parametrických a neparametrických testů. Celkové vyhodnocení souboru je závislé na vyhodnocení testů vycházející z numerických výpočtů v bodě 1. a poté na samotném hodnocení všech grafů. Obě části se řídí Bayesovým pravidlem, které říká, že minimálně 3 testy nebo grafy mají normální rozdělení, aby se potvrdilo normální rozdělení výběru dat souboru.

2.6 ZOBRAZENÍ GRAFŮ

Všechny grafy jsou zobrazovány v samostatných oknech s možností okamžitého hodnocení, zda graf má normální rozdělení. Grafy lze zobrazit pomocí menu viz obr. 2.3. Níže na obr. 2.8 je příklad zobrazení jednoho z grafů.



Obr. 2.8 – Zobrazení kvantilového grafu

Dialogové okno pro zobrazení graf obsahuje části vyobrazené a označené na obr. 2.8 podle bodů:

1. Při najetí kurzoru myši na hodnotu kvantilu se zobrazí box s údaji na x -ové a y -ové ose. V případě obr. 2.8 se zobrazuje hodnota pravděpodobnosti P_i a kvantilů.
2. Zobrazení celého grafu názvem grafu, případně legendou popisující trendy, označením x -ové a y -ové osy.

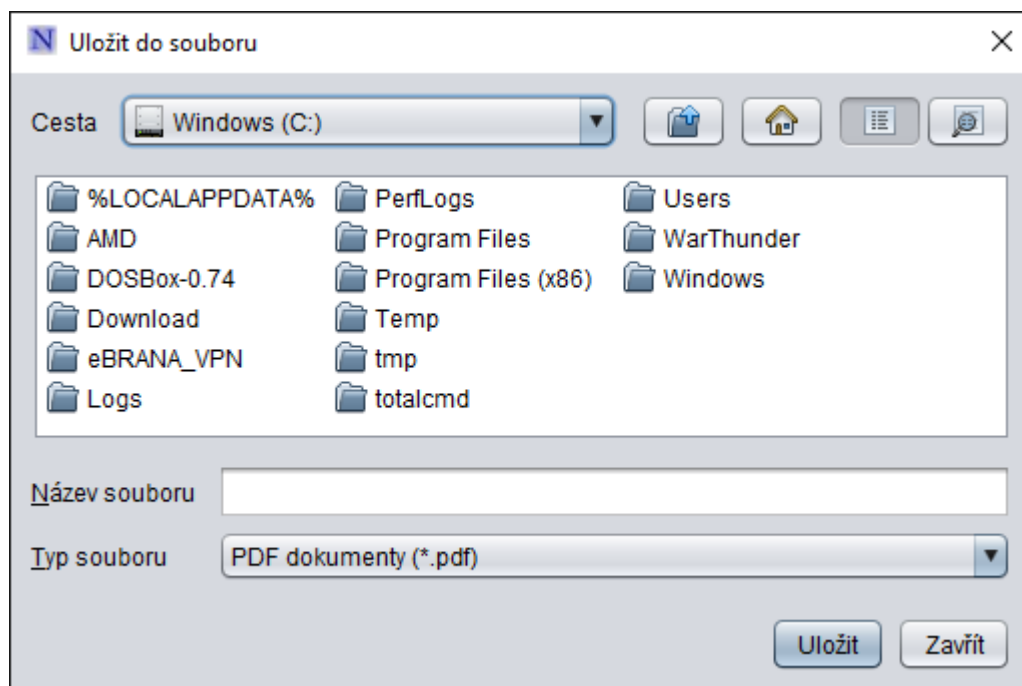
3. Panel s tlačítkem stavů, zda graf má nebo nemá normální rozdělení. Uživatel klikem na tlačítko mění stav tlačítka u grafu. Stav a hodnocení grafů jsou vypsány v tab. viz obr. 2.7 bod 2.
4. Pokud uživatel stiskem pravého tlačítka myši klikne do těla grafu, zobrazí se uvedená nabídka na obr. 2.8. V nabídce jsou uvedeny položky:
 - Vlastnosti – zobrazí se nové dialogové okno s nastavením daného grafu, pokud si uživatel potřebuje upravit vzhled grafu podle vlastních potřeb. Ve vlastnostech si může ovlivnit název grafu, změnit měřítko grafu na obou osách, popisky os, změnit barvu pozadí apod.
 - Kopírovat – stiskem na uvedenou položku se do paměti počítače zkopíruje obr. celého grafu bez spodního panelu, který je možné si uložit například v malování pomocí vložení (klávesy CTRL+V).
 - Save as – slouží k uložení grafů jako obr. ve formátu PNG, SVG a PDF. Opět se ukládá pouze samotný graf bez spodního panelu hodnocení grafu.
 - Tisk – zobrazí dialogové okno nastavení vzhledu grafu při tisku. Stiskem tlačítka OK se uživatel dostane do dalšího dialogového okna s výběrem tiskárny, nastavením počtu kopií apod.
 - Přiblížit a Oddálit – samotné názvy napovídají, jaké konkrétní funkčnosti zastupuje. Ve výběru obou položek lze přiblížit nebo oddálit jednu či druhou osu a také obě osy najednou. Položky jsou nazvané jako položka Osa oblasti zastupující x -ovou osu a Osa rozsahu zastupující y -ovou osu.
 - Auto rozměr – navrácí přiblížení nebo oddálení do výchozího stavu grafu. Opět lze ovlivnit, zda auto rozměr se má nastavit pro jednu z os nebo pro obě najednou.

2.7 UKLÁDÁNÍ DAT DO SOUBORU

Ukládání výsledků nebo dat úlohy do souboru je možné v menu vyobrazené na obr. 2.1, a také pomocí klávesové zkratky CTRL+S nebo CTRL+D. Následně se zobrazí dialogové okno, viz obr. 2.9 s tím, že všechny ikony dialogového okna mají stejnou funkčnost jako funkčnosti popsané na obr. 2.4. Jenom se liší v části tlačítka Uložit, které ukládá data programu do souboru. V základu je vybrán typ souboru ve formátu PDF nebo txt, pro data úlohy, protože ukládání dat programu lze provést pouze do formátu PDF nebo txt, a tedy není nutné vypisovat jiné soubory.

Při ukládání souboru je kontrolováno, zda název souboru již existuje. V takovém případě bude uživatel informován o existujícím souboru a bude vyzván k akci, zda se má uložit do stejnojmenného souboru či nikoliv.

Do souboru jsou uložena veškerá zaznamenaná data z výpisu výsledků viz obr. 2.7 včetně hodnocení grafů a celkového hodnocení výběru dat. Vše je opět přehledně uloženo do tabulek a odděleno nadpisy.



Obr. 2.9 – Dialogové okno pro ukládání dat do souboru

3 POUŽÍVÁNÍ PROGRAMU

Ve spuštěném programu je nutné si určit, zda uživatel bude zadávat data ručně nebo importovat ze souboru. V případě importu ze souboru je nutné kliknout na položku Import dat ze souboru, která je vyobrazena na obr. 2.1. Otevře se dialogové okno, viz obr. 2.5, a uživatel si vybere soubor ve formátu txt, ve kterém má uložena data. Program je schopný rozeznat, zda první řádek obsahuje název úlohy, ale jen v případě, že první řádek obsahuje alespoň jedno písmeno. V opačném případě provede import všech číselných hodnot souboru. Je nutné importovat soubor s daty zadávanými s jedním sloupcem v jednotlivých řádcích. Dokáže rozeznat i vícerozměrná data, která však nejsou zadáním diplomové práce, a tak bude uživatel upozorněn, že importuje vícerozměrná data. Při ručním zadávání dat provede uživatel výběr položky Zadat data ručně, viz na obr. 2.2. Otevře se dialogové okno, jako je na obr. 2.4, ve kterém uživatel zadá počet řádku a název tab..

Vytvoří se tabulka s nastavením parametrů, viz obr. 2.6. Uživatel si může do tab. přidat další data pomocí ovládacího panelu pro přidání/odebrání řádků. Tabulka je editovatelná, a proto uživatel má umožněno změnit i již vložená data. V polích pro hladinu významnosti a parametru α jsou předvyplněny hodnoty, které se nejčastěji využívají při exploratorní analýze dat. Jakmile má uživatel vše připravené, může tlačítkem v tab. spustit výpočet.

Provedou se veškeré numerické výpočty a zobrazí se výpis výsledků exploratorní analýzy, které jsou přehledně vypsány do tabulek, viz obr. 2.7. Uživatel si může také procházet výsledky všech grafů, které jsou otvírány do nových oken a lze zobrazit více grafů najednou. Zobrazený graf má možnost okamžitého hodnocení grafu, zda má nebo nemá normální rozdělení. Případně uživatel při exploratorní analýze identifikuje na grafech odlehlé body, které poté vyloučí při další analýze. Editace tab. dat lze zobrazit pomocí položky Editovat data, viz obr. 2.2 a uživateli se zobrazí stejná tabulka, jako je na obr. 2.6.

Jakmile uživatel provede kompletní vyhodnocení a již chce si vše uložit, tak lze využít položky pro uložení výsledků i dat do souborů. Proto provede klik na položku Uložit výsledky do souboru na obr. 2.1 a otevře se dialogové okno pro ukládání dat do souboru, viz obr. 2.9. V případě, že chce uživatel uložit i data úlohy z obr. 2.6, provede klik na položku Uložit data úlohy zvlášť do souboru a opět se zobrazí dialogové okno pro ukládání dat do souboru, viz obr. 2.9. V dialogovém okně je umožněno ukládat data pouze do souboru typu txt. Data se uloží do jednoho sloupce včetně názvu úlohy, který je uložen na první řádek souboru. Pokud uživatel provede ruční zadání dat a název bude pouze číselný s tím, že provede uložení dat úlohy do souboru, tak při opětovném importu souboru s uloženými daty, nebude první řádek souboru importován jako název úlohy, ale importuje se jako číselná hodnota pro exploratorní analýzu.

4 OTESTOVÁNÍ PROGRAMU

Program byl podroben testování vůči renomovaným programům jako je Adstat, Statistica a QCExpert v trialové verzi. V případě testování parametrických a neparametrických testů je nutno použít testované příklady z použité literatury, protože uvedené renomované programy nedisponují kompletním seznamem všech využitých testů. Jako hlavní testovaný program je využit Adstat.

4.1 OVĚŘOVÁNÍ PEARSONOVA CHÍ-KVADRÁT PRO TEST DOBRÉ SHODY

Vypočtené hodnoty Chí-kvadrátu vycházejí ze vztahů teoretické části, kde je tento test popsán v pododdíle 1.2.1. Pro testování je použita úloha B202 pro hladinu významnosti 5 %, která je popsána níže v pododdíle 3.7.1. V tab. 4.1 níže je uveden výpis hodnot všech testovaných programů. Je patrné, že testovaný program disponuje podobnými výsledky jako u renomovaných programů.

Tab. 4.1 - Hodnoty testování Pearsonova Chí-kvadrát testu dobré shody

Program	Experimentální hodnota	Kritická hodnota	Normalita
Testovaný program	11,714	5,9915	Zamítnuta
Adstat	13,259	5,9915	Zamítnuta
QCExpert	6,065	5,9915	Zamítnuta
Statistica	12,549	neuveдено	neuveдено

4.2 OVĚŘOVÁNÍ NORMALITY PODLE SHAPIRA A WILKA

Vypočtené hodnoty testu podle Shapira a Wilka vycházejí ze vztahů teoretické části, kde je tento test popsán v pododdíle 1.2.2. Pro testování je použita úloha B202 pro hladinu významnosti 5 %, která je popsána níže v pododdíle 3.7.1. V tab. 4.2 níže je uveden výpis hodnot testovaného programu a programu Statistica. Ostatní renomované programy nedisponují testem pod Shapira a Wilka.

Tab. 4.2 - Hodnoty testování testu podle Shapira a Wilka

Program	Experimentální hodnota	Kritická hodnota	Normalita
Testovaný program	0,861	0,905	Přijata
Statistica	0,862	neuveдено	neuveдено

4.3 OVĚŘOVÁNÍ KOEFICIENTŮ ŠIKMOSTI A ŠPIČATOSTI

Vypočtené hodnoty testu koeficientů šikmosti a špičatosti vycházejí ze vztahů teoretické části, kde je tento test popsán v pododdíle 1.2.3. Pro testování je použita úloha B202 pro hladinu významnosti 5 %, která je popsána níže v pododdíle 3.7.1. V tab. 4.3 níže je uveden výpis hodnot všech testovaných programů.

Tab. 4.3 - Hodnoty testování koeficientů šikmosti a špičatosti

Program	Experimentální hodnota		Kritická hodnota		Normalita
	šikmost	špičatost	šikmost	špičatost	
Testovaný program	-1,469	2,226	0,927	$\langle -1,635; 1,349 \rangle$	Zamítnuta
Statistica	-1,469	2,226	neuveдено	neuveдено	neuveдено

4.4 OVĚŘOVÁNÍ ANDERSONOVA-DARLINGOVA TESTU

Vypočtené hodnoty Andersonova-Darlingova testu vycházejí ze vztahů teoretické části, kde je tento test popsán v pododdíle 1.2.4. Pro testování jsou použita data použita z literatury (Meloun, 2004), která jsou níže vypsána v obr. 4.1 ve červeném rámečku.

PŘÍKLAD 2.2 *Generování náhodných výběrů z pěti teoretických rozdělení*
 K ilustraci jednotlivých metod průzkumové analýzy dat bylo vybráno pět výběrů z rozdělení vyskytujících se v instrumentálních úlohách. Výběry rozsahu $n = 50$ byly generovány pro předem zvolené střední hodnoty μ a rozptyly σ^2 :

- výběr z rovnoměrného rozdělení na intervalu $[0, 1]$ s $\mu = 0.5$ a $\sigma^2 = 1/12$,
- výběr z normovaného normálního rozdělení s $\mu = 1$ a $\sigma^2 = 1$,
- výběr z Laplaceova rozdělení s $\mu = 1$ a $\sigma^2 = 2$,
- výběr z exponenciálního rozdělení $\mu = 1$ a $\sigma^2 = 1$,
- výběr z logaritmickeo-normálního rozdělení s $\mu = 2.718$ a $\sigma^2 = 47.209$.

