

FamStat

Statistická nadstavba systému Famulus

Jan Hanousek, Pavel Charamza, Marek Malý, Karel Zvára

Zatímco pojem *famulus* se nejčastěji překládá jako 'pomocník', 'sluha' či 'služebník', prostě někdo, kdo je stále po ruce a ochotný pomoci, označuje původně latinský pojem *statistika* obor lidské činnosti zabývající se zpracováním dat. Statistika je věda, ale také trochu umění generování či shromažďování dat, jejich popisu, analýzy, sumarizace a interpretace; statistika umožňuje rozšíření našich znalostí o datech a o populaci, z které data pocházejí. Spojení obou dvou pojmů v sobě skrývá FamStat - statistická nadstavba systému Famulus, Váš statistický pomocník a služebník. Představujeme Vám FamStat jako původní český ucelený systém, který umožňuje provádět statistické zpracování i značně rozsáhlých dat a názornou prezentaci výsledků ve formě tabulek a grafů. Je dostatečně srozumitelný pro ty, kteří jsou méně zbláhli ve statistické analýze, ale i dostatečně obsáhlý na to, aby se stal užitečným nástrojem v ruce profesionálního statistika. Toto integrované statistické prostředí zahrnuje jednak celobrazovkový datový editor, jednak systém menu umožňující snadnou aktivaci implementovaných procedur a funkcí. Statistická nadstavba umožňuje jednak interaktivní práci v systému pomocí menu s možností připojení vlastních procedur, jednak přímé výpočty pomocí modelů (vhodné zejména pro rutinní a opakované zpracování dat) a jednak programování vlastních výpočetních postupů (s výhodou použití již vytvořených procedur).

Ze zpracovávaných témat uveďme zejména: Editace a transformace dat (editor), základní popisné statistiky, průzkumová analýza dat, frekvenční tabulky (histogramy, polygony) i pro neekvidistantní dělení, distribuční a kvantilové funkce známých rozdělání, simulace náhodných veličin s různým rozdělením, základní parametrické a neparametrické testy, testy dobré shody, korelační analýza, regresní analýza, analýza rozptylu, časové řady (obecné klouzavé průměry, EDA, Boxova-Jenkisonova metodologie), kontingenční tabulky, faktorová analýza.

V dodatku lze nalézt seznam implementovaných procedur a funkcí (včetně jejich začlenění do knihoven). Věříme, že z anglické mnemoniky lze získat lepší představu o obsahu knihoven.

Statistický komplet důsledně využívá vlastností systému Famulus a díky tomu zejména:

- umožňuje práci s nedefinovanými hodnotami proměnné,
- má výraznou grafickou podporu (více grafů na jedné obrazovce, možnost jejich exportu v různých formátech, odečítání hodnot z grafu, atp.)
- pro snadnou orientaci uživatele využívá interaktivní kontextové nápovědy systému 'help' (pochopitelně v češtině),
- využívá možnosti psaní (a editace) zpráv o provedeném zpracování, (pomocí záznamníku) atd.

- velikost zpracovávaných dat je limitována pouze velikostí virtuálního souboru (a činí cca 30 MB), při velkých datových souborech rychlost zpracování záleží na velikosti EMS.
- je požadováno minimálně 640 kB RAM, HD, grafická karta CGA/Hercules/EGA/VGA, DOS 3.0 a vyšší, Famulus 3.5. Pro náročnější výpočty doporučujeme EMS a koprocesor.

(V dodatku můžete též nalézt stručnou charakteristiku systému FAMULUS.)

Jelikož je statistická nadstavba systému Famulus budována pro co nejširší užití v mnoha aplikačních oblastech, vzniká samozřejmě potřeba reagovat na požadavky z praxe doplněním či rozšířením FamStatu o nové možnosti, které nebudou zachyceny v původní verzi. Autoři si vyhrazují právo takovéto úpravy provádět a uvítají veškeré vaše podněty a připomínky. Budeme rádi, pokud se FamStat skutečně stane Vaším pomocníkem a přispěje ke zdárnému vyhodnocení Vašich dat.

