2
ODHADY PARAMETROV ZÁKLADNÉHO SÚBORU

2.1 Bodové odhady

Každý zákon rozdelenia pravdepodobnosti diskrétnej aj spojitej náhodnej premennej závisí od jedného alebo viacerých **parametrov**. V praxi často hľadáme vhodný pravdepodobnostný model rozdelenia určitej náhodnej premennej, napr. počtu poistných plnení pri poistení motorových vozidiel, výšky príjmov domácností a pod. Často máme rozumné dôvody predpokladať určitý typ funkcie pravdepodobnosti diskrétnej, alebo hustoty pravdepodobnosti spojitej premennej. Pretože obyčajne nepoznáme vektor parametrov $\boldsymbol{\Theta}$ predpokladaných pravdepodobnostných rozdelení, musíme ich odhadnúť pomocou náhodného výberu $\left(X_1, X_2, \ldots\ldots, X_n\right)$.

BODOVÝ ODHAD

Bodový odhad parametra Θ spočíva v jeho nahradení hodnotou vhodne zvolenej výberovej charakteristiky U_n, čo symbolicky zapisujeme:

$$U_n = est\,\Theta = \hat{\Theta}$$

2.1.1 Vlastnosti bodových odhadov

Pre odhad parametra Θ sa snažíme zvoliť takú výberovú charakteristiku, ktorá čo najlepšie aproximuje hodnotu Θ a poskytuje tak jej „najkvalitnejší“ odhad.

Základné vlastnosti, ktoré majú zabezpečiť „kvalitu“ bodového odhadu parametra Θ, sú tieto:

- neskreslenosť (nevychýlenosť) odhadu,
- konzistentnosť odhadu,
- výdatnosť odhadu,

- dostatočnosť odhadu,
- robustnosť odhadu.

NESKRESLENÝ ODHAD

Výberová charakteristika U_n je neskresleným odhadom parametra Θ, ak sa jej stredná hodnota rovná skutočnej hodnote odhadovaného parametra, teda:

$$E(U_n) = \Theta \tag{2.1}$$

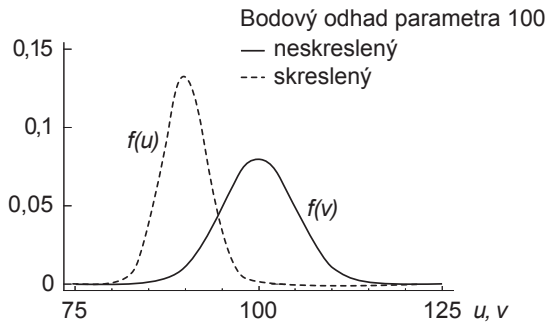
Skreslením Δ_n bodového odhadu parametra Θ pomocou výberovej charakteristiky U_n je rozdiel jej strednej hodnoty $E(U_n)$ a skutočnej hodnoty parametra Θ, teda:

$$\Delta_n = E(U_n) - \Theta$$

Ak je výberová charakteristika U_n neskresleným odhadom parametra Θ, zrejme pre skreslenie platí $\Delta_n = 0$.

Výberová charakteristika U_n je **asymptoticky neskresleným odhadom** parametra Θ, ak platí:

$$\lim_{n \to \infty} E(U_n) = \Theta \tag{2.2}$$



Obr. 2.1 **Bodový odhad parametra Θ = 100**

Zdroj: Vlastné spracovanie.