○ *Data:* a) Výběr z rovnoměrného rozdělení $R(0.5, 0.083)$:
 0.531 0.677 0.171 0.065 0.848 0.021 0.380 0.760 0.524 0.283 0.841 0.631 0.645 0.567 0.594 0.141
 0.994 0.998 0.211 0.487 0.595 0.751 0.231 0.012 0.487 0.794 0.358 0.823 0.414 0.087 0.147 0.559
 0.053 0.217 0.385 0.755 0.853 0.707 0.266 0.878 0.040 0.407 0.839 0.171 0.325 0.295 0.842 0.636
 0.172 0.924;

b) Výběr z normálního rozdělení $N(0, 1)$:
 -1.008 -0.500 0.749 1.723 0.076 0.569 -1.389 0.087 1.112 -0.235 0.519 0.279 -0.758 -0.588 -0.594
 -0.885 -0.072 1.980 0.063 0.016 -0.673 -0.993 0.752 0.092 0.236 -2.962 0.109 -1.285 0.634 -0.383
 1.134 -0.711 -1.825 2.374 0.500 -1.380 0.046 -0.544 -0.150 -1.129 1.173 1.401 -2.121 0.521 0.280
 1.440 -0.415 -0.443 -0.384 0.690;

Obr. 4.1 – Hodnoty výběru pro testování Andersonova-Darlingova testu

Jak je patrné z použitých dat, jedná se o výběr z normálního rozdělení a logicky musí mít tento test potvrzenou normalitu. Testování se provádí při hladině významnosti 5 %. V tab. 4.4 je uveden výpis hodnot testovaného programu vůči výsledkům dat z literatury (Meloun, 2004, s. 132). Je patrný rozdíl v experimentální hodnotě, což je zapříčiněno nutností zaokrouhlovat data pro hledání hodnoty normovaného normálního rozdělení, kde dostupné hodnoty jsou tabelovány na 2 desetinná místa. Při 2 desetinných místech se zdají být hodnoty shodné, což způsobuje uvedenou chybovost.

Tab. 4.4 - Hodnoty testování Andersonova-Darlingova testu

Program	Experimentální hodnota	Kritická hodnota	Normalita
Testovaný program	0,175	0,739	Přijata
Adstat	0,24	neuvedeno	Přijata

4.5 OVĚŘOVÁNÍ WILCOXONOVA POŘADOVÉHO TESTU

Vypočtené hodnoty Wilcoxonova pořadového testu vycházejí ze vztahů teoretické části, kde je tento test popsán v pododdíle 1.2.5. Pro testování jsou použita data z literatury (Hendl, 2012, s. 82), které jsou zobrazeny na obr. 4.2 níže pro hladinu významnosti 5 %.

Tab. 6.7 Příklad provádění Wilcoxonova testu střední hodnoty

Délka skoku x_j [m]	3,75	3,75	3,9	3,65	3,65	3,65	3,6	4,05	4,05	3,5	3,45	4,2	3,3
$x_j - \bar{\mu}_0$	-0,05	-0,05	0,1	-0,15	-0,15	-0,15	-0,2	0,25	0,25	-0,3	-0,35	0,4	-0,5
Seřazené $ x_j - \bar{\mu}_0 $	0,05	0,05	0,1	0,15	0,15	0,15	0,2	0,25	0,25	0,3	0,35	0,4	0,5
Pořadí	1,5	1,5	3	5	5	5	7	8,5	8,5	10	11	12	13

Obr. 4.2 – Tabulka s hodnotami pro příklad provádění Wilcoxonova testu

Data na výše uvedeném obr. 4.2 vykazují jinou hodnotu průměru, protože se jedná o výtah dat z celkové tab. literatury (Hendl, 2012, s. 82). Proto se v programu musí nastavit průměr na hodnotu 3,8 m viz příklad 6.11 literatury (Hendl, 2012, s. 234) a také se musí upravit podmínka testu. A to způsobem, že není brán v potaz počet dat výběru, aby se mohla otestovat testovací charakteristika ze vztahu 1.35. V tab. 4.5 níže je uveden výpis hodnot testovaného programu vůči výsledkům dat z literatury (Hendl, 2012, s. 234).

Tab. 4.5 - Hodnoty testování Wilcoxonova testu

Program	Experimentální hodnota	Kritická hodnota	Normalita
Testovaný program	0,943	1,96	Zamítnut
Příklad z literatury (Hendl, 2012, s. 234)	0,943	1,96	Zamítnut

4.6 OVĚŘOVÁNÍ TESTU PODLE KOLMOGOROVA-SMIRNOVA

Vypočtené hodnoty testu podle Kolmogorova-Smirnova vycházejí ze vztahů teoretické části, kde je tento test popsán v pododdíle 1.2.6. Pro testování jsou použita data z literatury (Hendl, 2012, s. 243), které jsou zobrazeny na obr. 4.3 níže pro hladinu významnosti 5 %.

Tab. 6.14 Příklad postupu při provádění Kolmogorova-Smirnova testu normality

x_i	n_i	$Kum n_i$	$\hat{F}(x)$	z	$F_0(x)$	$ D $
50	1	1	0,1111	-2,33	0,0099	0,1012
68	1	2	0,2222	-1,20	0,1151	0,1071
77	1	3	0,3333	-0,53	0,2981	0,0796
81	1	4	0,4444	-0,67	0,2514	0,0352
83	1	5	0,5556	-0,13	0,4483	0,1073
84	1	6	0,6667	-0,07	0,4721	0,1946
91	1	7	0,7778	0,40	0,6554	0,1224
92	1	8	0,8889	0,46	0,6772	0,2117
94	1	9	1,0000	0,60	0,7257	0,2743

Obr. 4.3 – Tabulka s hodnotami pro příklad postupu testu podle Kolmogorova-Smirnova

Data na obr. 4.3 jsou seřazena vzestupně viz první sloupec x_i , což je důležité pro sestavení empirické distribuční funkce. Výběr dat obsahuje 9 měření, kde statisticky významná data jsou od 10 hodnot výběru a výše, což v tomto ohledu daný testovaný výběr nesplňuje. Avšak pro otestování podle dat v obr. 4.3 se musí program upravit, aby provedl výpočet dat i pro tak malý výběr.

Tabulka na obr. 4.3 ukazuje postup řešení testu podle Kolmogorova-Smirnova, kde je vyznačena největší hodnota testovacího kritéria D vycházejícího ze vztahu (1.36). V tab. 4.6 níže je uveden výpis hodnot testovaného programu vůči výsledkům dat z literatury (Hendl, 2012, s. 244).

Tab. 4.6 - Hodnoty testování testu podle Kolmogorova-Smirnova

Program	Experimentální hodnota	Kritická hodnota	Normalita
Testovaný program	0,274	0,43	Přijata
Příklad z literatury (Hendl, 2012, s. 244)	0,274	0,43	Přijata

4.7 TESTOVÁNÍ PROGRAMU NA PŘEDEM STANOVENÝCH ÚLOHÁCH

Testování samotného programu probíhá za pomoci již vytvořených úloh z literatury (Meloun, 2002), které jsou porovnávány s výsledky z renomovaných programů jako je Adstat a QCExpert. Hlavním porovnávací program je Adstat, který je celkovou předlohou pro realizaci programu. Data všech úloh se porovnávají s výsledky z programů s tím, že se nehodnotí data výběrů souborů, zda mají nebo nemají normální rozdělení, a také se neodpovídá na otázky v úlohách.

4.7.1 Úloha B2.02

Úloha je zaměřena na symetrii rozdělení obsahu účinné látky v tabletě. Celé zadání úlohy níže v obr. 4.4 i s tabulkou použitých dat.

Úloha B2.02 Symetrie rozdělení obsahu účinné látky v tabletě

V jistém přípravku byl stanoven obsah účinné látky A v mg na tabletu. Zkonstruuje bariérově-číslíkové schéma formou sedmipísmenového zápisu výběru a rozhodněte o typu rozdělení. Určete kvantilové parametry polohy a rozptýlení a prověřte předpoklady o výběru, kladené na reprezentativní náhodný výběr. Je rozdělení výběru symetrické? Obsahuje výběr nějaké odlehlé body? Určete, jaká je hloubka dolního a horního kvartilu?

Data: Obsah účinné látky A [mg/tbl]:

0.6544	0.6121	0.6438	0.6510	0.6592	0.6525	0.6515	0.6545	0.6519	0.6504
0.6416	0.6656	0.6626	0.6546	0.6342	0.6413	0.6588	0.6531	0.6461	0.6196

Obr. 4.4 – Zadání úlohy B2.02

U všech programů se testují pouze numerické výsledky, které jsou vypsány do jednotlivých tabulek, a též se porovnávají grafy pro testování normality dat. Vše je testováno na hladině významnosti 5 %.

Tab. 4.7 – Klasické odhady parametrů úlohy B2.02

	Testovaný program	Adstat	QCExpert
Medián	$6,517 \cdot 10^{-1}$	$6,517 \cdot 10^{-1}$	$6,517 \cdot 10^{-1}$
Průměr	$6,479 \cdot 10^{-1}$	$6,479 \cdot 10^{-1}$	$6,479 \cdot 10^{-1}$
Rozptyl	$1,775 \cdot 10^{-4}$	$1,775 \cdot 10^{-4}$	$1,775 \cdot 10^{-4}$
Šikmost	-1,356	-1,356	-1,356
Špičatost	4,421	4,421	4,421
Směrodatná odchylka	$1,332 \cdot 10^{-2}$	$1,332 \cdot 10^{-2}$	$1,332 \cdot 10^{-2}$
3. centrální moment	$-2,971 \cdot 10^{-6}$	$-3,127 \cdot 10^{-6}$	neuveďeno
4. centrální moment	$1,258 \cdot 10^{-7}$	$1,324 \cdot 10^{-7}$	neuveďeno

V tab. 4.7 je patrné, že většina testovaných parametrů mají stejné hodnoty s porovnávanými programy. Rozdíl lze najít pouze u třetího a čtvrtého centrálního momentu s tím, že je obtížné určit z jakých vztahů vychází Adstat, ale testovaný program vychází ze vztahu (1.6). Program QCExpert nedisponuje výpisem hodnot obou momentů.

Tab. 4.8 – Písmenové hodnoty úlohy B2.02

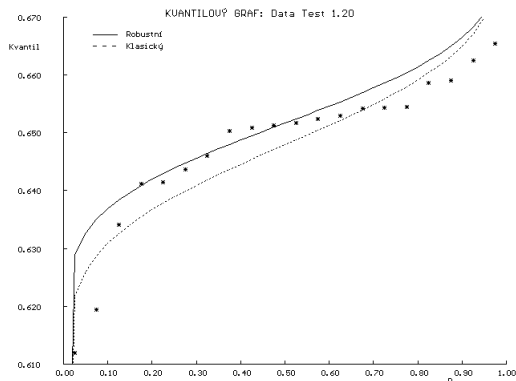
Kvantily	Meze	Testovaný program	Adstat
Sedecil	Spodní	$6,226 \cdot 10^{-1}$	$6,223 \cdot 10^{-1}$
	Horní	$6,620 \cdot 10^{-1}$	$6,620 \cdot 10^{-1}$
Oktil	Spodní	$6,365 \cdot 10^{-1}$	$6,369 \cdot 10^{-1}$
	Horní	$6,592 \cdot 10^{-1}$	$6,591 \cdot 10^{-1}$
Kvartil	Spodní	$6,433 \cdot 10^{-1}$	$6,433 \cdot 10^{-1}$
	Horní	$6,545 \cdot 10^{-1}$	$6,545 \cdot 10^{-1}$

Tab. 4.9 – Kvantilové míry úlohy B2.02

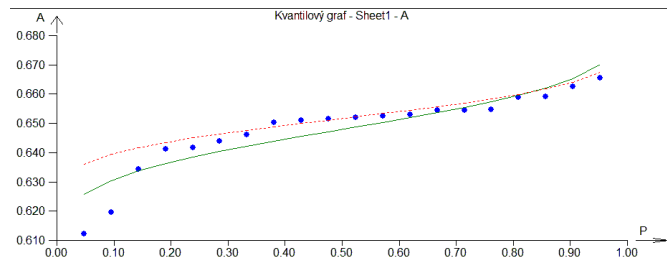
Kvantily	Kvantily	Testovaný program	Adstat
Rozsah	Sedecil	$3,934 \cdot 10^{-2}$	$3,963 \cdot 10^{-2}$
	Oktil	$2,271 \cdot 10^{-2}$	$2,219 \cdot 10^{-2}$
	Kvartil	$1,128 \cdot 10^{-2}$	$1,128 \cdot 10^{-2}$
Polosuma	Sedecil	$6,423 \cdot 10^{-1}$	$6,422 \cdot 10^{-1}$
	Oktil	$6,478 \cdot 10^{-1}$	$6,480 \cdot 10^{-1}$
	Kvartil	$6,489 \cdot 10^{-1}$	$6,489 \cdot 10^{-1}$
Délka konců	Sedecil	1,250	1,257
	Oktil	$7,003 \cdot 10^{-1}$	$6,769 \cdot 10^{-1}$
	Kvartil	0	0
Šikmost	Sedecil	$2,395 \cdot 10^{-1}$	$1,612 \cdot 10^{+1}$
	Oktil	$1,678 \cdot 10^{-1}$	$2,894 \cdot 10^{+1}$
	Kvartil	$2,494 \cdot 10^{-1}$	$5,706 \cdot 10^{+1}$
Pseudosigma	Sedecil	$1,286 \cdot 10^{-2}$	$1,295 \cdot 10^{-2}$
	Oktil	$9,875 \cdot 10^{-3}$	$9,647 \cdot 10^{-3}$
	Kvartil	$8,364 \cdot 10^{-3}$	$8,364 \cdot 10^{-3}$

Tab. 4.8 a 4.9 obsahují data pouze testovaného programu a programu Adstat. QCExpert ani Statistica nemají výpočty písmenných hodnot a kvantilových mír. Testovaný program vychází ze vztahu 1.4 a vztahů uvedených v tab. 1.1. Není tedy známo, z jakých vzorců vychází program Adstat, ale výsledné hodnoty obou programů se liší v minimálních hodnotách. V případě kvantilové míry šikmosti je rozdíl větší, ale pro testovaný program se využil ověřený vztah v tab. 1.1 převzatý z literatury (Meloun, 2004).

Níže jsou vypsány všechny grafy testovaného programu, který se porovnává s renomovanými programy Adstat a QCExpert.

