

VYUŽITÍ MATLAB WEB SERVERU PRO INTERNETOVOU VÝUKU ANALÝZY DAT A ŘÍZENÍ JAKOSTI

Aleš Linka¹, Petr Volf²

¹Katedra textilních materiálů, FT TUL,

²Katedra aplikované matematiky, FP TUL

ABSTRAKT. Internetové interaktivní učebnice se v současné době stávají nezbytnou součástí systému výuky nejen matematiky, ale zejména matematické statistiky a analýzy dat. Dokladem toho jsou například interaktivní úvod do statistiky „MMstat“ (viz odkaz [8]) nebo projekt „E-stat“ pro tvorbu interaktivního kursu aplikované statistiky (viz [4]). Začleněním těchto interaktivních textů do výukového procesu se tento proces stává dynamičtější, čtenáři získávají zpravidla možnost si interaktivně vyzkoušet základní poznatky z učebnice na modelových příkladech a získané zkušenosti mohou použít při statistické analýze jednoduchých vlastních dat. V našem konkrétním případě jsme k vytvoření internetového interaktivního systému použili MATLAB a MATLAB Web Server.

1. ÚVOD

S pojmy jakost a kvalita se dnes setkáváme téměř na každém kroku. Otázkou však je, co se těmito pojmy vlastně rozumí. Pro porozumění výrobních procesů, pro jejich řízení a hodnocení jakosti je při rigorózním přístupu nezbytný matematicko-statistický aparát. Při přípravě internetového interaktivního učebního textu pro předmět „Řízení jakosti v textilu“ jsme se proto zaměřili především na vytvoření jednoduchých aplikací pro základní statistickou analýzu dat. Vzniklý text lze dále rozšiřovat a doplňovat nejen o další textovou část, ale především o další interaktivní aplikace a praktické demonstrace teoretických poznatků. Systém je v místech s přístupem na Internet velmi rychle dostupný a okamžitě použitelný pro jednoduché výpočty. Učební textu doplňují interaktivní aplikace vytvořené v prostředí MATLAB Web Serveru (viz [6]). Aplikace lze používat parametricky, což usnadňuje včlenění dynamických prvků přímo do interaktivního učebního textu. Vlastní text je členěn do následujících kapitol a podkapitol:

1. Základní rozdělení pravděpodobnosti

Číselné charakteristiky
Základní spojitá rozdělení
Základní diskrétní rozdělení

2. Základy matematické statistiky

Průzkumová analýza dat
Testy hypotéz
Neparametrické testy
Regresní analýza (pdf soubor)
Testy dobré shody (připravuje se v HTML formě)

3. Řízení jakosti

Regulační diagramy
Indexy způsobilosti (připravuje se v HTML formě)