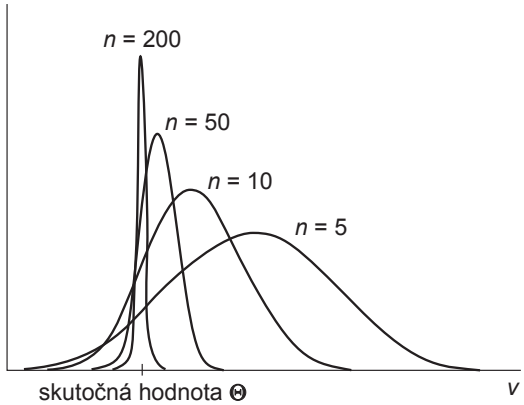
Na obr. 2.1 je znázornený odhad parametra základného súboru Θ = 100 pomocou dvoch výberových charakteristík *U*, *V*. Platí *E*(*U*) = 95 a *E*(*V*) = 100, preto neskresleným odhadom parametra Θ je výberová charakteristika *V*.

KONZISTENTNÝ ODHAD

Výberová charakteristika U_n je **konzistentným odhadom** parametra Θ, ak platí:

$$\lim_{n\to\infty} P\left(\left|U_n - \Theta\right| < \varepsilon \right) = 1 \tag{2.3}$$

teda ak sa so zväčšovaním rozsahu n výberového súboru výberová charakteristika U_n takmer isto líši od parametra Θ menej ako o ľubovoľne malé kladné číslo ε.



Obr. 2.2 **Vlastnosť konzistencie bodových odhadov**

Zdroj: Vlastné spracovanie.

Na obr. 2.2 vidíme, ako sa s rastúcim n rozdelenie výberovej charakteristiky V sústreďuje okolo skutočnej hodnoty parametra Θ. S rastúcim rozsahom výberového súboru n preto rastie aj pravdepodobnosť, že sa hodnota v bodového odhadu bude málo líšiť od skutočnej hodnoty Θ.

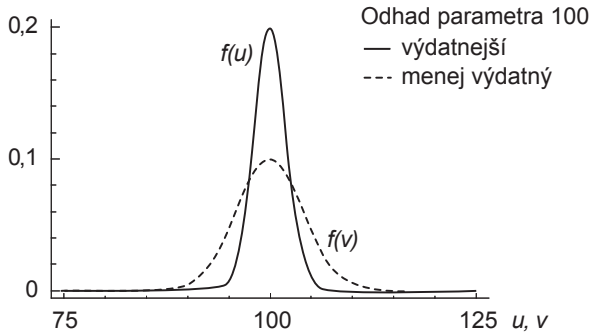
VÝDATNÝ ODHAD

Výdatným odhadom parametra Θ nazývame takú výberovú charakteristiku $\hat{U}_n$, ktorá zo všetkých výberových charakteristík U_n poskytujúcich neskreslený bodový odhad parametra Θ má *najmenší rozptyl*.

Ak z nejakých dôvodov nemôžeme určiť takú výberovú charakteristiku $\hat{U}_n$, ktorá je výdatným odhadom parametra Θ, odhadujeme tento parameter výberovou charakteristikou U_n, ktorá sa najviac blíži výdatnému odhadu. **Mierou výdatnosti** tejto charakteristiky je pomer:

$$e\left(U_n\right) = \frac{D\left(\hat{U}_n\right)}{D\left(U_n\right)} \tag{2.4}$$

Podľa obr. 2.3 je výberová charakteristika U s rozptylom $\sigma_U^2 = 4$ zrejme výdatnejším odhadom parametra Θ = 100 ako výberová charakteristika V s rozptylom $\sigma_V^2 = 16$.



Obr. 2.3 **Bodové odhady *U*, *V* s rôznou mierou výdatnosti**

Zdroj: Vlastné spracovanie.

Čím viac sa miera výdatnosti blíži k hodnote 1, tým je odhad parametra Θ pomocou výberovej charakteristiky U_n výdatnejší. Odhad, pre ktorý platí:

$$\lim_{n\to\infty} e\left(U_n\right) = 1 \tag{2.5}$$

sa nazýva **asymptoticky výdatným odhadom**.

ROBUSTNÝ ODHAD

Výberová charakteristika U_n je robustným odhadom parametra Θ, ak jej rozdelenie nie je ovplyvnené nedodržaním základných predpokladov, resp. jej výberové rozdelenie nie je citlivé na zmenu východiskových predpokladov.