Dodatky

A. Stručný popis knihoven

Knihovna:	Obsah:
ANOVA	- analýza rozptylu
APPQUANT	- aproximace kvantilů některých rozdělení
CHARSORT	- abecední řazení (třídění) řetězců
CHIFIT	- chí-kvadrát test dobré shody
COMBI	- kombinatorika
COMPO	- analýza hlavních komponent
CONTI	- čtyřpolní a kontingenční tabulky
CORREL	- korelační analýza
DEFAULT	- pomocné procedury a funkce, globální konstanty
DESCRIPT	- základní popisné statistiky
DISTR	- distribuční funkce a hustoty rozdělení
EDA	- průzkumová analýza dat
EIGEN	- vlastní čísla a vektory symetrické matice
FACTOR	- faktorová analýza
FIT	- testy dobré shody a normality
FREQ2	- dvojrozměrné frekvenční tabulky a histogramy
FREQHIST	- frekvenční tabulky a histogramy
GENTLE	- ortogonální rotace, inverze matic
GRAPHPLT	- rozptylové diagramy
LIRA	- lineární regresní model
MAHA	- posouzení vícerozměrné normality
NONPARAM	- neparametrické testy
PARAMET	- parametrické testy
PARESTIM	- odhady parametrů rozdělení
QUANTILE	- kvantily rozdělení
RESIDS	- zpracování reziduí v regresních modelech
ROTATE	- rotace dat a histogramů
SHAPIRO	- Shapiro-Wilkův test normality
SIMUL	- generování náhodných čísel z různých rozdělení
SMOOTH	- vyhlazování časových řad
SPECTRAL	- spektrální analýza časových řad
SPREAD	- test shody více výběrů
TESTHELP	- aproximace dosažené hladiny významnosti

B. Stručný seznam procedur a funkcí FamStatu

Knihovna ANOVA

AnalysisOfCoVarBase
AnalysisOfCovarStatist
BaseOfDiagnostics
BaseOfTable
CompareTermsB
DecompositionOfSS
DecompositionOfSSReg
DenominatorTermB
FactorLevelsComparison
FindTermsExpectations
Friedman
GetModel
IndInv
KruskalWallis
OneTermExpectation
Ranks
SetIndex
SetRow
TermDescription
TermExpectation
TotalSum
TwoWay

Knihovna APPQUANT

ApproxChiInv
ApproxFiInv
ApproxNormalInv
ApproxStudentInv

Knihovna CHARSORT

CharSort2
CharSort
CompareStrings

Knihovna CHIFIT

EqualExpFreq
ExpectedFreq
FitChi2TestKnihovna

COMBI

Comb
CombWithRep
Perm
PermWithRep

Knihovna COMPO

PrincipalComponentsCorrel
PrincipalComponentsCovar
ProportionCorrel
ProportionCovar
ReducedCMat

Knihovna CONTI

ArmitageTrendTest
AssocChar
BowkerSymmetryTest
ContiIndepTest
ContiSums
Crosstabulation
DandekarTest
FisherExactTest
FixedMarginalTest
FourfoldIndepTest
FreemanTukeyChi2
GartFTTest
IntOdds
IntRelativeRisk
LeRoyEquivTest
LikelihoodChi2
MantelHaenszelTest
McNemarTest
MinExpectedFreq
OddsRatioTest
PearsonChi2
PearsonChi2EC
PhiCoefTest
StuartSymmetryTest
YatesTest

Knihovna CORREL

Biserial
ConcorCoef
CornerTest
CorrelMat
CorrMat
Covariance
CovarMat
Fisher_z
Fisher_zInv
HotellingF
IndependenceTest
KendallTest
MultipleCCEqualityTest
MultipleCorrelCoefTest
PartialCorrelCoefTest
Pearson
RankingList
rTest
SpearmanTest
Ties
z2CommonTest
z2Test
zkTest1
zkTest2
zTest

Knihovna DEFAULT

CompString
DelLeftBlanks
DelRightBlanks
ErrorCode
Halt
iMax
iMin
InIt
iSwap
iVal
rMax
rMin
round
SetBit
SetUp
Significance
TestBitB

Knihovna DESCRIPT

AverageMinMax
Modus
Moments2_3_4
PercentIn1_2_3_StdDev
Quantiles
SortData
Stat
Sum
WhiskerLength

Knihovna DISTR

ALTER, alter, AlterRnd
BETA, beta, BetaRnd
betacf
BINOM, binom, BinomRnd,
CAUCHY, cauchy,
CauchyRnd
chi, CHI, ChiRnd
EXPONENTIAL,
exponential, ExpRnd
F, f, FRnd
fi, FI, FiRnd
gama, GAMA, GamaRnd
gamp
gcf
geom, GEOM, GeomRnd
gser
LnGama
normal, NORMAL,
NormalRnd
poisson, POISSON,
PoissonRnd
student, STUDENT,
StudentInv
uniform, UNIFORM,
UniformRnd