4. Kvalita a textil

Textil a řízení jakosti

5. Generované příklady

6. Matlab Web Server aplikace

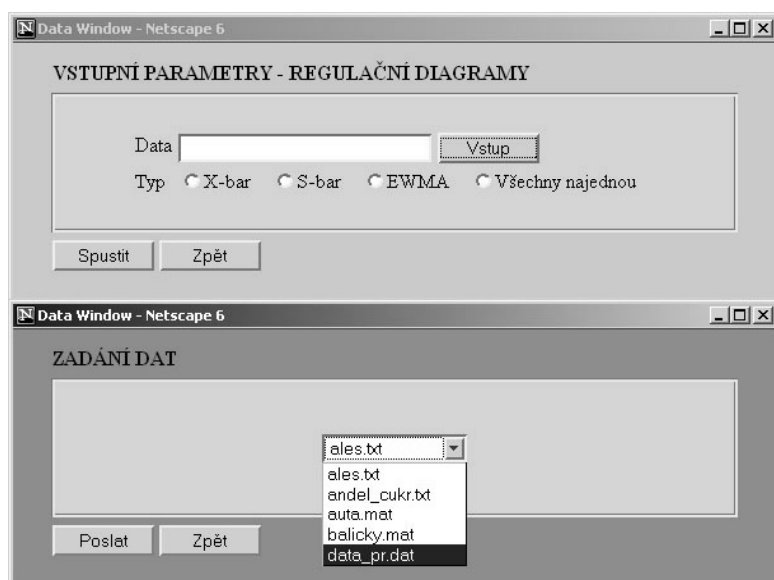
7. Zdroje

2. APLIKACE PRO MATLAB WEB SERVER

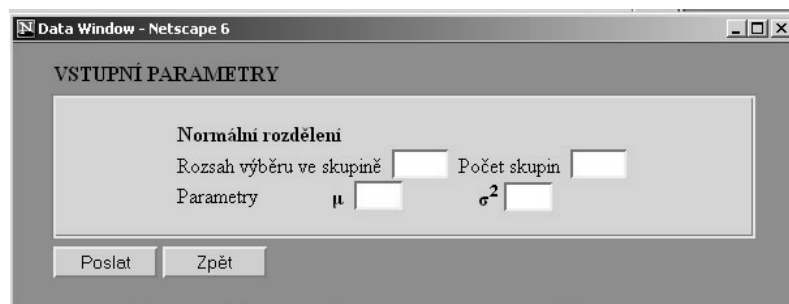
Při vlastní tvorbě aplikací se ukázalo, že k tomu aby mohly být statistické postupy efektivně pomocí prostředí Matlab Web Serveru demonstrovány, bude nutné vytvořit alespoň elementární blok pro vstup dat (načtení dat). Tento blok obsahuje tyto možnosti získání dat:

- generování náhodných dat,
- ze souboru z přístupného adresáře na serveru,
- přímo z klávesnice,
- ze souboru z lokálního disku klienta.

Ukázka vstupního okna a okna pro načtení dat ze souboru umístěného na serveru je zobrazena na obr. 1. Výběr dat nemusí být v tomto případě omezen pouze na data umístěná na serveru tj. data, která do příslušného adresáře umístí správce, pokud do adresáře není přístup, ale lze do vstupního okna poslat i data ze souboru na lokálním disku klienta, data zadat přímo do vstupní řádky nebo data vygenerovat pomocí aplikace Matlab Web Serveru (viz obr. 2 - vstupní okno pro generování dat z normálního rozdělení pro aplikaci „Regulační diagramy“).



Obr. 1 Příklad výstupních oken pro vstup dat ze souboru.



Obr. 2 Příklad vstupního okna pro generování dat.

Pro efektivní vytvoření tohoto bloku bylo nutné využít i skriptovacích jazyků PHP a JavaScript, s čímž sice v původně nepočítalo, ale pro zachování efektivnosti a interaktivnosti dynamických HTML stránek se tento krok ukázal jako nezbytný. Vlastní aplikace můžeme rozdělit zhruba do 3 skupin podle typu výstupu do WWW stránky:

- výstupem je grafika
- výstup je textový

c) výstup je smíšený (text i grafika)

Pro většinu uváděných aplikací lze popis použitých metod statistické analýzy dat najít v [1], [2] a [8], či alespoň částečně ve vlastním interaktivním textu „Řízení jakosti v textilu“. Konkrétně vytvořeným aplikací jsou věnovány následující odstavce.

2.1. GRAFICKÝ VÝSTUP

Rozdělení náhodné veličiny

Rozdělení náhodné veličiny je vlastně jednoduchá demonstrace 2D grafických možností Matlabu, demonstrující základní spojitá a diskrétní rozdělení. Výsledkem procedury je graf zobrazující dle výběru distribuční funkci, hustotu (pravděpodobnostní funkci) a kvantilovou funkci příslušného rozdělení. U spojitých rozdělení lze zobrazit 3 grafy současně.

2.2. TEXTOVÝ VÝSTUP

Generované příklady

V aplikaci jsou některé prvky v zadání příkladu náhodně generované, tj. mění se s novým načtením aplikace. Po zadání výsledku vypočteného uživatelem dojde pomocí MATLAB Web Serveru k zobrazení správného výpočtu získaného Matlabem i ke kontrole zadaného výsledku. Vytvořeny byly aplikace pro ověření základních znalostí týkajících rovnoměrného, normálního (viz obr. 3) a Poissonova rozdělení.