DOSTATOČNÝ ODHAD

Výberová charakteristika U_n je dostatočným odhadom parametra Θ, ak využíva všetky informácie z náhodného výberu, ktoré majú vzťah k odhadovanému parametru Θ. Teda všetky poznatky o parametri Θ, ktoré je možné získať z jednotlivých hodnôt výberového súboru, poskytuje aj samotná výberová charakteristika U_n.

2.1.2 Overovanie vlastností bodových odhadov

Vlastnosť konzistentnosti overíme pomocou postačujúcej podmienky konzistentnosti bodových odhadov (pozri Bakytová – Hátle – Novák – Ugron, 1986, s. 61-62).

POSTAČUJÚCA PODMIENKA KONZISTENTNOSTI

Výberová charakteristika U_n je konzistentným odhadom parametra Θ, ak má tieto vlastnosti:

- má konečný rozptyl, teda $D(U_n) < \infty$ pre všetky n,
- je to neskreslený, alebo aspoň asymptoticky neskreslený odhad parametra Θ,
- pre $n \to \infty$ rozptyl výberovej charakteristiky $D(U_n)$ konverguje k nule, t. j.

$$\lim_{n\to\infty} D(U_n) = 0$$

Vlastnosť výdatnosti overujeme pomocou tzv. **Rao-Cramérovej nerovnosti**.

RAO-CRAMÉROVA NEROVNOSŤ

Pre každý neskreslený bodový odhad U_n parametra Θ platí nerovnosť:

$$D(U_n) \geq \frac{1}{n \cdot E\left(\left(\frac{\partial \ln f(x;\Theta)}{\partial \Theta}\right)^2\right)} \tag{2.6}$$

Z tejto nerovnosti a z definície výdatných bodových odhadov bezprostredne vyplýva, že výberová charakteristika $\boldsymbol{U_n}$ **je výdatným bodovým odhadom** parametra Θ **práve vtedy**, ak vo vzťahu (2.6) dostaneme rovnosť:

$$D(U_n) = \frac{1}{n \cdot E\left(\left(\frac{\partial \ln f(x;\Theta)}{\partial \Theta}\right)^2\right)} \tag{2.7}$$

2.1.3 Bodový odhad strednej hodnoty μ základného súboru

Strednú hodnotu μ základného súboru, alebo presnejšie strednú hodnotu rozdelenia pravdepodobnosti sledovanej náhodnej premennej X odhadujeme **pomocou výberového priemeru** $\bar{X}$.

Overíme požadované vlastnosti bodových odhadov pre túto výberovú charakteristiku. Zo vzťahu (1.6) vyplýva, že výberový priemer $\bar{X}$ **je neskresleným odhadom** strednej hodnoty μ pre ľubovoľné rozdelenie pravdepodobnosti náhodnej premennej X.

Pretože prvé dve vlastnosti konzistentnosti sú splnené a podľa (1.7) platí:

$$\lim_{n\to\infty} D\left(\bar{X}\right) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0$$

výberový priemer $\bar{X}$ **je konzistentným odhadom** strednej hodnoty μ.

Ukážeme pomocou vzťahu (2.7), že výberový priemer $\bar{X}$ je výdatným odhadom strednej hodnoty základného súboru, ak $X \sim Po(\lambda)$. V takomto prípade platí:

$$\mu = E(X) = \lambda \text{ a } f(x;\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$$

Postupnosťou krokov riešenia pravej strany rovnosti (2.7) dostaneme:

$$\ln f(x;\lambda) = x \ln \lambda - \ln x! - \lambda$$

$$\frac{\partial \ln f(x;\lambda)}{\partial \lambda} = \frac{x}{\lambda} - 1 = \frac{x-\lambda}{\lambda}$$

$$E\left(\left[\frac{x-\lambda}{\lambda}\right]^2\right) = \frac{1}{\lambda^2} E\left(\left[x-\lambda\right]^2\right) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

Po dosadení do (2.7) pravá strana tejto rovnosti je $\frac{\lambda}{n}$.