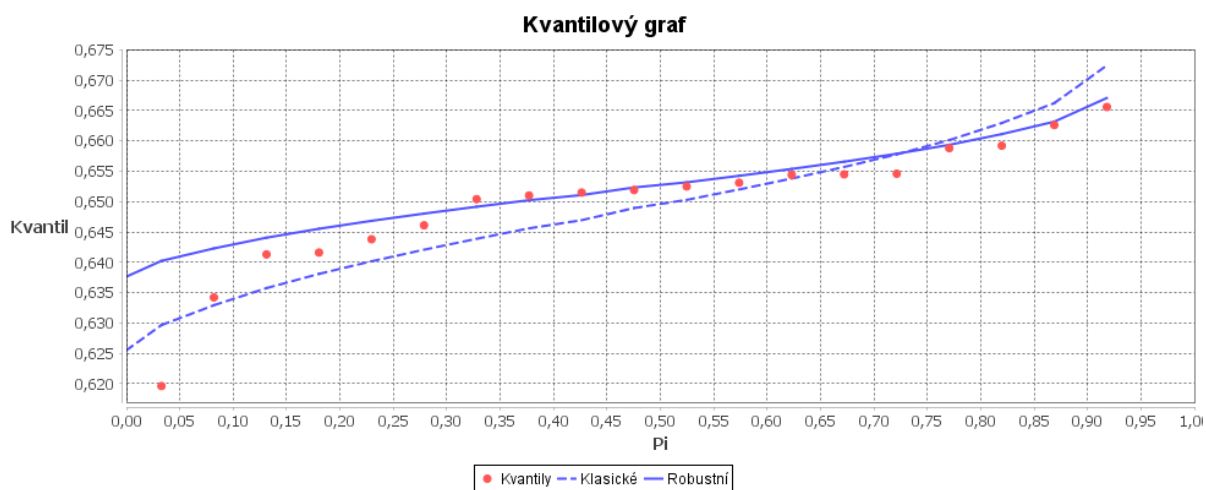


Obr. 4.5 – Kvantilový graf úlohy B2.02 Adstat

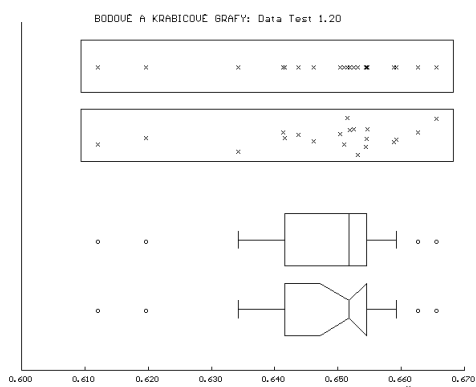


Obr. 4.6 – Kvantilový graf úlohy B2.02

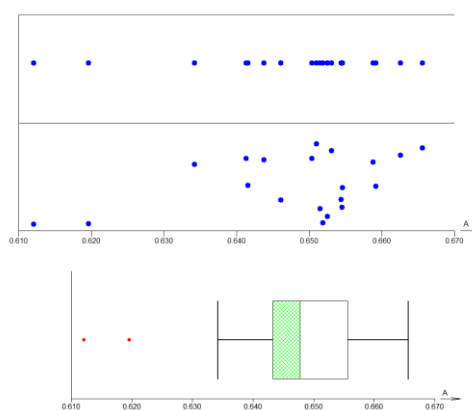
QCExpert



Obr. 4.7 – Kvantilový graf úlohy B2.02 testovaného programu

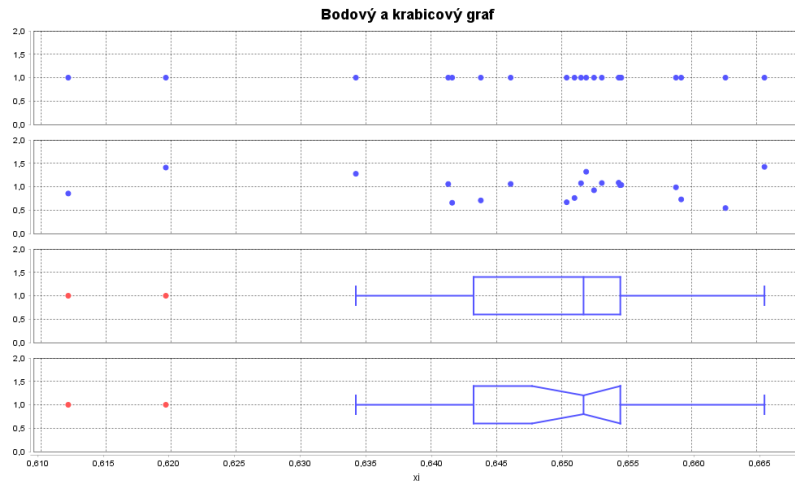


Obr. 4.8 – Bodový a krabicový graf úlohy B2.02 Adstat

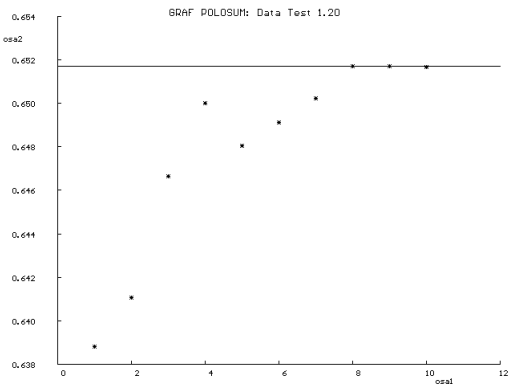


Obr. 4.9 – Bodový a krabicový graf úlohy B2.02

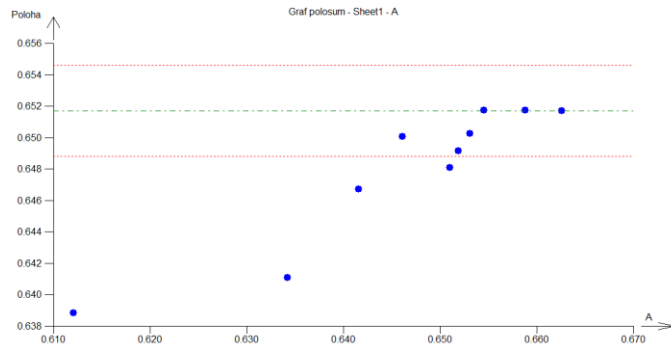
QCExpert



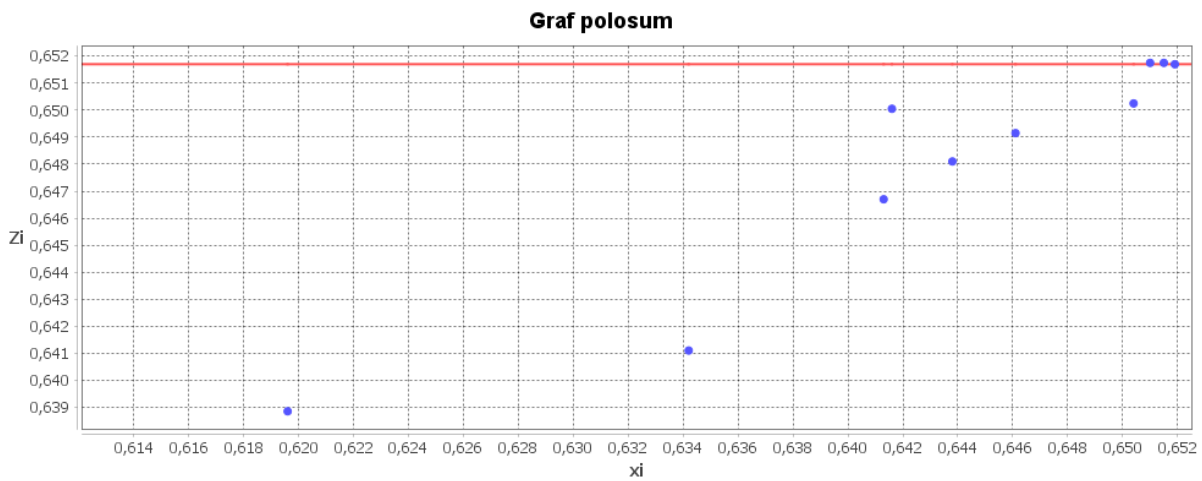
Obr. 4.10 – Bodový a krabicový graf úlohy B2.02 testovaného programu



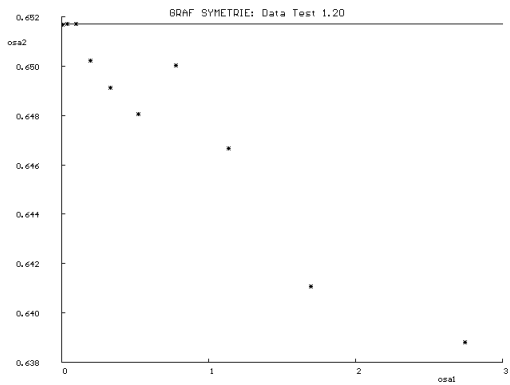
Obr. 4.11 – Graf polosum úlohy B2.02 Adstat



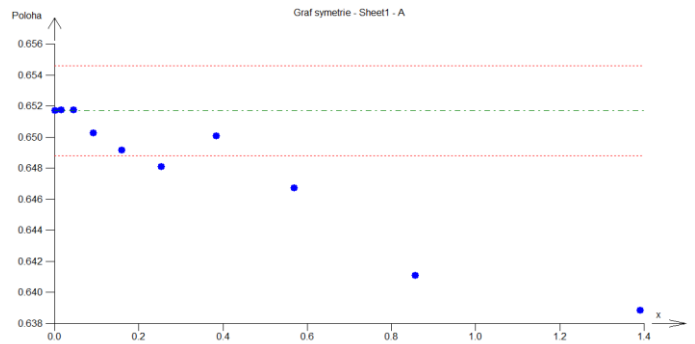
Obr. 4.12 – Graf polosum úlohy B2.02 QCExpert



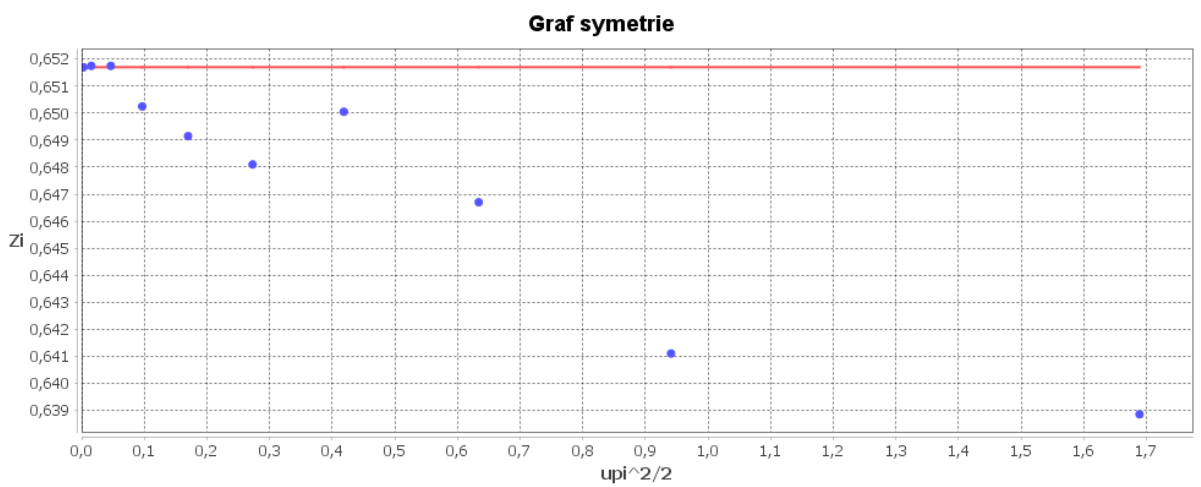
Obr. 4.13 – Graf polosum úlohy B2.02 testovaného programu



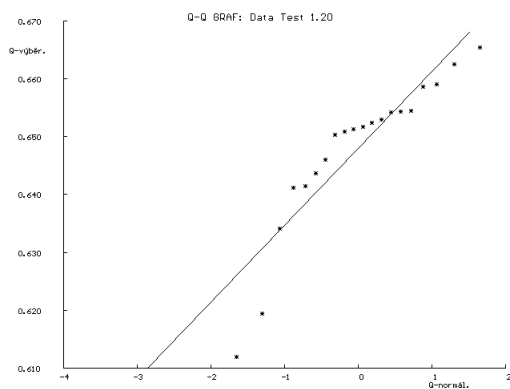
Obr. 4.14 – Graf symetrie úlohy B2.02 Adstat



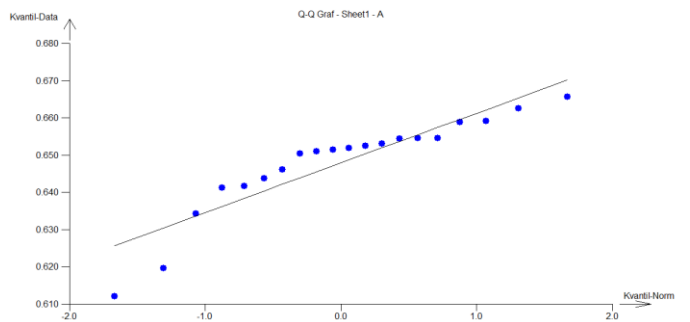
Obr. 4.15 – Graf symetrie úlohy B2.02 QCExpert



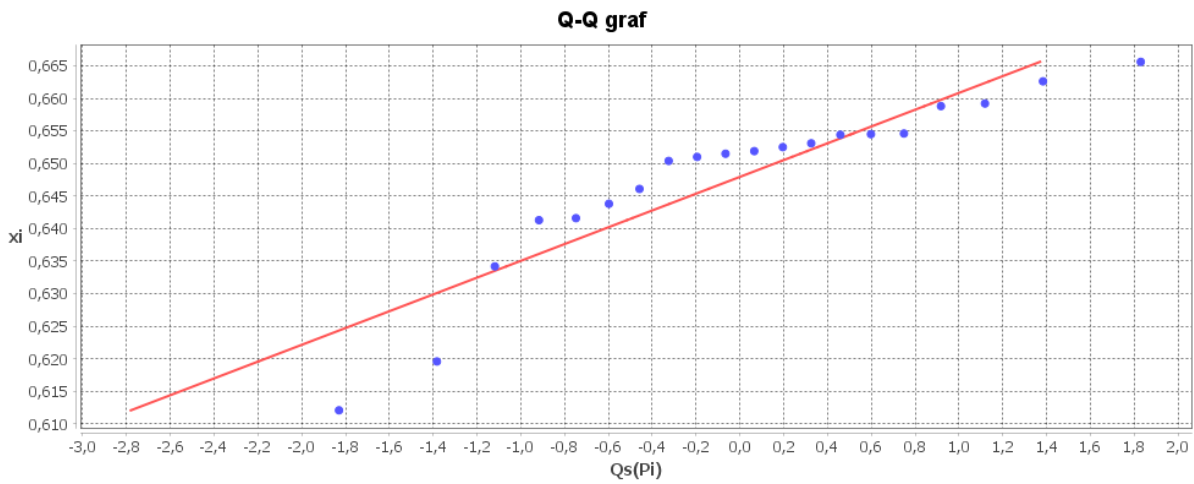
Obr. 4.16 – Graf symetrie úlohy B2.02 testovaného programu