Příklad. Předpokládejme, že měřený znak X je náhodná veličina s normálním rozdělením $N(-1,1)$. Vypočítejte $P(X \geq 0)$?

Odpověď. $P(X \geq 0) = 0.56$

Submit

Řešení. Ke správnému výsledku můžeme dojít několika různými postupy. Například můžeme využít vztahu (1.1). Pro hledanou pravděpodobnost dostáváme

$$P(X \geq 0) = 1 - F_X(0) = 0.15866.$$

kde F_X je distribuční funkci rozdělení $N(-1,1)$. Jinou alternativou může být hledanou pravděpodobnost odříznout tak, že využijeme normování náhodné veličiny X podle vztahu (2.2) a tabelovaných hodnot distribuční funkce normovaného normálního rozdělení

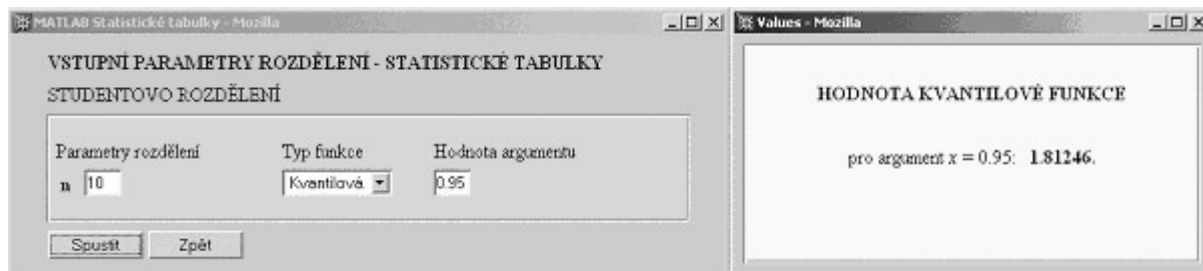
$$P(X \geq 0) = P(U \geq 1) = 1 - \Phi(1) = 1 - 0.84134 = 0.15866.$$

Kontrola odpovědi. Zadaná hodnota $P(X \geq 0)$ je nesprávná zaokrouhleno na čtyři platné číslice.

Obr. 3 Příklad výstupního okna s generovaným příkladem.

Statistické tabulky

Procedura počítá hodnoty hustot, distribučních a kvantilových funkcí základních pravděpodobnostních rozdělení pro zadanou hodnotu argumentu. Funkce se liší svým definičním oborem. Kvantilová funkce je definovaná na intervalu $(1,0)$, distribuční funkce je definovaná pro všechna reálná čísla a hustota je na nosiči zvoleného pravděpodobnostního rozdělení. Jedno ze vstupních oken a okno výstupní je obr. 4.