Pretože $X \sim Po(\lambda)$, platí $D(X) = \lambda$ a podľa (1.7) pre ľavú stranu vzťahu (2.7) dostaneme tiež:

$$D\left(\bar{X}\right) = \frac{\lambda}{n}$$

Tým sme dokázali, že $\bar{X}$ **je výdatným odhadom parametra λ Poissonovho rozdelenia**.

Predpokladajme, že v základnom súbore $X \sim N\left(\mu;\sigma^2\right)$. Ukážeme pomocou platnosti vzťahu (2.7), že aj v tomto prípade je $\bar{X}$ výdatným odhadom strednej hodnoty μ.

6
RIEŠENÉ PRÍKLADY V SYSTÉME SAS

6.1 Úvod do práce v systéme SAS Enterprise Guide

Štatistický analytický systém **SAS** (*Statistical Analytical System* – SAS) je softvérový produkt firmy SAS Institute Inc., Cary, NC 27513, USA. Softvér vznikal v rokoch 1966 – 1976 ako vedecký projekt na Štátnej univerzite v Severnej Karolíne (*North Carolina State University*). Spoločnosť SAS založili v roku 1976 **Jim Goodnight**, **John Sall** a **Jim Barr**, pracovníci tejto univerzity. Z malej kancelárie so štyrmi zamestnancami sa stala časom veľká spoločnosť s takmer 13 700 pracovníkmi. SAS má dnes zákazníkov v 140 krajinách sveta v rôznych inštitúciách a organizáciách, kde im pomáha zbierať najdôležitejšie údaje a tieto transformovať do využiteľných strategických informácií o zákazníkoch, o dodávateľoch a o vlastnej organizácii. Na Slovensku vznikla SAS pobočka[1] v roku 1995 a v súčasnosti má asi 50 zamestnancov. SAS Slovakia, s. r. o. ponúka na slovenskom trhu komplexné konzultačné služby, technickú podporu a školenia.

Systém SAS je rozsiahly, profesionálne orientovaný softvérový systém, ktorý sa skladá z viac ako 200 modulov (častí, komponetov). **SAS Enterprise Guide** (ďalej **SAS EG**) je grafické užívateľské rozhranie (*Graphic User Interface* – GUI) pre systém SAS, ktoré bolo dodané do verzie 9 v roku 2004. Je to interaktívny nástroj k systému SAS a bez základného systému SAS nemôže fungovať. Pri nákupe SAS EG si treba kúpiť licenciu minimálne na základný modul SAS-u (Base SAS) a pre štatistické analýzy na modul SAS/STAT. Sú to moduly, ktoré sú potrebné na riešenie príkladov z tejto kapitoly.

Použitie interaktívneho rozhrania SAS EG nevyžaduje znalosť programovania, čiže znalosť syntaxe SAS programovacieho jazyka. SAS kód (program) je v SAS EG automaticky generovaný na základe práce používateľa v interaktívnych oknách úloh, čiže v ponuke (menu). Po spustení analýzy je SAS kód vykonaný na pozadí a je takisto súčasťou jednotlivých SAS výstupov. Užívateľ sa tak môže oboznámiť s obsahom a štruktúrou vykonaných SAS kódov, a tak sa učiť aj programovať v SASe. Nezáleží na tom, či SAS kód pre analýzu údajov napíšeme a spustíme v programovom režime, alebo túto ana-

[1] SAS Slovakia, s. r. o., Lazaretská 12, 811 08 Bratislava 1. Dostupné na internete: <http://www.sas.com/offices/europe/slovakia/contact/>.

lýzu „naklikáme“ v SAS EG, dostaneme rovnaké výsledky. SAS programový kód napísaný v prostredí systému SAS môžeme bez problémov použiť v aplikácii SAS EG. Je tu plná kompatibilita pri dodržaní niektorých základných pravidiel umiestnenia SAS dátových súborov.