Knihovna EDA

BoxWhisker2
 BoxWhisker
 DetectPos1
 DimenH
 DrawLine
 Hann
 Median
 PlotArrayLine
 RunningMedHelp
 RunningMedRep
 Split
 SteamLeaf
 TableGraph
 Transform
 TwoWayTab

Knihovna EIGEN

Eigsrt
 Jacobi
 TranspSqMa

Knihovna FACTOR

Communality
 FactorLoadings
 FactorScoresCorrel
 FactorScoresCovar
 NumberOfGtValue
 OrthogonalRotation
 ProportionCorrel
 ProportionC
 Uniquen

Knihovna FIT

AndersonDarlingTest0
 AndersonDarlingTestE
 AndersonDarlingTestN
 BasicStats
 BiweightNormalityTest
 Cramer_vonMisesTest0
 Cramer_vonMisesTestE
 Cramer_vonMisesTestN
 KS1Test
 KS2Test
 KS
 KSPlot
 KurtosisTest
 OmnibusTest
 QQdiagram
 QuickKurtosisTest
 RatioTest
 SkewnessTest
 TukeyFitTest
 WatsonTest0
 WatsonTestE
 WatsonTestN

Knihovna FREQ2

Freq2
 Histogram3D
 HistogramPolygon3D
 Polygon3D

Knihovna FREQHIST

ChangeClassesRepresentati
 on
 FindClasses
 Freq
 Histogram
 HistogramPolygon
 Polygon

Knihovna GENTLE

Cij
 CijSub
 ConFound
 Escalator
 Gii
 ijSym
 IncludeRow
 InverzeV
 InverzeVSub
 Regress
 RegressSub
 SSDecomp
 SwapRows

Knihovna GRAPHPLT

PlotArrays
 PlotExprs

Knihovna LIRA

LinearRegression
 - " - BaseOfDiagnostics
 LinearRegressionEstimates
 LinearRegressionStatistics
 RestoreEstimates
 SubmodelTesting

Knihovna MAHA

Mahalanobis
 MultiKurtosisTest
 MultiSkewnessTest
 PlotNormalEllipse
 QD2diagram

Knihovna NONPARAM

AnsariBradleyTest
 DixonTest
 IterationTest
 MannWhitneyTest
 RunUpDownTest
 SignTest
 WilcoxonTest

Knihovna PARAMET

Chi2Test
 FTest
 TestHelpH
 TestMi
 TTest
 TwoSample

Knihovna PARESTIM

AlterE
 BasicEstim
 BetaE
 BinomE
 CauchyE
 ChiE
 ExponentialE
 FE
 GamaE
 GeomE
 NormalE
 PoissonE
 StudentE
 UniformE

Knihovna QUANTILE

BetaInv
 BinomInv
 BInv
 CauchyInv
 Chi2Inv
 ExpInv
 Filnv
 FInv
 Gamalnv
 GeomInv
 NormalInv
 PoissonInv
 StudentInv

Knihovna RESIDS

DurbinWatson
 PartialResiduals
 Residuals
 Sf

Knihovna ROTATE

Rotate
 RotateAndDo
 RotateAndDo2
 RotateGroups
 RotatePoints

Knihovna SHAPIRO

ExpNormScores
 PolynomValue
 ShapiroWilkTest

Knihovna SMOOTH

AditWinters1
AditWinters2
AditWintersSSEH
CenterMovingAve
ComputeWeightsH
CreateWeightMatrix2H
CreateWeightMatrixH
DoubleExp1
DoubleExp2
DoubleExpSSEH
InitAditWintersH
InitMultWintersH
LinRegrEquiH
MovingAverages2H
MovingAveragesH
MovingAverages
MultiWinters1
MultiWinters2
MultiWintersSSEH
QRegrEquiH
SimpleExp1
SimpleExp2
SimpleExpSSEH
SumPower
TripleExp1
TripleExp2
TripleExpSSEH

Knihovna SPECTRAL

ARCov
ARMACov
ARMA SpecDens
ARSpecDens
Autocor2
AutoCor
AutoCov
DetermH
FisherTest2
FisherTest
Invertible
MACov
MASpectDens
Parzen
Periodogram2
Periodogram
SpecDensEst
SpecDensEstimH
Stationary
SumOfExpH
Knihovna SPREAD
GenSiegelTukeyTest
Knihovna TESTHELP
ApproxProbabValue1
ApproxProbabValue2
SplineTab