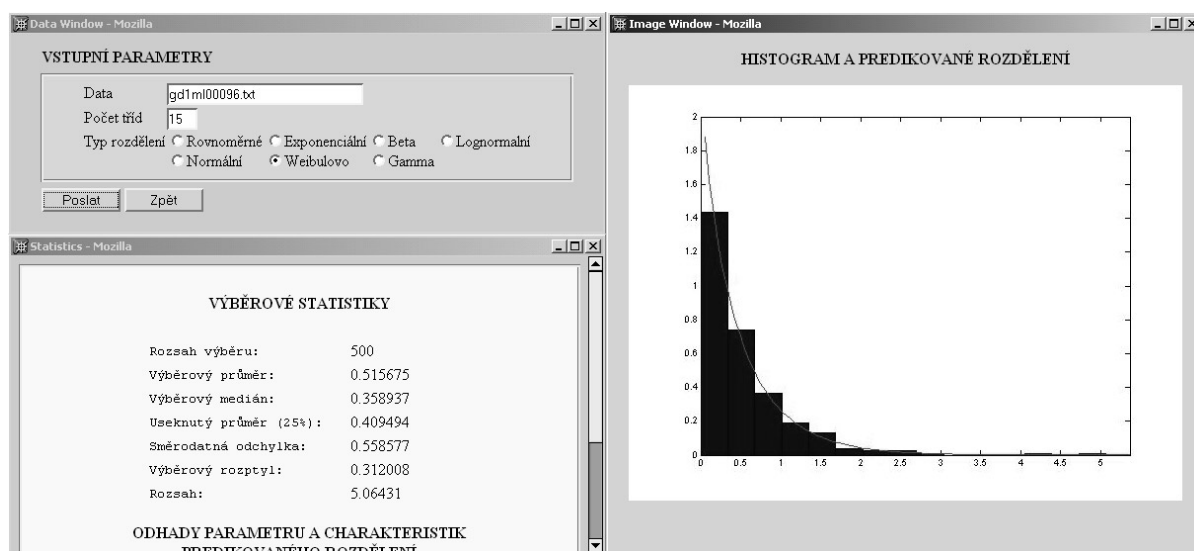


Obr. 4 Vstupní a výstupní okno pro aplikaci „Statistické tabulky“.

2.3. SMÍŠENÝ VÝSTUP

Průzkumová analýza dat (EDA) I

Procedura umožňuje zobrazit v textovém okně hodnoty základních výběrových charakteristik a odhady parametrů zadaného rozdělení na základě vstupního souboru dat. V grafickém okně procedura zobrazí histogram spolu s hustotou predikovaného rozdělení (viz obr. 5).



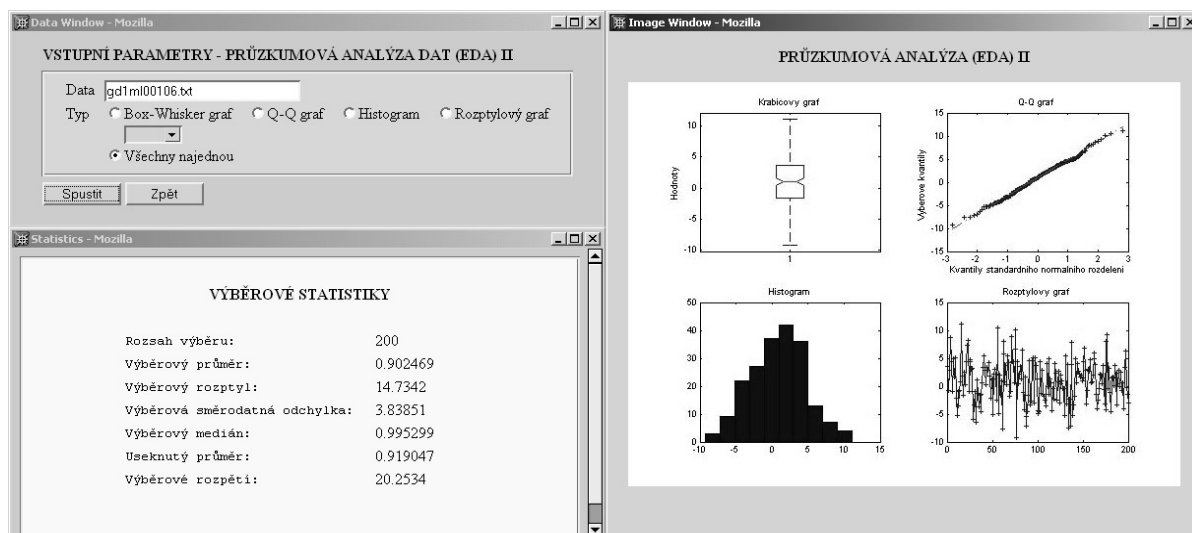
Obr. 5 Vstupní a výstupní okno pro aplikaci „Průzkumová analýza dat (EDA) I“.