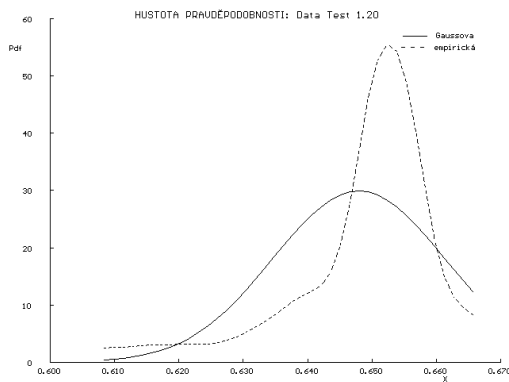
Obr. 4.17 – Q-Q graf úlohy B2.02 Adstat



Obr. 4.18 – Q-Q graf úlohy B2.02 QCExpert

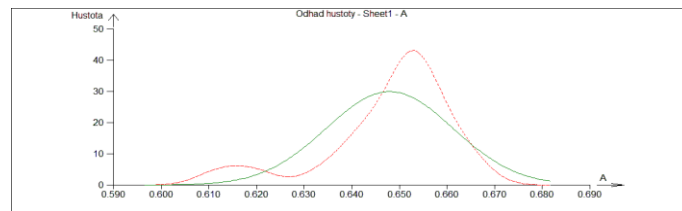


Obr. 4.19 – Q-Q graf úlohy B2.02 testovaného programu

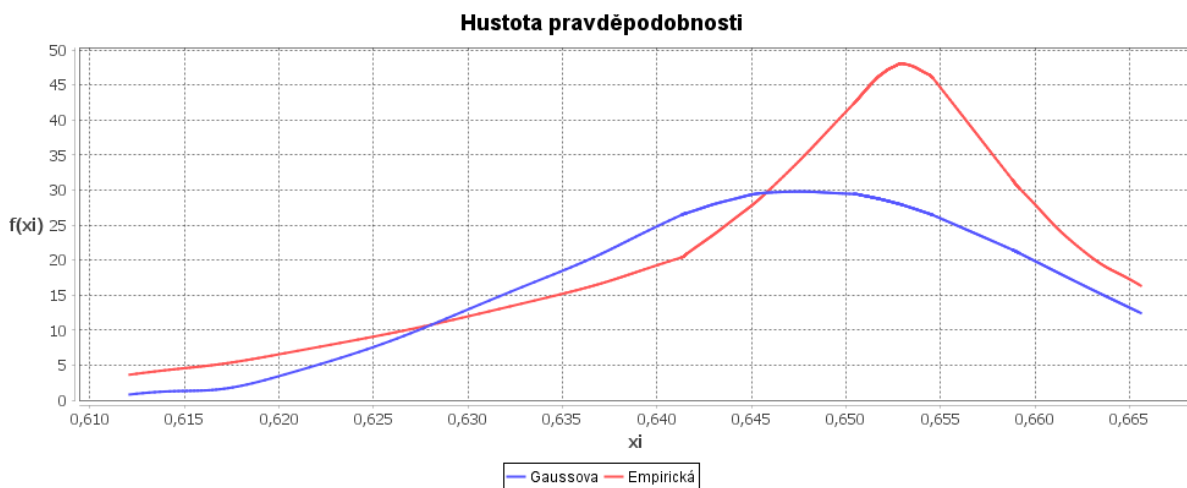


Obr. 4.20 – Graf hustoty pravděpodobnosti úlohy B2.02

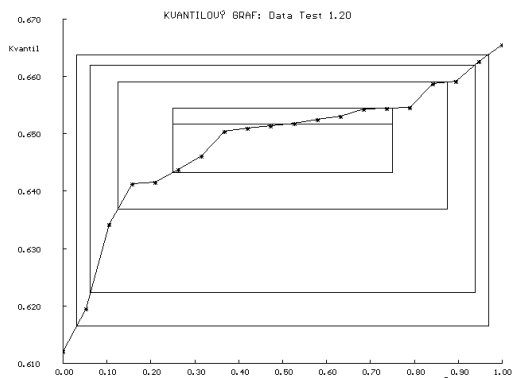
Adstat



Obr. 4.21 – Graf hustoty pravděpodobnosti úlohy B2.02 QCExpert

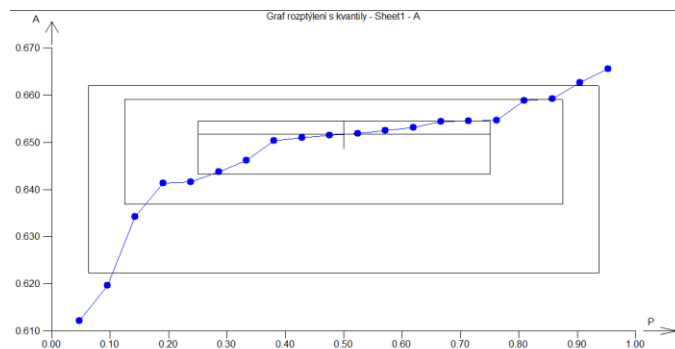


Obr. 4.22 – Graf hustoty pravděpodobnosti úlohy B2.02 testovaného programu



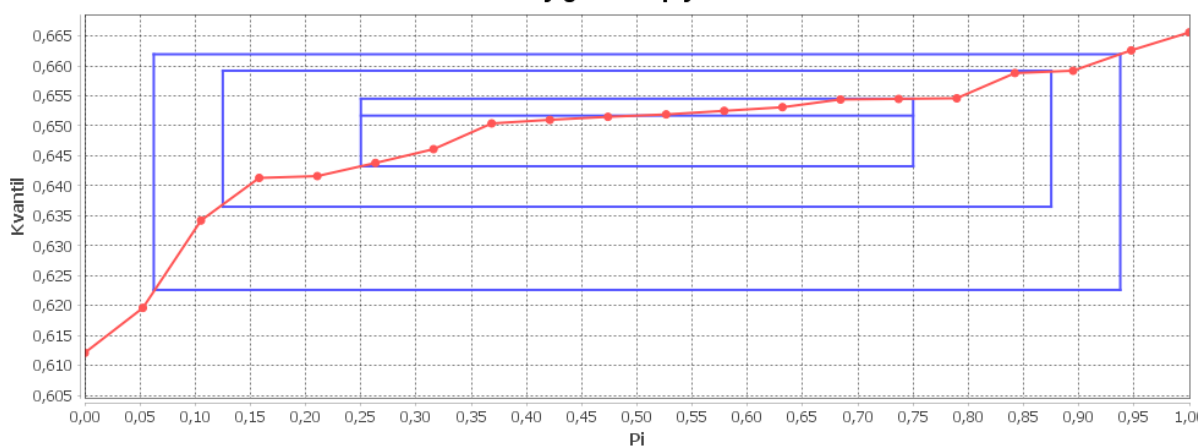
Obr. 4.23 – Q-Q graf úlohy B2.02

Adstat



Obr. 4.24 – Q-Q graf úlohy B2.02 QCExpert

Kvantilový graf s rozptýlením



Obr. 4.25 – Kvantilový graf rozptýlení úlohy B2.02 testovaného programu

Všechny výše uvedené grafy testovaného programu se shodují s výsledky s programů Adstat a QCExpert. Na kvantilových grafech je patrné, že jediný Adstat se odlišuje ve spojnici robustních dat, kde je logické, že robustní data jsou méně náchylná na vybočující body, a tedy by se měla správně více přimykát k bodům. Toto tvrzení potvrzují grafy testovaného programu i QCExpert. V případě krabicového grafu je patrné, že Adstat chybně zobrazuje přilehlé body vnitřních hradeb viz obr. 1.1, kde přilehlé body jsou sedecily nikoliv 1,5 násobku kvartilů, který označuje vnitřní hradby. Opět je patrné, že testovaný program se shoduje s renomovaným programem QCExpert. Hustota pravděpodobnosti má totožné výsledky mezi testovaným programem a QCExpertem, ale program Adstat se odlišuje, protože spojnice trendu hustoty pravděpodobnosti má vyšší hodnoty. Ostatní grafy se vzájemně shodují nebo mají minimální odlišnosti.

4.7.2 Úloha B2.12

Úloha je zaměřena na vyšetření rozdělení přírůstku hmotnosti chovných jalovic. Celé zadání úlohy níže v obr. 4.26 i s tabulkou použitých dat.

Úloha B2.12 *Vyšetření rozdělení přírůstku hmotnosti chovných jalovic*

Při vyhodnocování výsledků nové metody chovu jalovic byly zjištěny přírůstky hmotnosti v kg za určité období u 100 jalovic. Komentujte rozdělení výběru a rozhodněte, který kvantilový odhad výběrového parametru polohy a rozptýlení zde bude nejlepší. Vyšetřete statistické zvláštnosti (stupeň symetrie, špičatosti, lokální koncentrace dat, vybočující hodnoty)? Vyšetřete tvar a symetrii rozdělení na základě grafu polosum, symetrie, špičatosti a diferencního kvantilového grafu. Kolik procent jalovic dosáhlo maximálního přírůstku 80 kg?

Data: Přírůstky hmotnosti jalovic [kg]:

59.8	61.6	62.1	62.3	62.8	65.2	65.4	65.7	66.7	67.4
..
92.8	93.7	94.2	94.7	94.9	96.1	96.9	98.1	99.1	100.4

Obr. 4.26 – Zadání úlohy B2.12

U všech programů se testují pouze numerické výsledky, které jsou vypsány do jednotlivých tabulek, a též se porovnávají grafy pro testování normality dat. Vše je testováno na hladině významnosti 5 %.

Tab. 4.10 – Klasické odhady parametrů úlohy B2.12

	Testovaný program	Adstat	QCExpert
Medián	$7,980 \cdot 10^{+1}$	$7,980 \cdot 10^{+1}$	$7,980 \cdot 10^{+1}$
Průměr	$7,999 \cdot 10^{+1}$	$7,999 \cdot 10^{+1}$	$7,999 \cdot 10^{+1}$
Rozptyl	$8,351 \cdot 10^{+1}$	$8,351 \cdot 10^{+1}$	$8,351 \cdot 10^{+1}$
Šikmost	$7,500 \cdot 10^{-3}$	$7,506 \cdot 10^{-3}$	$7,506 \cdot 10^{-3}$
Špičatost	2,579	2,579	2,579
Směrodatná odchylka	9,139	9,139	9,139
3. centrální moment	5,639	5,700	neuvedeno
4. centrální moment	$1,763 \cdot 10^{+4}$	$1,781 \cdot 10^{+4}$	neuvedeno

V tab. 4.7 je patrné, že většina testovaných parametrů mají stejné hodnoty s porovnávanými programy. Rozdíl lze najít opět u třetího a čtvrtého centrálního momentu jako úlohy B2.02. U šikmosti je u testovaného programu rozdílná hodnota o 0,0008 %, což je velmi zanedbatelná chybovost.

Tab. 4.11 – Písmenové hodnoty úlohy B2.12

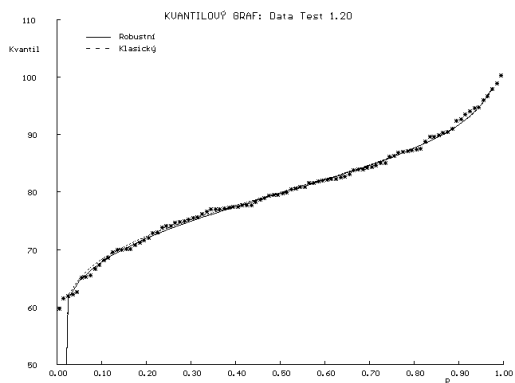
Kvantily	Meze	Testovaný program	Adstat
Sedecil	Spodní	$6,586 \cdot 10^{+1}$	$6,546 \cdot 10^{+1}$
	Horní	$9,435 \cdot 10^{+1}$	$9,461 \cdot 10^{+1}$
Oktil	Spodní	$6,919 \cdot 10^{+1}$	$6,985 \cdot 10^{+1}$
	Horní	$9,118 \cdot 10^{+1}$	$9,056 \cdot 10^{+1}$
Kvartil	Spodní	$7,430 \cdot 10^{+1}$	$7,430 \cdot 10^{+1}$
	Horní	$8,633 \cdot 10^{+1}$	$8,633 \cdot 10^{+1}$

Tab. 4.12 – Kvantilové míry úlohy B2.12

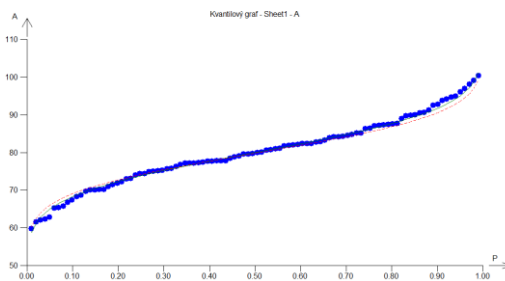
Kvantily	Kvantily	Testovaný program	Adstat
Rozsah	Sedecil	$2,849 \cdot 10^{+1}$	$2,915 \cdot 10^{+1}$
	Oktil	$2,198 \cdot 10^{+1}$	$2,071 \cdot 10^{+1}$
	Kvartil	$1,203 \cdot 10^{+1}$	$1,203 \cdot 10^{+1}$
Polosuma	Sedecil	$8,011 \cdot 10^{+1}$	$8,003 \cdot 10^{+1}$
	Oktil	$8,019 \cdot 10^{+1}$	$8,021 \cdot 10^{+1}$
	Kvartil	$8,031 \cdot 10^{+1}$	$8,031 \cdot 10^{+1}$
Délka konců	Sedecil	$8,626 \cdot 10^{-1}$	$8,855 \cdot 10^{-1}$
	Oktil	$6,033 \cdot 10^{-1}$	$5,438 \cdot 10^{-1}$
	Kvartil	0	0
Šikmost	Sedecil	$-1,079 \cdot 10^{-2}$	2,411
	Oktil	$-1,768 \cdot 10^{-2}$	3,418
	Kvartil	$-4,262 \cdot 10^{-2}$	5,894
Pseudosigma	Sedecil	9,311	9,526
	Oktil	9,558	9,005
	Kvartil	8,921	8,921

Tab. 4.11 a 4.12 obsahují data pouze testovaného programu a programu Adstat. QCExpert ani Statistica nemají výpočty písmenných hodnot a kvantilových mír. Testovaný program vychází ze vztahu 1.4 a vztahů uvedených v tab. 1.1. Není tedy známo, z jakých vzorců vychází program Adstat, ale výsledné hodnoty obou programů se liší v minimálních hodnotách. V případě šikmosti je rozdíl větší, ale pro testovaný program se využil ověřený vztah v tab. 1.1 převzatý z literatury (Meloun, 2004).