Základnou výhodou systému SAS v porovnaní s inými podobnými softvérmi je to, že vie priamo pracovať s dátovými súbormi rôznych typov. Aj aplikácia SAS EG umožňuje jednoduchý prístup k lokálnym alebo vzdialeným SAS údajom (SAS dátové súbory verzie 6, 8 a 9) a aj k lokálnym alebo vzdialeným údajom iného typu (napr. Microsoft Excel, Microsoft Access, Lotus, Paradox, Text, HTML, ODBC, k tabuľkám z databáz ako Oracle, DB2, OLE DB a pod.).

Táto časť je zameraná na obsluhu SAS Enterprise Guide, verzie 4.2 až 5.1. Ponuka (menu) a obrazovky pre iné (staršie, resp. novšie) verzie SAS EG môžu byť v niečom odlišné. Obsahom nie je kompletný návod na obsluhu SAS EG, ale len výber niektorých častí, ktoré považujeme za potrebné bližšie vysvetliť vzhľadom k ich použitiu pri riešení príkladov v tejto kapitole. Užívateľ môže využiť aj systém ***Help***, ktorý je pre jednotlivé časti SAS EG podrobne vypracovaný. Na webovej stránke spoločnosti SAS je dostupná SAS dokumentácia, ktorá je takisto výborným zdrojom informácií[2].

6.1.1 Pracovná plocha SAS Enterprise Guide

SAS EG ako interaktívne rozhranie k systému SAS je samostatná aplikácia, ktorá sa inštaluje pod operačným systémom Windows. V SAS EG užívateľ vytvára svojou interaktívnou prácou SAS EG projekt. Projekt je základná štruktúra v SAS EG, s ktorou užívateľ pracuje. Každá nová analýza môže predstavovať nový SAS EG projekt, ktorý možno uložiť ako súbor s príponou egp (*nazov_projektu.egp*)[3], neskôr otvoriť a pokračovať v práci.

Aplikáciu SAS EG spustíme tak, že vyhľadáme tento program cez Štart menu operačného systému Windows. Otvorí sa úvodné okno, ktoré umožňuje (obr. 6.1):

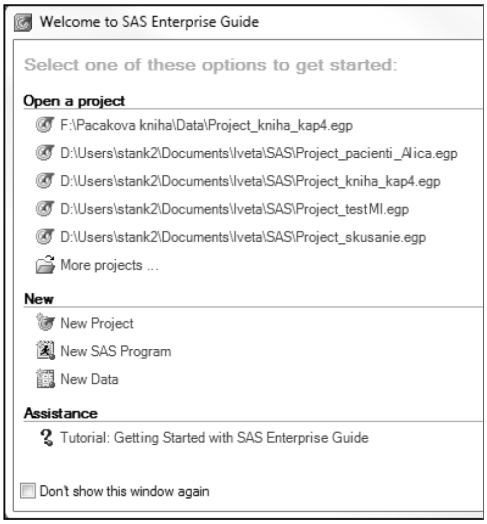
1. Vybrať SAS EG projekt zo zoznamu existujúcich projektov. Názvy existujúcich projektov, s ktorými užívateľ pracoval na svojom PC, sú zobrazené v časti *Open a project*.
2. Vytvoriť nový projekt (voľba *New Project* v časti *New*).
3. Vytvoriť nový SAS programový kód (voľba *New SAS Program* v časti *New*).
4. Vytvoriť nový dátový súbor (voľba *New Data* v časti *New*).
5. V časti *Assistance* ponúka výukový program pre SAS EG (*Tutorial Getting Started with SAS Enterprise Guide*).
6. V prípade, že užívateľ nechce, aby sa úvodné okno zobrazovalo, označí voľbu *Don't show this window again* v spodnej časti okna (obr. 6.1).