Průzkumová analýza dat (EDA) II

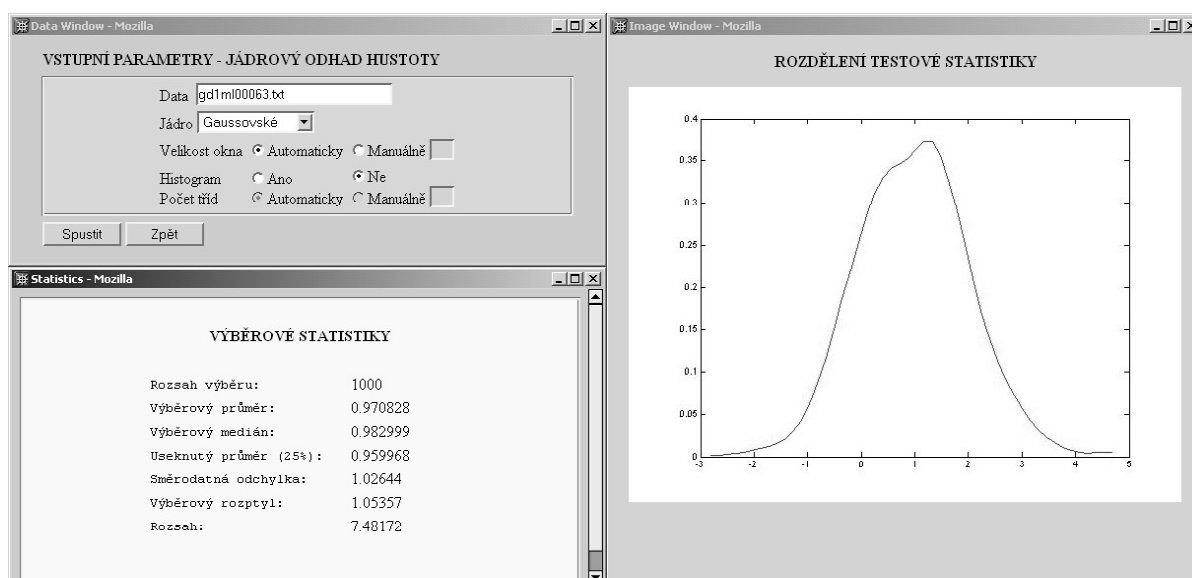
Aplikace zobrazuje v textovém okně hodnoty základních výběrových statistik, v grafickém okně pak lze získat Box-Whisker krabicový graf, Q-Q graf, histogram a rozptylový graf. Q-Q graf je nastaven pro ověření normality (viz obr. 6).

Jádrový odhad hustoty

Aplikace „Jádrový odhad hustoty“ zobrazuje v textovém okně hodnoty základních výběrových statistik. V grafickém okně, pak jádrový odhad hustoty a volitelně můžeme zároveň zobrazit histogram (viz obr. 7). Zatím lze vybírat ze tří jader (gaussovské, trojúhelníkové a Epanechnikovovo), volit šířku okna či použít automatickou volbu. Podobné je to u histogramu.



Obr. 6 Vstupní a výstupní okno pro aplikaci „Průzkumová analýza dat (EDA) II“.



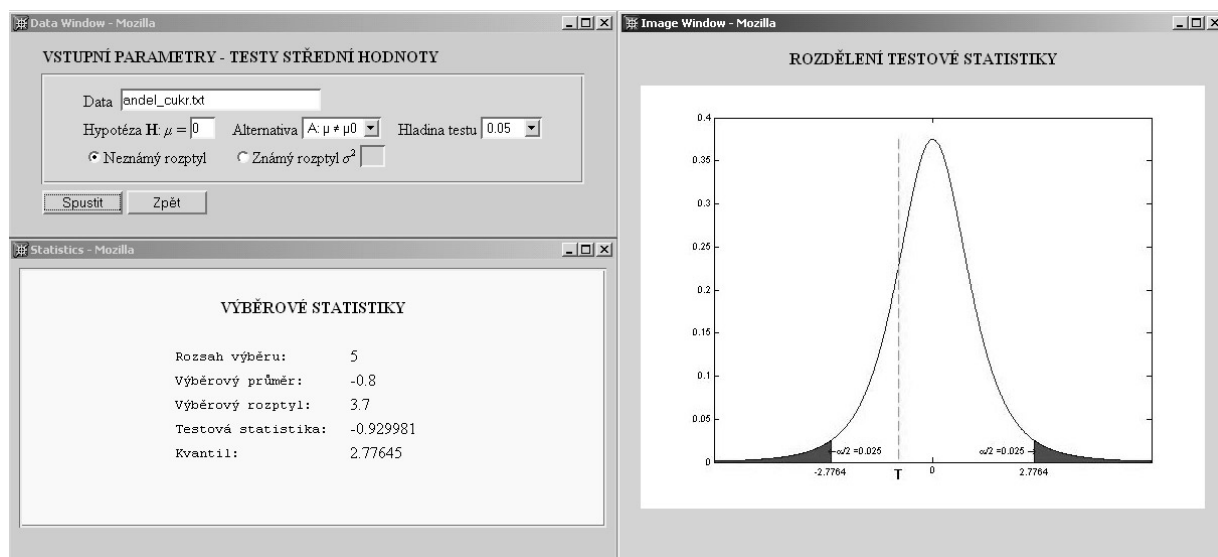
Obr. 7 Vstupní a výstupní okno pro aplikaci „Jádrový odhad hustoty“.