Níže jsou vypsány všechny grafy testovaného programu, který se porovnává s programy Adstat a QCExpert.

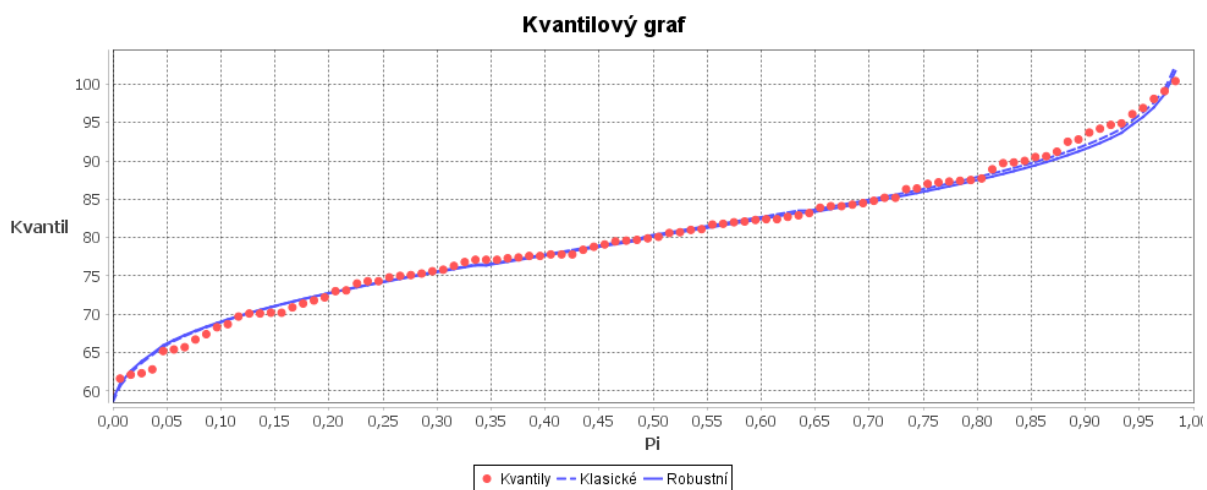


Obr. 4.27 – Kvantilový graf úlohy B2.12 Adstat

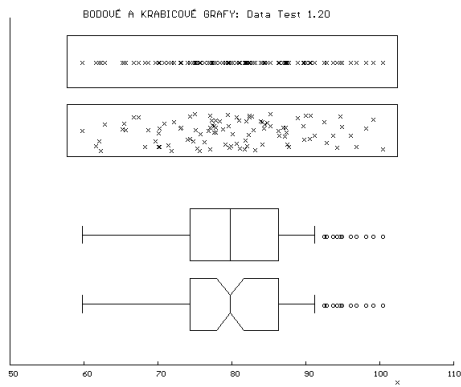


Obr. 4.28 – Kvantilový graf úlohy B2.12

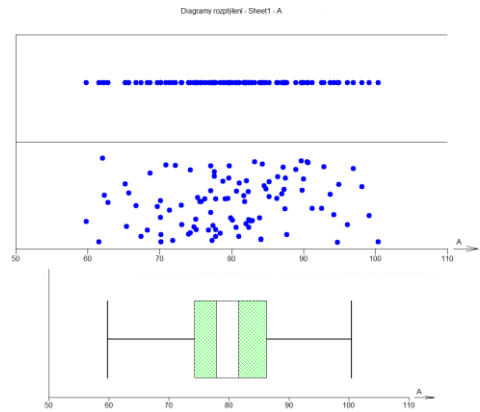
QCExpert



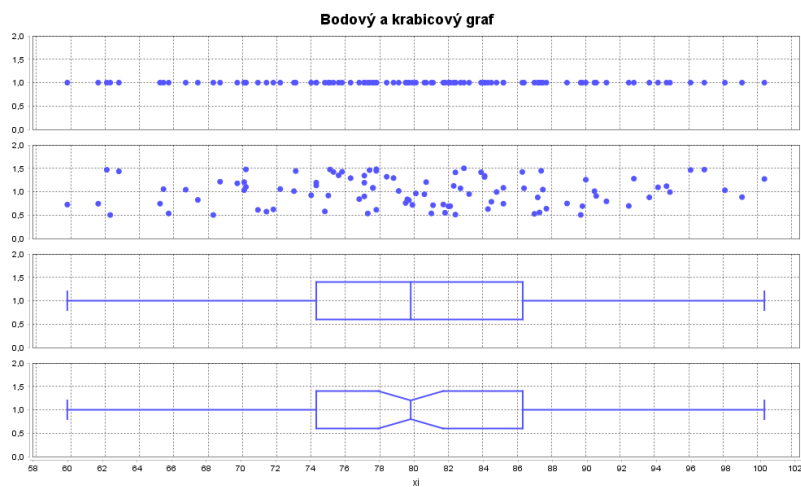
Obr. 4.29 – Kvantilový graf úlohy B2.12 testovaného programu



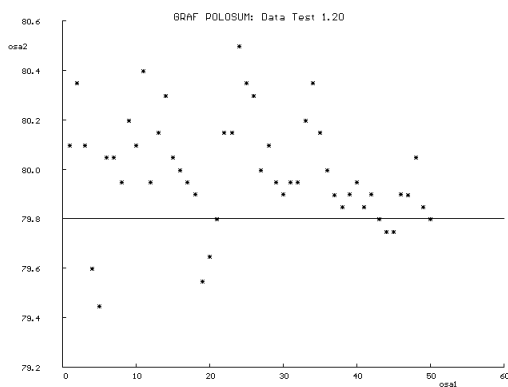
Obr. 4.30 – Bodový a krabicový graf úlohy B2.12 Adstat



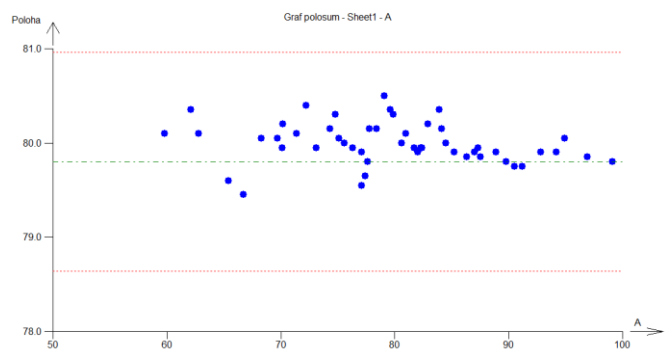
Obr. 4.31 – Bodový a krabicový graf úlohy B2.12 QCExpert