Po výbere voľby *New Project* sa na obrazovke objavia tri základné okná prostredia SAS EG (obr. 6.2): *Project Tree, Workspace, Server List.*

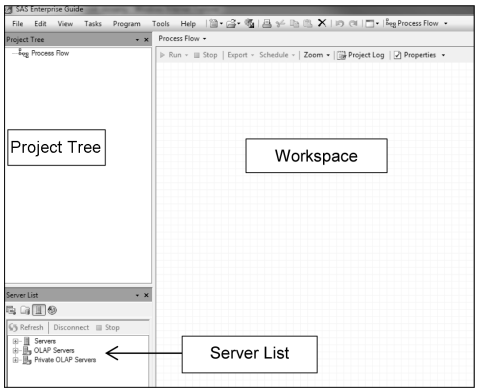
---

[2] SAS Online Documentation. Dostupné na internete: <http://support.sas.com/onlinedoc/913/docMainpage.jsp>.

[3] Projekty v SAS EG od verzie 4.0 majú príponu *.*egp*. V starších verziách SAS EG mali projekty príponu *.*seg*.



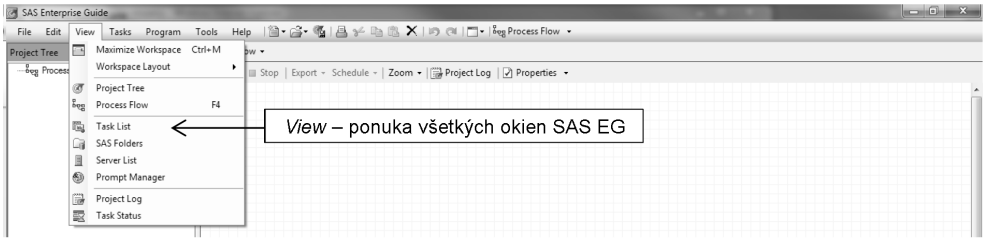
Obr. 6.1 **Úvodné okno SAS EG**



Obr. 6.2 **Základné okná v SAS EG**

Na začiatku je pre nový práve otvorený SAS EG projekt okno *Project Tree* prázdne (pozri obr. 6.2). Postupne sa tu ukladajú jednotlivé prvky analýzy v stromovej štruktúre, ktoré užívateľ interaktívne tvorí cez ponuku SAS EG. Okno obsahuje všetky vstupné i výstupné dáta, SAS programy, chybové hlásenia o spustených analýzach a výsledky analýz, ktoré boli vytvorené alebo použité v SAS projekte. Vytvorený projekt užívateľ prvýkrát pomenuje a uloží postupnosťou krokov: *File – Save Project As.*

V pracovnom prostredí (*Workspace*) sa automaticky po naštartovaní SAS EG otvorí okno *Project Designer*, ktorého súčasťou je záložka *Process Flow.* Je to grafické zobrazenie priebehu projektu prostredníctvom diagramu s ikonami. V pracovnom prostredí sú zobrazované aj výsledky jednotlivých analýz po vykonaní SAS úloh, resp. programov zvyčajne v HTML formáte alebo ako SAS Reporty. Môžu tu byť zobrazené aj obsahy jednotlivých systémových hlásení (*Log*) alebo SAS kódov (*Code*).



Obr. 6.3 **Okná SAS EG – ponuka voľby *View***

Všetky okná sa dajú ľubovoľne posúvať, dá sa meniť ich veľkosť a prípadne ich zatvárať, resp. otvárať. Okná zatvárame cez „krížik“ umiestnený v ich pravom hornom rohu a spätne otvárame cez hlavnú ponuku, záložka *View* (obr. 6.3), kliknutím na príslušný

názov okna, ktoré chceme otvoriť. Okrem troch úvodných okien sa v ponuke *View* nachádzajú aj ďalšie okná: *Task List, Project Log, Task Status*.