Testy hypotéz

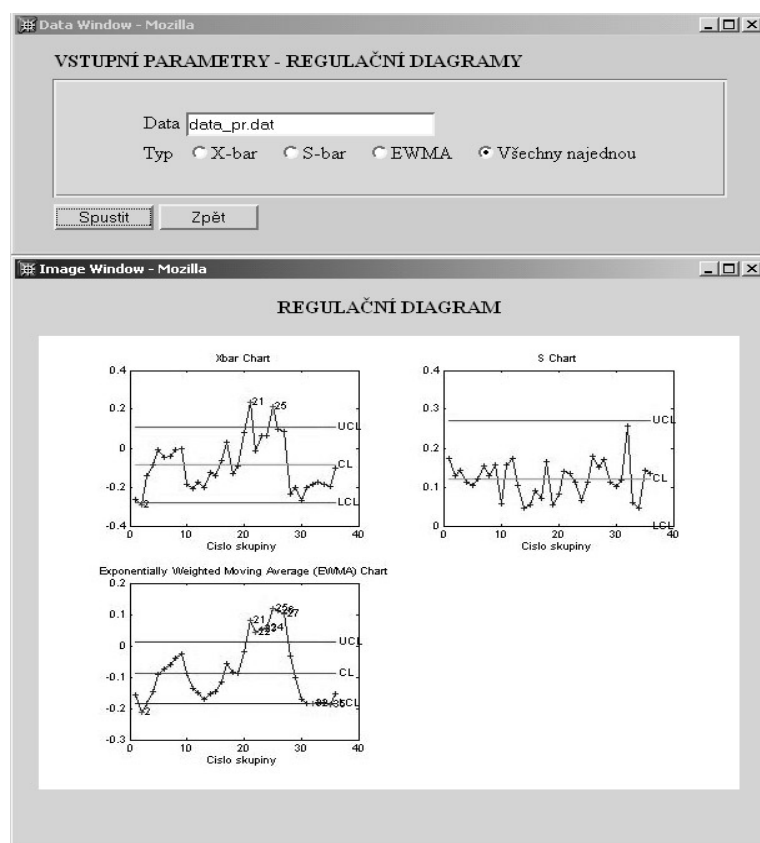
V aplikaci „Testy hypotéz“ se provádí test jednoduché hypotézy o střední hodnotě proti oboustranné a jednostranným alternativám v případě, kdy náhodný výběr pochází z normálního rozdělení. V textovém okně lze zobrazit hodnoty základních statistik, hodnotu testové statistiky, interval spolehlivosti pro testovaný znak. V grafickém okně se pak zobrazí grafická ilustrace testu, spolu s vyznačením testové statistiky, hladiny testu a kritického oboru (viz obr. 8).

Regulační diagramy

Tato procedura zobrazí meze pro regulační diagram, tj CL, UCL, LCL a jeden ze tří typů zatím dostupných regulačních diagramů, tj. Shewhartův regulační diagram pro průměr, diagram pro směrodatnou odchylku a EWMA diagram. Základní popis použitých metod lze najít v [9] nebo [10].