Obr. 4.32 – Bodový a krabicový graf úlohy B2.12 testovaného programu

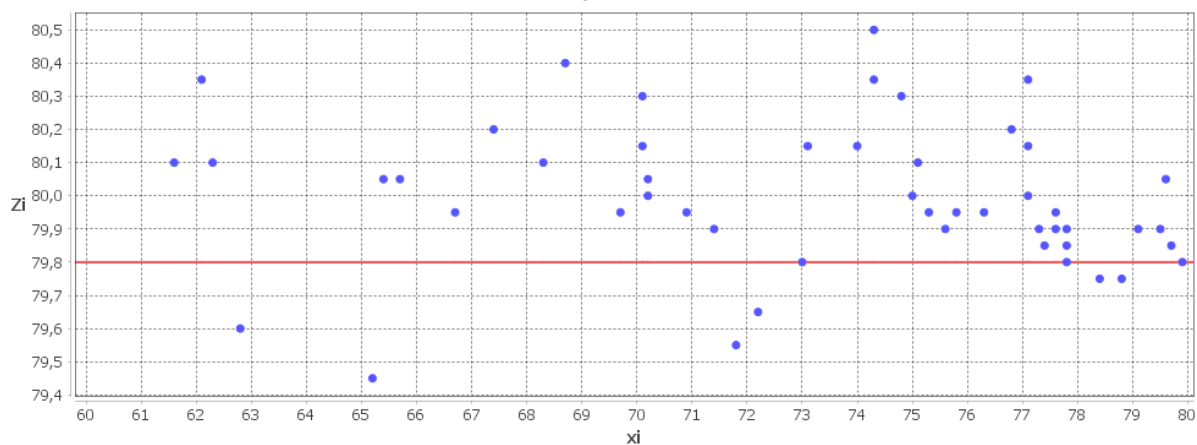


Obr. 4.33 – Graf polosum úlohy B2.12 Adstat

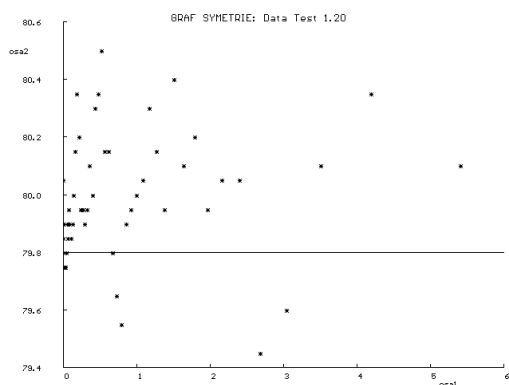


Obr. 4.34 – Graf polosum úlohy B2.12 QCExpert

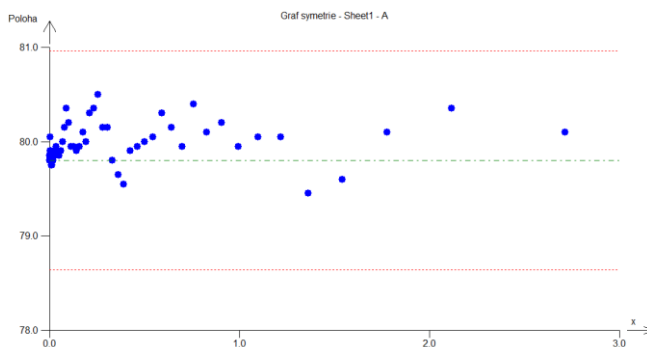
Graf polosum



Obr. 4.35 – Graf polosum úlohy B2.12 testovaného programu

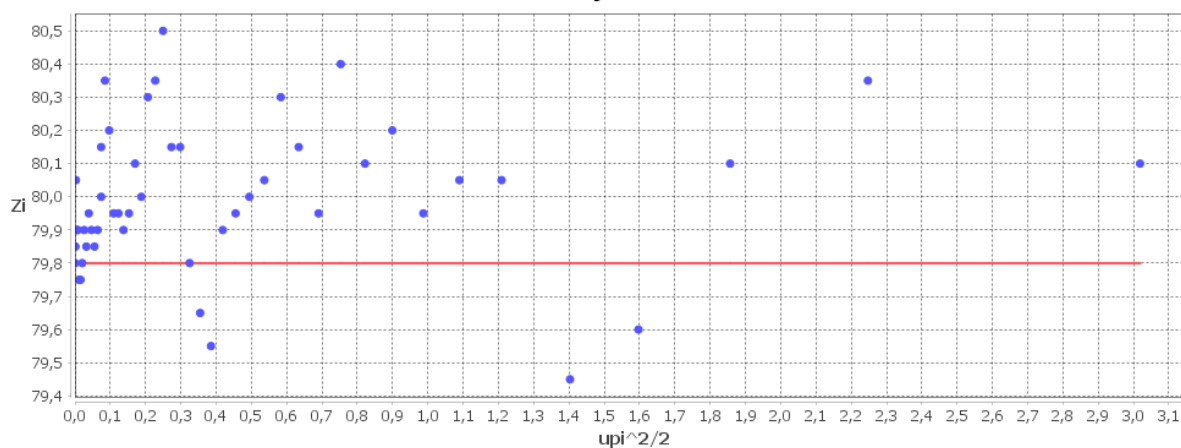


Obr. 4.36 – Graf symetrie úlohy
B2.12 Adstat

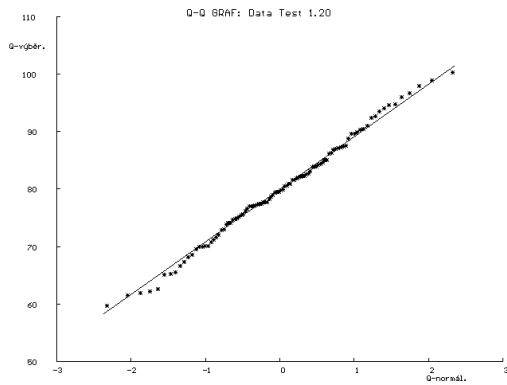


Obr. 4.37 – Graf symetrie úlohy B2.12
QCExpert

Graf symetrie

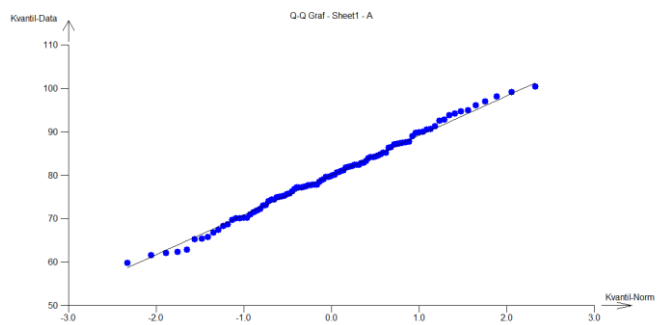


Obr. 4.38 – Graf symetrie úlohy B2.12 testovaného programu

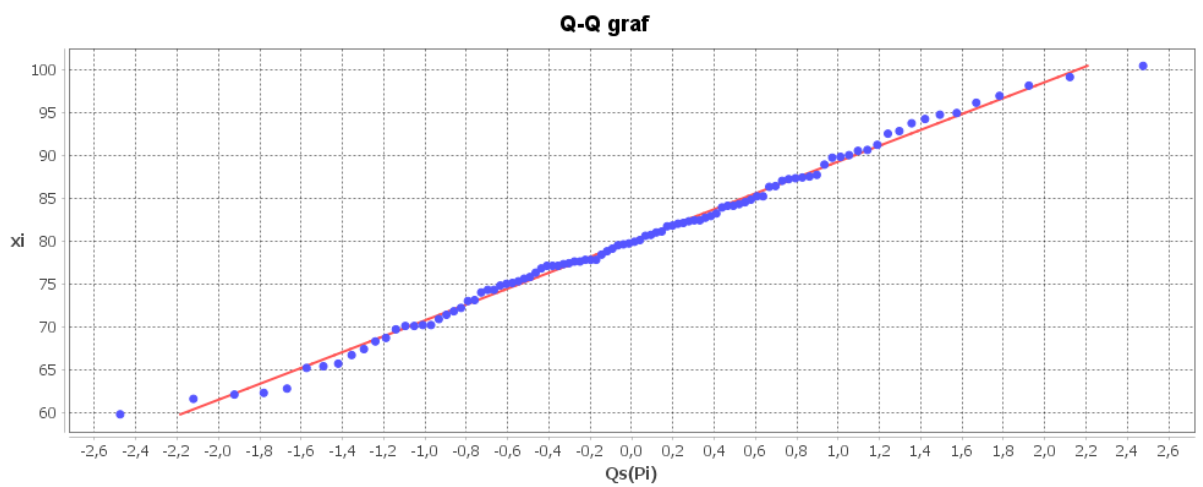


Obr. 4.39 – Q-Q graf úlohy B2.12

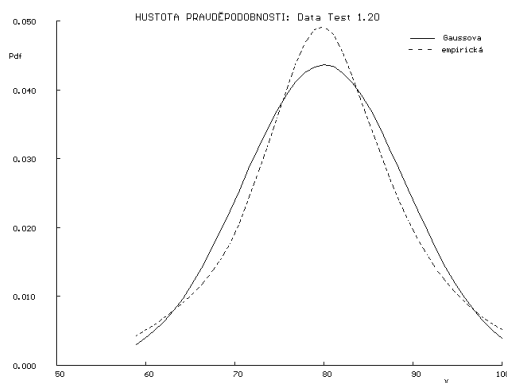
Adstat



Obr. 4.40 – Q-Q graf úlohy B2.12 QCExpert

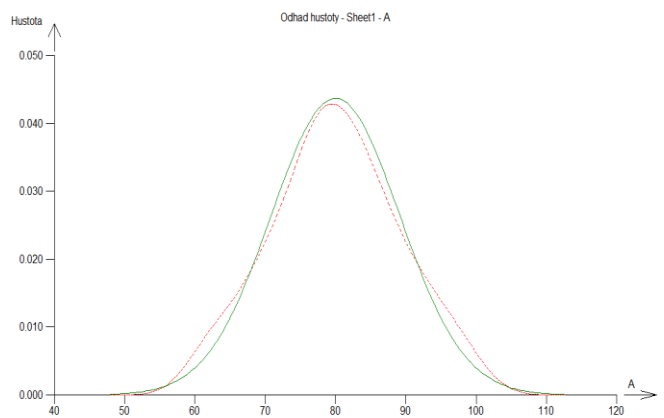


Obr. 4.41 – Q-Q graf úlohy B2.12 testovaného programu

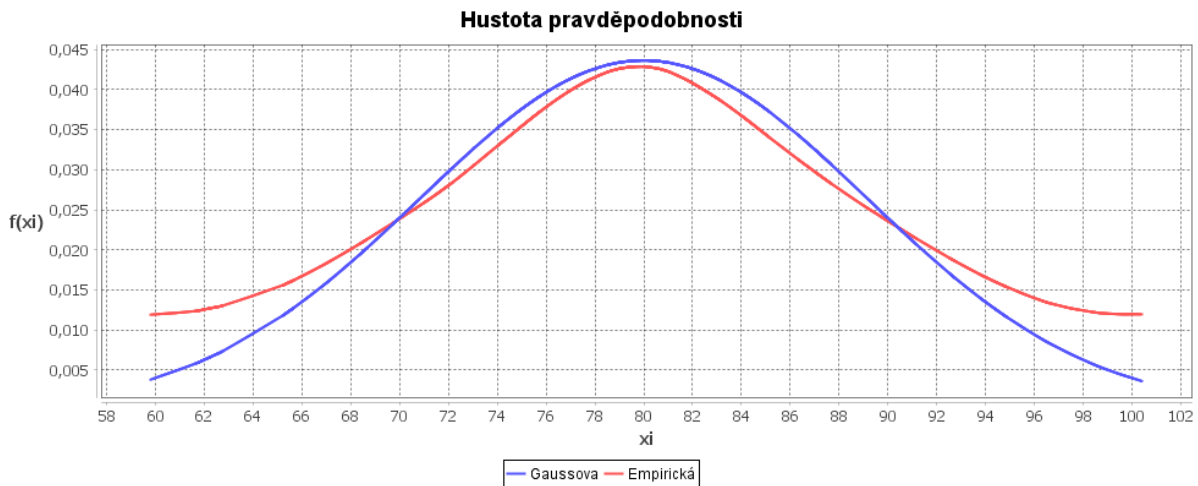


Obr. 4.42 – Graf hustoty
pravděpodobnosti úlohy B2.12

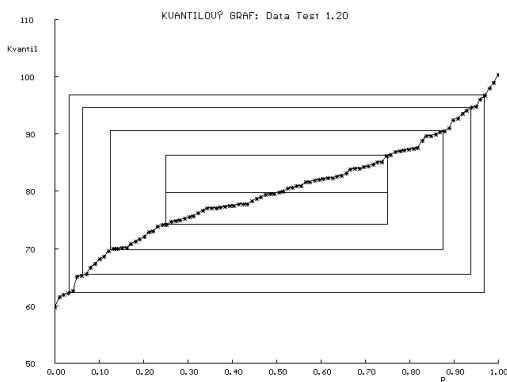
Adstat



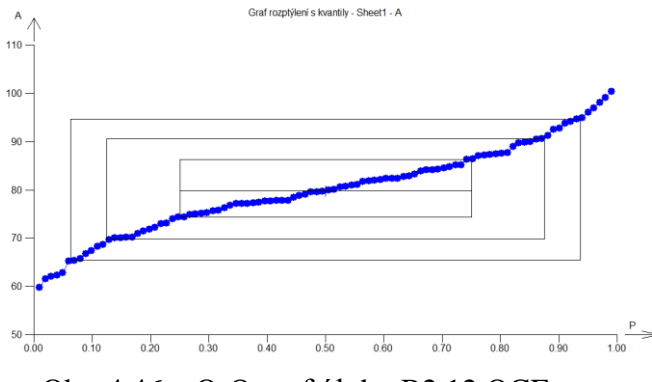
Obr. 4.43 – Graf hustoty pravděpodobnosti
úlohy B2.12 QCExpert



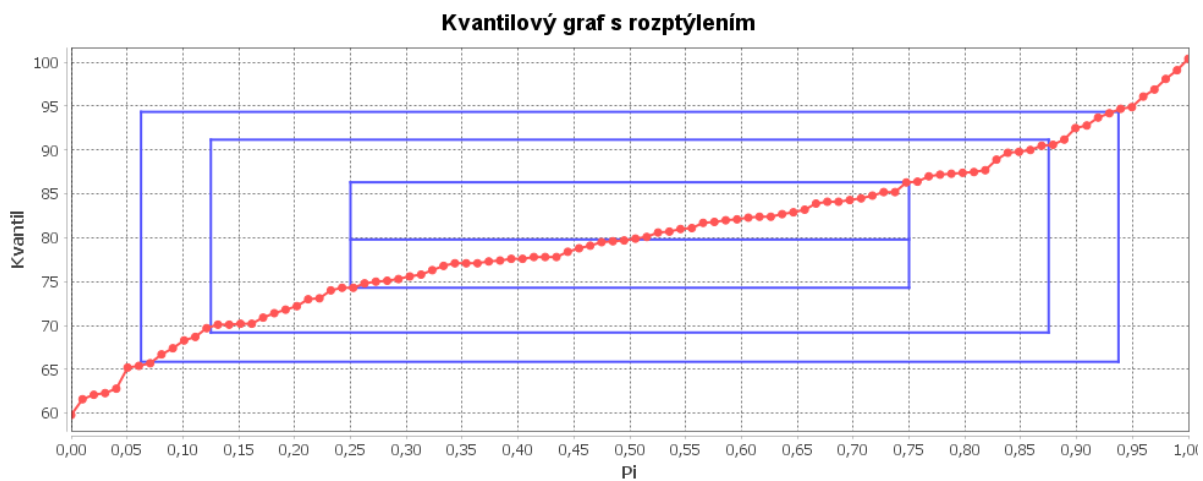
Obr. 4.44 – Graf hustoty pravděpodobnosti úlohy B2.12 testovaného programu



Obr. 4.45 – Q-Q graf úlohy B2.12
Adstat



Obr. 4.46 – Q-Q graf úlohy B2.12 QCExpert



Obr. 4.47 – Kvantilový graf rozptýlení úlohy B2.12 testovaného programu

Stejně odlišnosti v grafech jako u úlohy B2.02 jsou patrné i na grafech úlohy B2.12.

4.7.3 Úloha B2.22

Úloha je zaměřena na typ rozdělení obsahu kadmia ve vlasech. Celé zadání úlohy níže v obr. 4.48 i s tabulkou použitých dat.

Úloha B2.22 *Typ rozdělení obsahu kadmia ve vlasech*

U 98 náhodně vybraných lidí byla provedena analýza vlasů pro stanovení obsahu kadmia. Vlasy byly po promytí a odmaštění mineralizovány v přístroji Apion směsí kyslíku a amoniaku. Obsah kadmia byl v mineralizátu stanoven metodou AAS a přepočten na původní navážku vzorku vlasů. Nalezněte typ rozdělení, odlehle body a odhad střední hodnoty. Využijte také schéma sumarizace dat. Zdůvodněte, jakému odhadu dáte přednost.

Data: Obsah kadmia ve vlasech [mg/kg]:

0.558	0.063	0.442	0.049	0.041	0.044	0.380	0.630	0.179	0.179
..
0.164	0.064	0.229	0.065	0.157	0.060	0.951	0.374		

Obr. 4.48 – Zadání úlohy B2.22

U všech programů se testují pouze numerické výsledky, které jsou vypsány do jednotlivých tabulek, a též se porovnávají grafy pro testování normality dat. Vše je testováno na hladině významnosti 5 %.

Tab. 4.13 – Klasické odhady parametrů úlohy B2.22

	Testovaný program	Adstat	QCExpert
Medián	$1,790 \cdot 10^{-1}$	$1,800 \cdot 10^{-1}$	$1,635 \cdot 10^{-1}$
Průměr	$3,067 \cdot 10^{-1}$	$3,067 \cdot 10^{-1}$	$2,219 \cdot 10^{-1}$
Rozptyl	$1,823 \cdot 10^{-1}$	$1,823 \cdot 10^{-1}$	$4,766 \cdot 10^{-2}$
Šikmost	3,167	3,167	1,602
Špičatost	$1,379 \cdot 10^{+1}$	$1,379 \cdot 10^{+1}$	$4,953 \cdot 10^{+1}$
Směrodatná odchylka	$4,269 \cdot 10^{-1}$	$4,269 \cdot 10^{-1}$	$2,533 \cdot 10^{-1}$
3. centrální moment	$2,427 \cdot 10^{-1}$	$2,452 \cdot 10^{-1}$	neuveďeno
4. centrální moment	$4,489 \cdot 10^{-1}$	$4,536 \cdot 10^{-1}$	neuveďeno

V tab. 4.13 je patrné, že testovaný program má všechny hodnoty podobné s programem Adstat, ale naopak QCExpert vykazuje 60 % rozdíl oproti oběma programům. Naopak Adstat a testovaný program se více shodují ve třetím a čtvrtém centrálním momentu. Podle předchozích měření to vypadá, že čím větší počet, tím se bude více shodovat Adstat s testovaným programem, ale naopak QCExpert je v numerických datech více rozdílný.