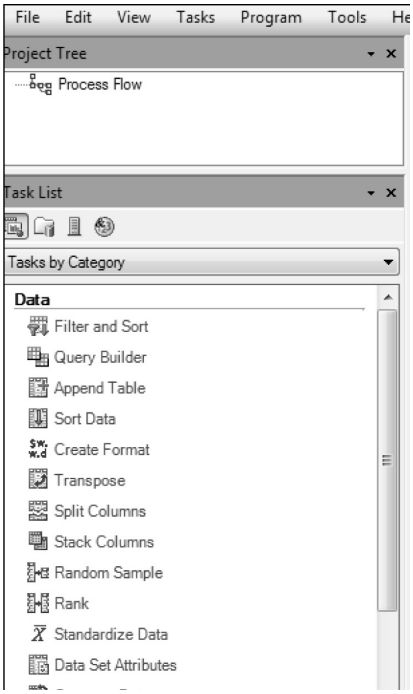
Okno *Task List* umožňuje otvorenie dialógu úloh (procedúr), ktoré poskytuje SAS Enterprise Guide. Jednotlivé úlohy obsahujú ponuku a nastavenia pre konkrétne analýzy, ktoré možno spúšťať aj cez hlavný ponukový riadok. Je možné využívať záložku podľa jednotlivých kategórií analýz, podľa ich zamerania (*Tasks by Category,* obr. 6.4) alebo podľa názvov SAS procedúr, ktoré jednotlivé úlohy reprezentujú (*Tasks by Name,* obr. 6.5). Druhú možnosť uvítajú užívatelia, ktorí poznajú SAS procedúry a ich jazyk. Použitie jednotlivých úloh (procedúr) závisí od konkrétnej štruktúry licencovaných SAS modulov systému SAS, ktoré aplikácia SAS EG využíva.

Okno *Task Status* zobrazuje informácie o momentálnom stave vykonávania jednotlivých úloh, ktoré boli spustené na vykonanie.

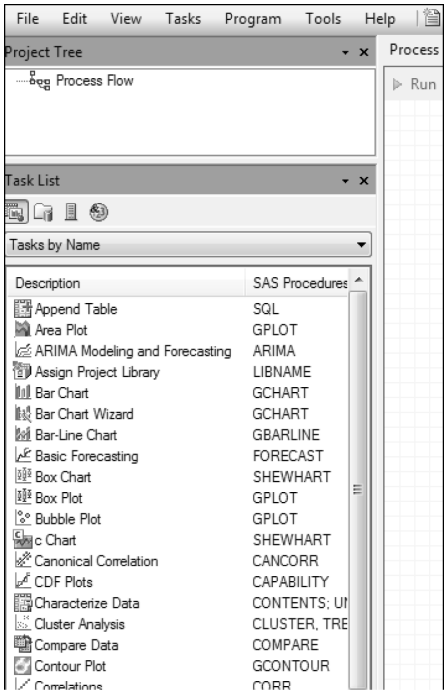
Okno *Project Log* zobrazuje systémové informácie o priebehu vykonania jednotlivých úloh (procedúr), ktoré užívateľ poslal na vykonanie pomocou tlačidla *Run*.

Nové aj existujúce projekty sa dajú načítať dvoma spôsobmi:

1. Ak začíname prácu so systémom SAS EG, zvolíme príslušnú voľbu pre načítanie nového alebo existujúceho projektu spomínanú v predchádzajúcom texte (pozri obr. 6.1).
2. Ak sa už nachádzame v SAS EG, tak zvolíme nasledujúcu postupnosť príkazov:
   a) ak ide o nový projekt: *File – New – Project;*
   b) ak ide o existujúci projekt: *File – Open – Project – Local Computers* alebo *SAS Servers.*