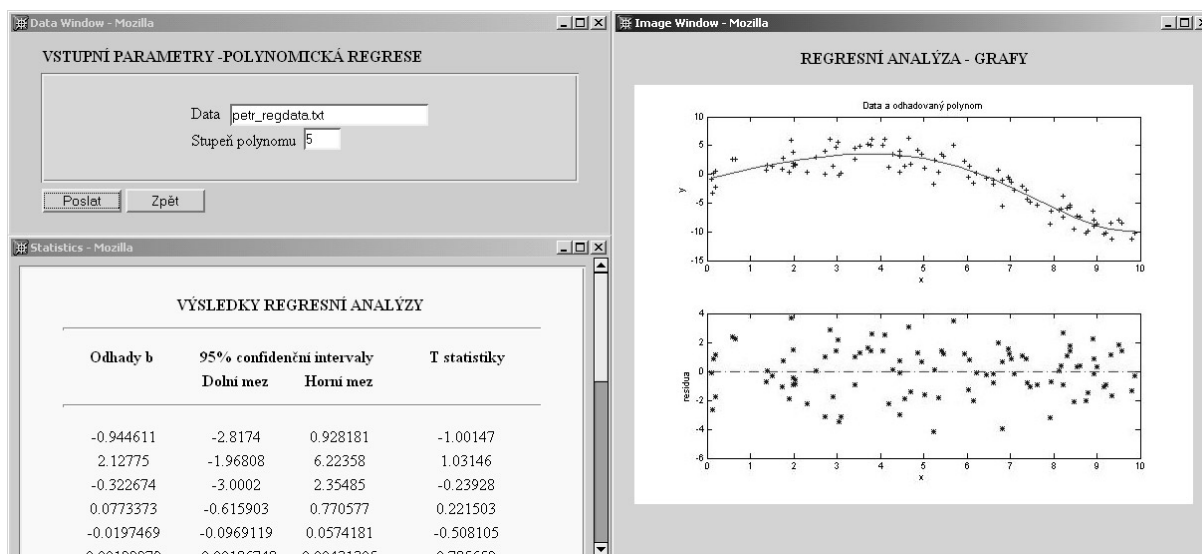
Obr. 8 Vstupní a výstupní okno pro aplikaci „Testy hypotéz“.



Obr. 9 Vstupní a výstupní okno pro aplikaci „Regulační diagramy“.