Tab. 4.14 – Písmenové hodnoty úlohy B2.22

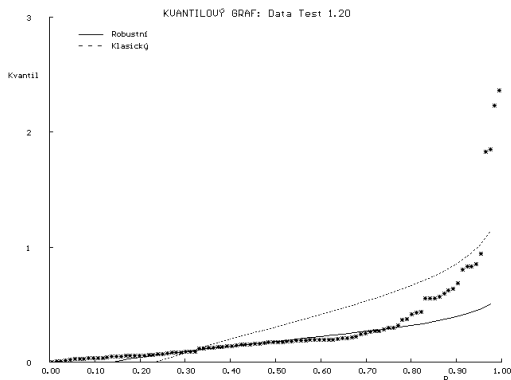
Kvantily	Meze	Testovaný program	Adstat
Sedecíl	Spodní	$3,245 \cdot 10^{-2}$	$3,419 \cdot 10^{-2}$
	Horní	$8,092 \cdot 10^{-1}$	$8,360 \cdot 10^{-1}$
Oktil	Spodní	$4,920 \cdot 10^{-2}$	$4,950 \cdot 10^{-2}$
	Horní	$6,036 \cdot 10^{-1}$	$5,969 \cdot 10^{-1}$
Kvartil	Spodní	$7,800 \cdot 10^{-2}$	$7,850 \cdot 10^{-2}$
	Horní	$2,958 \cdot 10^{-1}$	$3,013 \cdot 10^{-1}$

Tab. 4.15 – Kvantilové míry úlohy B2.22

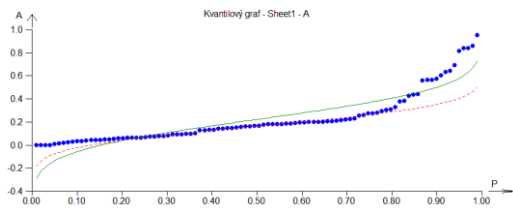
Kvantily	Kvantily	Testovaný program	Adstat
Rosah	Sedecíl	$7,767 \cdot 10^{-1}$	$8,018 \cdot 10^{-1}$
	Oktil	$5,544 \cdot 10^{-1}$	$5,474 \cdot 10^{-1}$
	Kvartil	$2,178 \cdot 10^{-1}$	$2,228 \cdot 10^{-1}$
Polosuma	Sedecíl	$4,208 \cdot 10^{-1}$	$4,351 \cdot 10^{-1}$
	Oktil	$3,264 \cdot 10^{-1}$	$3,232 \cdot 10^{-1}$
	Kvartil	$1,869 \cdot 10^{-1}$	$1,899 \cdot 10^{-1}$
Délka konců	Sedecíl	1,272	1,281
	Oktil	$9,345 \cdot 10^{-1}$	$8,991 \cdot 10^{-1}$
	Kvartil	0	0
Šikmost	Sedecíl	$-3,113 \cdot 10^{-1}$	$-1,036 \cdot 10^{-1}$
	Oktil	$-2,659 \cdot 10^{-1}$	$-1,078 \cdot 10^{-1}$
	Kvartil	$-3,617 \cdot 10^{-2}$	$6,175 \cdot 10^{-2}$
Pseudosigma	Sedecíl	$2,538 \cdot 10^{-1}$	$2,620 \cdot 10^{-1}$
	Oktil	$2,410 \cdot 10^{-1}$	$2,380 \cdot 10^{-1}$
	Kvartil	$1,615 \cdot 10^{-1}$	$1,652 \cdot 10^{-1}$

Tab. 4.14 a 4.15 obsahují data pouze testovaného programu a programu Adstat. QCExpert ani Statistica nemají výpočty písmenných hodnot a kvantilových mír. Testovaný program vychází ze vztahu 1.4 a vztahů uvedených v tab. 1.1. Není tedy známo, z jakých vzorců vychází program Adstat, ale výsledné hodnoty obou programů se liší v minimálních hodnotách. V případě šikmosti je rozdíl větší, ale pro testovaný program se využil ověřený vztah v tab. 1.1 převzatý z literatury (Meloun, 2004) a tedy i správné hodnoty.

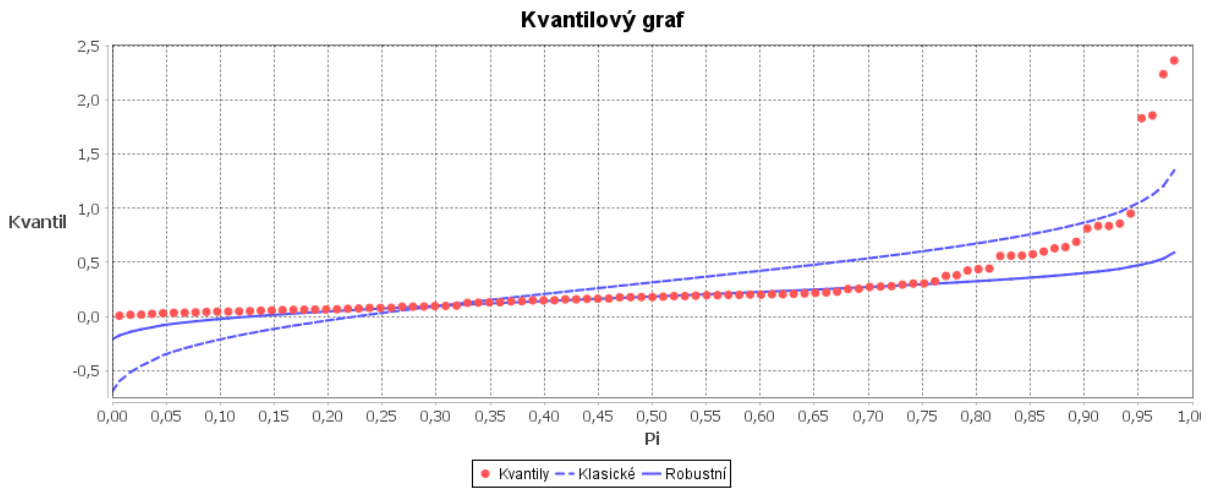
Níže jsou vypsány všechny grafy testovaného programu, který se porovnává s renomovanými programy Adstat a QCExpert.



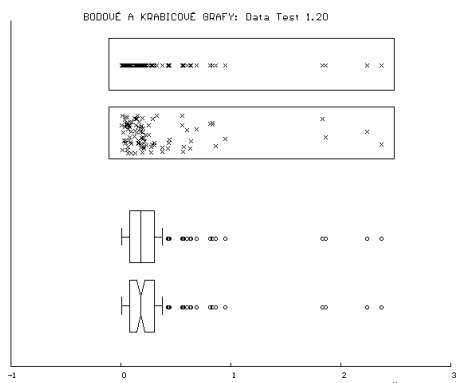
Obr. 4.49 – Kvantilový graf úlohy B2.22 Adstat



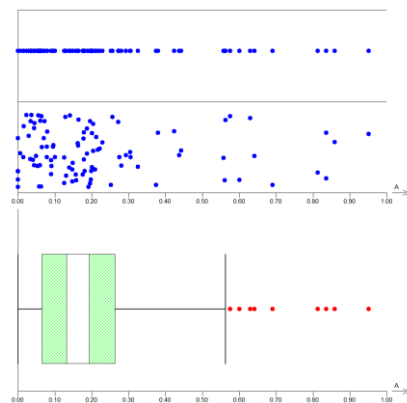
Obr. 4.50 – Kvantilový graf úlohy B2.22 QCExpert



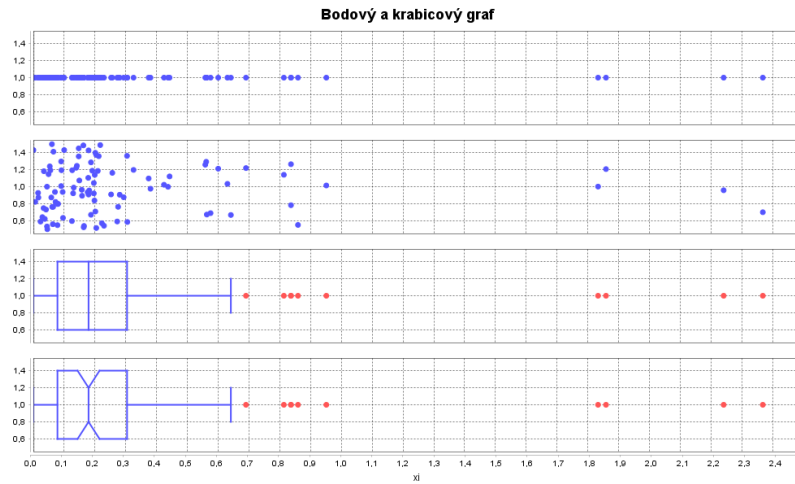
Obr. 4.51 – Kvantilový graf úlohy B2.22 testovaného programu



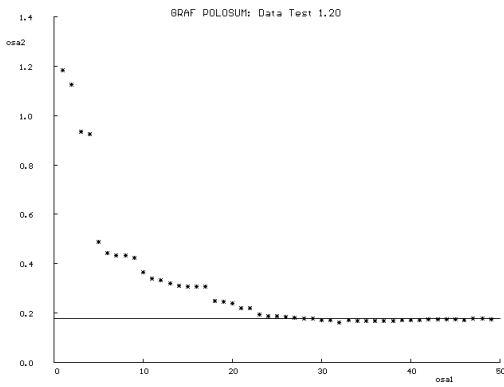
Obr. 4.52 – Bodový a krabicový graf úlohy B2.22 Adstat



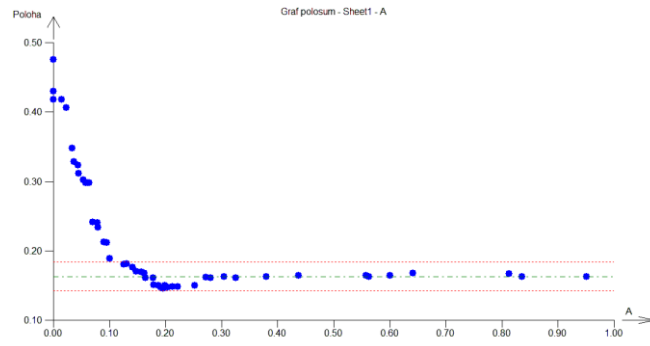
Obr. 4.53 – Bodový a krabicový graf úlohy B2.22 QCExpert



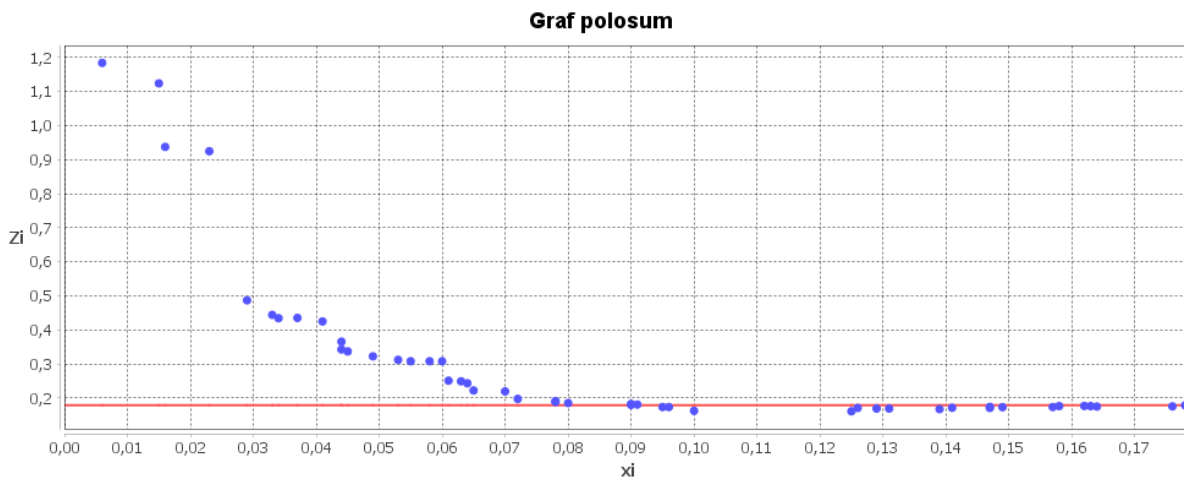
Obr. 4.54 – Bodový a krabicový graf úlohy B2.22 testovaného programu



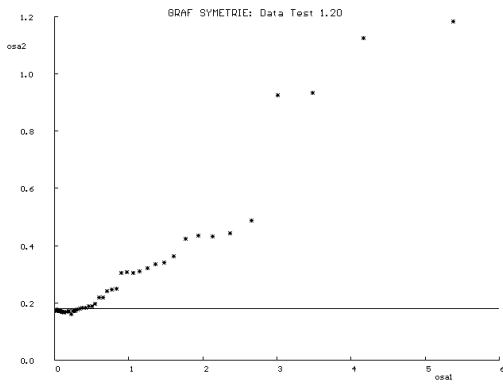
Obr. 4.55 – Graf polosum úlohy B2.22 Adstat



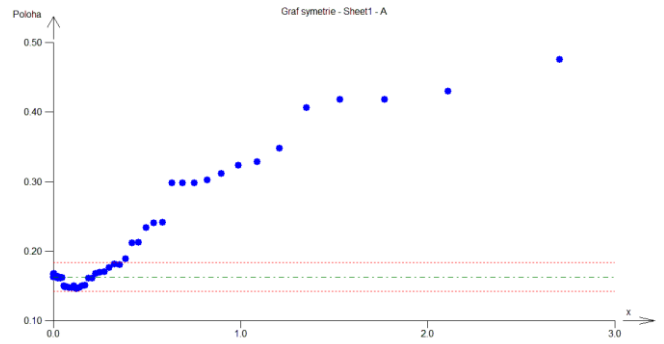
Obr. 4.56 – Graf polosum úlohy B2.22 QCExpert



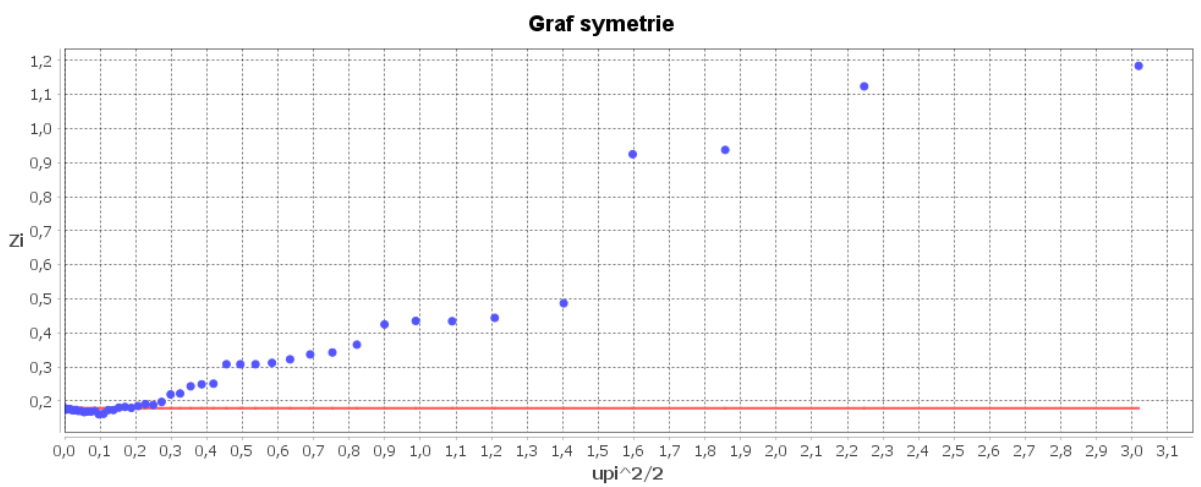
Obr. 4.57 – Graf polosum úlohy B2.22 testovaného programu



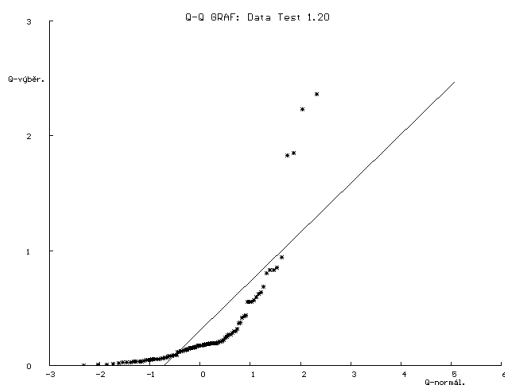
Obr. 4.58 – Graf symetrie úlohy B2.22 Adstat