C. Stručný souhrn vlastností a možností systému Famulus

- **Výpočty:** Přesnost výpočtů je 19 platných míst, rozsah čísel přes 10^{-1000} . K dispozici jsou vícerozměrná pole, komplexní čísla, proměnné a funkce, standardní funkce zahrnují funkce trigonometrické, cyklometrické, exp, ln, log a další. Výpočetní možnosti jsou rozšiřitelné pomocí aplikačních knihoven.
- **Práce s daty:** Famulus umožňuje čtení numerických hodnot z ASCII souborů ve značně flexibilních formátech, psaní na disk, zadávání z klávesnice s nápovědou hodnot, editaci jednorozměrných i dvourozměrných polí a další operace.
- **Výstup výsledků** může mít formu až 4 okének na obrazovce. V každém může být graf, tabulka ev. text modelu či záznamníku. Parametry předdefinovaných grafů a tabulek se nastavují interaktivně.
- **Grafika:** Jednoduše lze zobrazovat grafy funkcí, křivky, včetně více křivek v grafu. K dispozici je 16 druhů značek, volitelné barvy, pohyblivé značky umožňující prvky animace, grafické procedury a funkce, 3D-grafika, volba fontů a velikosti písma při psaní do grafů.
- **Tisk výsledků:** Vytvořené grafy lze tisknout v kvalitě omezené pouze rozlišením tiskárny. Výstupní zařízení zahrnují černobílé i barevné tiskárny 9- a 24-jehličkové (Epson a kompatibilní, IBM Proprinter a další), tryskové (DeskJet, PaintJet, ...) a laserové kompatibilní s HP-LaserJet i postscriptové.

- **Výstup výsledků na plotter:** ve formátu HP-GL na řadu plotterů (včetně Colorgrafu Aritma).
- **Export grafů do souborů** umožňuje jejich pohodlné začleňování do dalších dokumentů, zpracování DTP programy apod. K dispozici je řada pokrývající většinu běžných vektorových i rastrových formátů (převážně jak v černobílých, tak v barevných verzích): ANSI CGM, AutoCad DXF, Gem IMG, HPGL (7475), Postscript, TIFF, PCX, VideoShow (ANSI), WordPerfect WPG, PCX a BMP (Windows).
- **Ovládání systémem menu** s podporou myši a okamžitou nápovědou ev. pomocí 'hotkeys'. Práci usnadňuje kontextově závislý interaktivní systém 'help'. Ovládání, nápovědy i vysvětlující informace jsou samozřejmě v češtině.
- **Programování a ladění:** Famulus poskytuje univerzální jazyk orientovaný na numerické výpočty, velmi blízký Pascalu a Fortranu. Zahrnuje i procedury a funkce s možností rekurze, parametry typu pole a lokální pole s proměnnými mezemi, zobecnění parametrů typu funkce a procedura a další vylepšení pro numerickou práci (zejména práce s nedefinovanou hodnotou). Integrovaný debugger umožňuje mimo jiné nastavit kontrolní body, krokovat výpočet a sledovat hodnoty proměnných a výrazů (se současným zobrazením grafů).
- **Interaktivní práci** usnadňuje záměrný kříž pro odečítání výsledků z grafů a kalkulátor pro výpočet výrazů, zjišťování hodnot proměnných a interaktivní zadávání příkazů. Kalkulátor umožňuje využívat i funkce a algoritmy z knihoven.
- **Základní aplikační knihovny** dodávané se systémem zahrnují: fyzikální konstanty a převody jednotek, řešení soustav rovnic, maticový počet, integrály, řešení obyčejných diferenciálních rovnic, hledání minim a nulových bodů funkcí, speciální funkce, rychlou Fourierovu transformaci, interpolaci, třídění, generování náhodných čísel, jednoduchou statistiku, dvou- a třírozměrnou grafiku. Se systémem je dodávána i knihovna pro tvorbu uživatelských menu a dialogových boxů.
- **Rozšiřitelnost:** Všechny knihovny jsou ve zdrojových textech a uživatel je může prohlížet a upravovat. Může též vytvářet vlastní knihovny zaměřené na svou aplikační oblast.
- **Programování a práce s knihovnamí:** práce s řetězci a poli řetězců, další grafické procedury, ukládání přeložených knihoven na disk (při pozdějším linkování se již nepřekládá ze zdrojového textu).
- **Práce s pamětí:** využití rozšířené paměti (EMS), virtuální paměť na disku (do max. rozsahu 32 MB).

System Famulus vyžaduje minimálně 640 kB RAM, HD, graf.karta CGA/Hercules/EGA/VGA, DOS 3.0 a vyšší. Pro náročnější výpočty doporučujeme EMS a koprocesor.