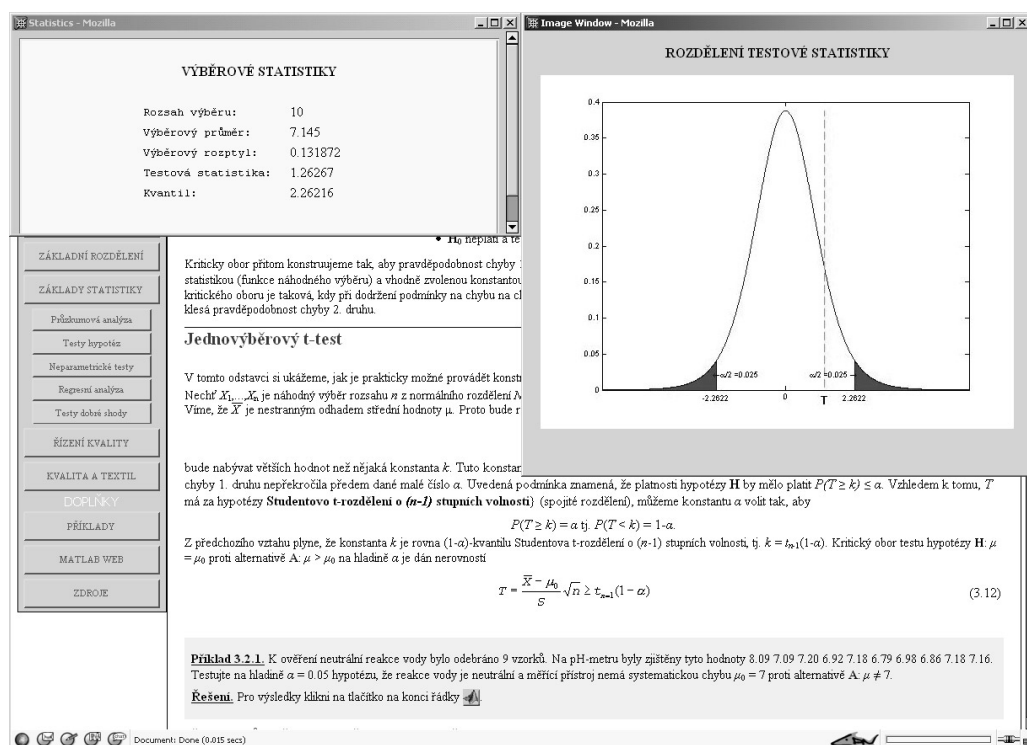
Polynomická regrese

Tato aplikace počítá základní regresní analýzu pro regresní funkce ve tvaru polynomu. Parametrem modelu je stupeň regresního polynomu. V grafickém okně je zobrazen graf predikovaného polynomu a rezidua.



Obr. 10 Vstupní a výstupní okno pro aplikaci „Polynomická regrese“.

Text je průběžně doplňován o další moduly. Zároveň jsou vytvářeny nové aplikace typu generovaných příkladů. Postupně zřejmě dojde k doplnění a úpravě stávajících modulů, neboť se ukazuje nutností doplnit moduly o stručnou nápovědu pro zadávání parametrů a dat. V řadě případech by se rovněž hodilo zobrazení datového souboru. Jak již bylo zmíněno na začátku, lze používat moduly parametricky přímo v textu. S tím odpadá mezičlánek v podobě okna (html formuláře) pro vstup parametrů, což prezentováno na obr.11. V neposlední řadě o zřejmě doplnění interaktivních testů z databáze příkladů (viz [3])



Obr. 11 Příklad parametrického vložení aplikace do interaktivního textu.

Závěrem lze konstatovat, že Matlab Web Server je určitě užitečná pomůcka pro oživení WWW stránek. Některé další možnosti využití Matlab Web Serveru pro analýzu dat, ale i jako didaktickou pomůcku při výuce lze najít na adrese

http://www.ft.vslib.cz/study/ma_study.htm,

kde umístěn odkaz interaktivní text „Řízení jakosti v textilu“.

Poděkování: Práce na tomto tématu byla podporována MŠMT - výzkumným záměrem č. MSM 245100303.

LITERATURA

- [1] ANDĚL, J. Matematická statistika. SNTL, 1985.
- [2] ANTOCH, J., VORLÍČKOVÁ D. *Vybrané metody statistické analýzy dat*. Academia, 1992.
- [3] Brzezina M. Internetová databáze příkladů. (<http://daemon.kap.vslib.cz/matematika>).
- [4] E-astat, (<http://www.emilea.de/>).
- [5] LINKA, A., PICEK J., VOLF P. *Teorie pravděpodobnosti*. Skripta TUL, 1998.
- [6] Matlab Web Server – The Language of Technical Computing, The MathWorks, Inc., 2000
- [7] MILITKÝ, J. *Statistické zpracování experimentálních dat*. East Publishing, 1998.
- [8] MMSTAT – An interactive introduction into the world of statistics.
(<http://www.quantlet.com/mdstat/mmstat.html>)
- [9] WETHERILL, G. B., BROWN D. W. *Statistical Process Control*. Chapman nad Hall, 1990.
- [10] TOŠENOVSKÝ J., NOSKIEČOVÁ D. *Statistické metody pro zlepšování jakosti*. Montex, 2000.

RNDr. Aleš Linka, CSc.
Katedra textilních materiálů
Technická univerzita v Liberci
Hálkova 6
e-mail: ales.linka@vslib.cz

Doc. Petr Volf, CSc
Katedra aplikované matematiky
Technická univerzita v Liberci
Hálkova 6
e-mail: petr.volf@vslib.cz