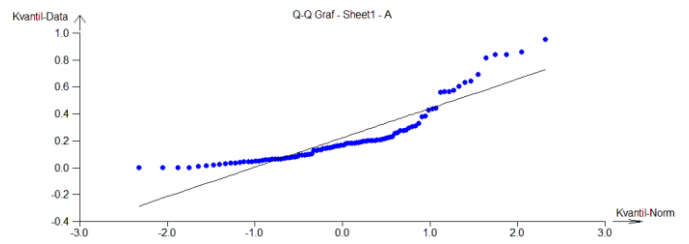
Obr. 4.59 – Graf symetrie úlohy B2.22 QCExpert



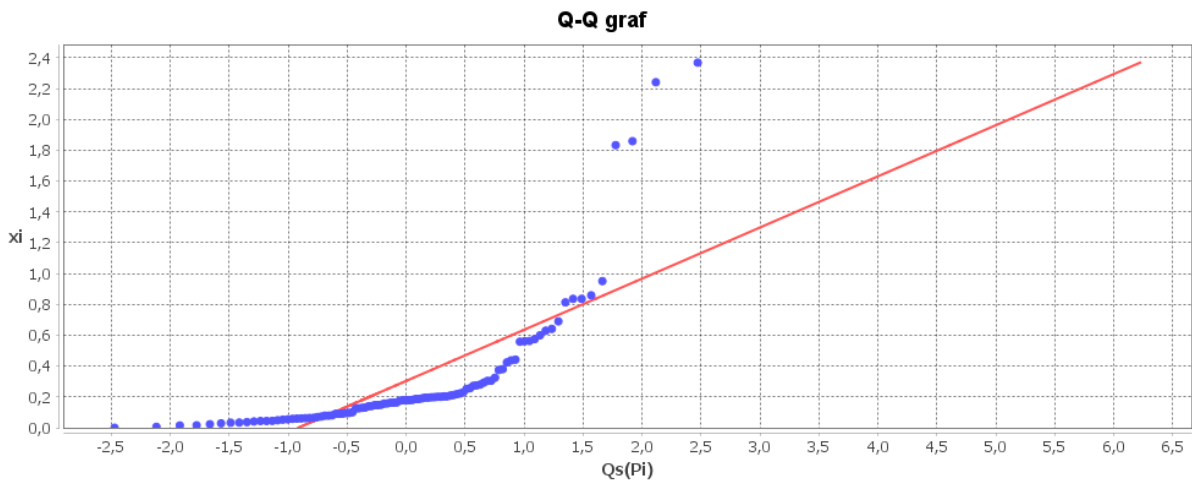
Obr. 4.60 – Graf symetrie úlohy B2.22 testovaného programu



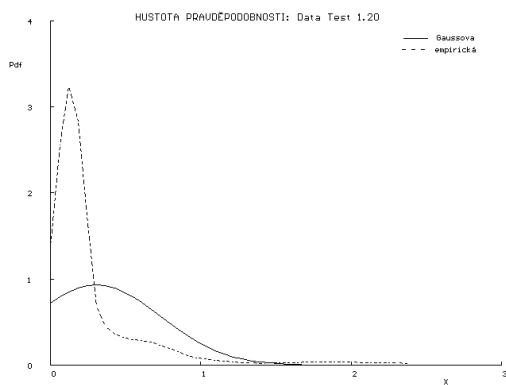
Obr. 4.61 – Q-Q graf úlohy B2.22 Adstat



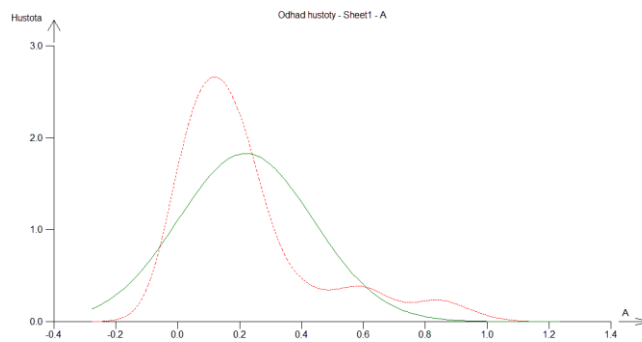
Obr. 4.62 – Q-Q graf úlohy B2.22 QCExpert



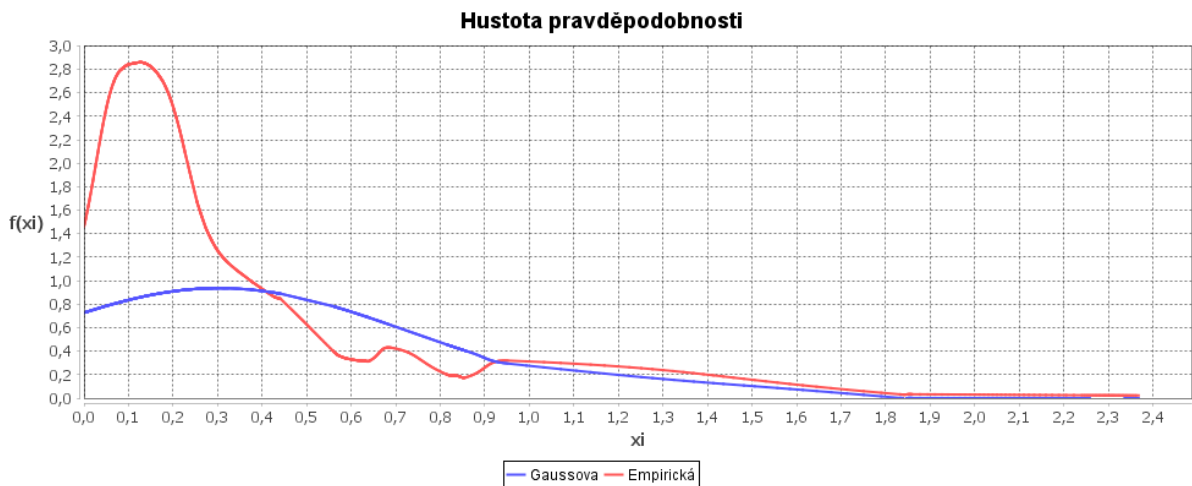
Obr. 4.63 – Q-Q graf úlohy B2.22 testovaného programu



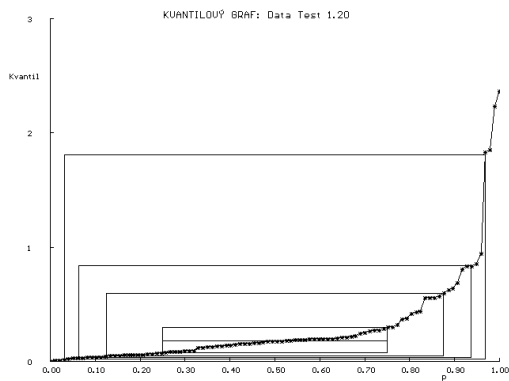
Obr. 4.64 – Graf hustoty pravděpodobnosti úlohy B2.22 Adstat



Obr. 4.65 – Graf hustoty pravděpodobnosti úlohy B2.22 QCExpert

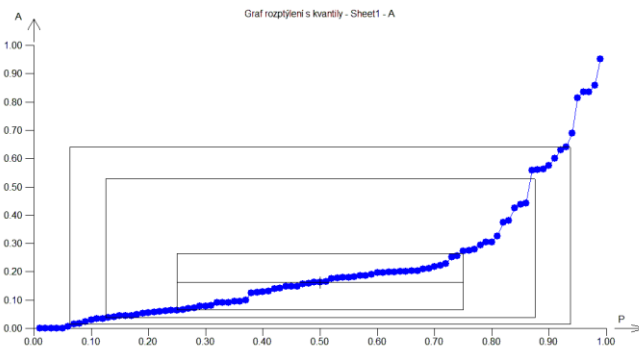


Obr. 4.66 – Graf hustoty pravděpodobnosti úlohy B2.22 testovaného programu



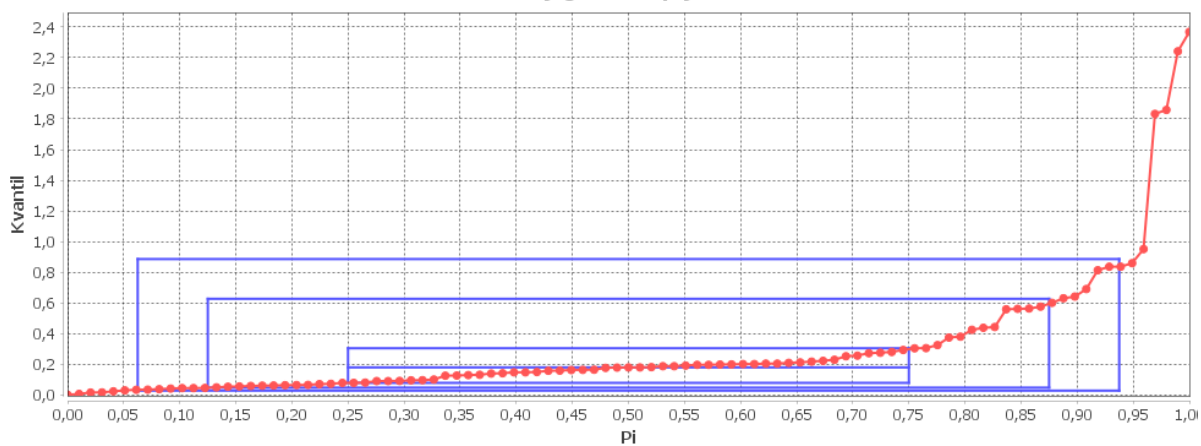
Obr. 4.67 – Q-Q graf úlohy B2.22

Adstat



Obr. 4.68 – Q-Q graf úlohy B2.22 QCExpert

Kvantilový graf s rozptýlením



Obr. 4.69 – Kvantilový graf rozptýlení úlohy B2.22 testovaného programu

Stejně odlišnosti v grafech jako u úlohy B2.02 jsou patrné i na grafech úlohy B2.22.

5 ZÁVĚR

Cílem diplomové práce bylo vytvoření programu na testování normality/nenormality jednorozměrných číselných dat s grafickým výstupem a porovnat jej s renomovanými programy.

Program je vytvořen v jazyce JAVA, který umožňuje vytvářet programy s grafickým rozhraním. Předlohou programu je Adstat s tím, že se využívá stejných numerických výpočtů i grafů doplněných o výpočty parametrických a neparametrických testů. Tyto testy jsou porovnávány s příklady z literatury, ze které byly využity vztahy pro program, protože samotný program Adstat nedisponuje kompletními parametrickými a neparametrickými testy

Program umí dávkově nahrát soubor s daty ve formátu txt, vypsat je přehledně do tab. programu, nastavovat parametry jako je například hladina významnosti, což vše určitě ocení uživatelé, kteří mohou program využít jako výukový pro exploratorní analýzu dat. Program je vhodný pro výuku, protože dává relevantní výstupy exploratorní analýzy dat, což potvrzují i programy Adstat, Statistica a QCExpert. Uživatel u programu jistě ocení, že lze výsledky analýzy uložit do PDF souborů nebo zadaná data úlohy lze uložit do souboru typu txt s tím, že je může dále využít pro další analýzu v jiném programu nebo na dalších statistických metodách.

S renomovanými programy se porovnává nově vytvořený program v oblasti přesnosti numerických výpočtu proměnných jako je průměr, medián, směrodatná odchylka atd. Také se porovnávají samotné grafy, kde rozdíly grafů lze hodnotit pouze vizuálním porovnáním.

Výchozím porovnávacím programem je Adstat, ve kterém se projevují rozdíly v případě numerických výpočtu třetího a čtvrtého centrálního momentu. To samé platí i pro sedecil a oktil, kde lze nalézt malou odchylku dat. Testovaný program vychází z ověřených vztahů, které jsou popsány v literatuře, kde odchylku může dát samotný využitý datový typ ve vytvořeném programu, zastaralost programu Adstat s chybnými výpočty nebo samotné zaokrouhlování dat, které je v programu využito v co nejmenším množství, aby se předešlo zvyšování chybovosti při zaokrouhlování. Z pohledu grafů je ověřeno, že v části kvantilového grafu a krabicového grafu je nalezena chyba u programu Adstat. Tato domněnka je potvrzena dalším renomovaným programem QCExpert, který se shoduje s vytvořeným programem.

Nově vytvořený program se shoduje s renomovanými programy, který má stejné či velmi podobné výsledky grafů a v případě numerických výsledků jsou odchylky v řádech setin až desetin procent.

POUŽITÁ LITERATURA

- MELOUN, M.; MILITKÝ, J. 2004. *Statistická analýza experimentálních dat*. Vyd. 2., upr. a rozš. Praha: Academia. 953 s. ISBN 80-200-1254-0.
- HEBÁK, P. 2015. *Statistické myšlení a nástroje analýzy dat*. 2. vydání. Praha: Informatorium. ISBN 978-80-7333-118-4.
- HENDL, J. 2012. *Přehled statistických metod: analýza a metaanalýza dat*. 4., rozš. vyd. Praha: Portál. 734 s. ISBN 978-80-262-0200-4.
- HEROUT, P. 2010. *Učebnice jazyka Java: úvod do používání*. 5., rozš. vyd. České Budějovice: Kopp. 386 s. ISBN 978-80-7232-398-2.
- KYSILKA, P. 2006. *Java (24) - úvod do grafiky a GUI*. [online]. Linuxsoft.cz. [cit. 07.03.2017].
Dostupné z: http://www.linuxsoft.cz/article.php?id_article=1184
- MELOUN, M; MILITKÝ, J. 2002. *Kompendium statistického zpracování dat: metody a řešené úlohy včetně CD*. Praha: Academia. ISBN 80-200-1008-4.

PŘÍLOHY

A - CD

Příloha k diplomové práci

Program pro testování normality číselných dat

Lukáš Vladyka

CD

Obsah

- 1 Text diplomové práce ve formátu PDF
- 2 Úplný zdrojový kód aplikace
- 3 Spustitelná aplikace ve formátu exe
- 4 Podkladové soubory testovaných úloh ve formátu txt