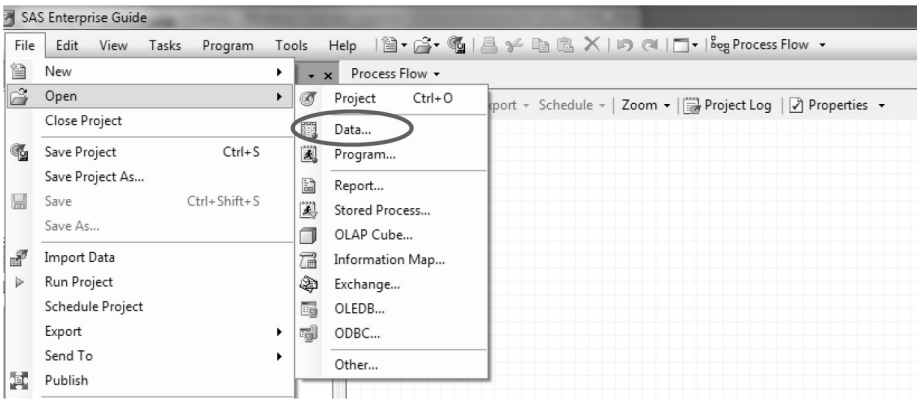
Obr. 6.4 ***Task List – Tasks by Category*** **(časť)**



Obr. 6.5 ***Task List – Tasks by Names*** **(časť)**

6.1.2 Práca s dátami v projekte

Vloženie SAS dátových súborov (prípona **.sas7bdat,* resp. staršie verzie ako 9 mali príponu **.sd2*) do projektu sa dá urobiť viacerými spôsobmi. Najjednoduchší spôsob je výber z hlavného ponukového riadku (obr. 6.6): *File – Open – Data – Local Computer* alebo *SAS Servers.*



Obr. 6.6 **Vloženie dátového súboru do EG projektu (*File – Open – Data*)**

Po vyhľadaní súboru na lokálnom počítači odklikneme *Open*. Súbor je pridaný do projektu a stáva sa automaticky aktívnym vstupným súborom. Ihneď po pridaní do projektu sa súbor štandardne otvorí v okne *Workspace* a je chránený pred prepisovaním, čiže je v stave *Protect data (read only)*. V projekte (okno *Project Tree*) alebo v okne *Process Flow* je SAS dátový súbor reprezentovaný štvorcovou ikonkou s červenou guličkou v pravom dolnom rohu (pozri napr. obr. 6.9).

Ďalší spôsob otvorenia existujúceho SAS dátového súboru je prostredníctvom tzv. SAS knižnice. SAS knižnica (*Library*) je v podstate predvolená fyzická cesta k súboru na disku či inom médiu. Vo verziách SAS EG 4.2 až 5.1 možno knižnice vytvárať priamo cez hlavnú ponuku *Tools – Assign Project Library* (existujú aj iné možnosti).

Objaví sa dialógové okno, ktoré umožní zadať názov knižnice (*Name,* obr. 6.7), vybrať server alebo lokálny disk (podľa možností inštalácie SAS EG) a fyzickú cestu (*Path,* obr. 6.8), kde sa táto knižnica bude nachádzať. Pri tvorbe knižnice vyberáme aj *Engine*, to znamená, že výberom z ponuky určíme, aké typy súborov budú v danej knižnici SAS EG viditeľné. Základnou výhodou softvéru SAS je to, že vie priamo pracovať s dátovými súbormi z rôznych iných, najmä databázových programov. V ponuke voľby *Engine* sú typy súborov napr. ACCESS, Oracle, DB2, SPSS a iné. Štandardná ponuka pre *Engine* je *BASE – Latest version of Base SAS* (obr. 6.8).

Okno pre tvorbu SAS knižnice sa skladá z viacerých častí a jeho správne vyplnenie je na záver prezentované žltou ikonkou (pozri napr. obr. 6.9), ktorá pribudne v stromovej štruktúre projektu (*Project Tree*, resp. *Process Flow*). Symbolizuje knižnicu, ktorej obsa-