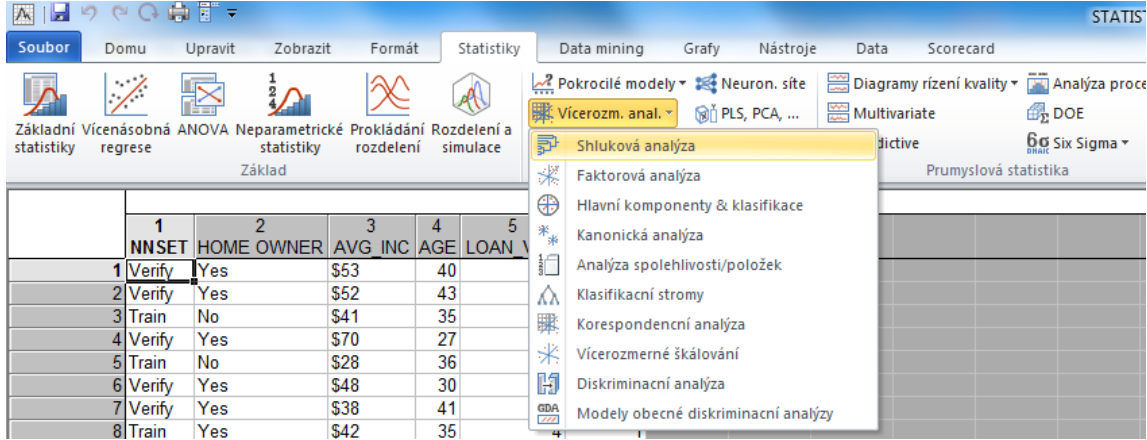


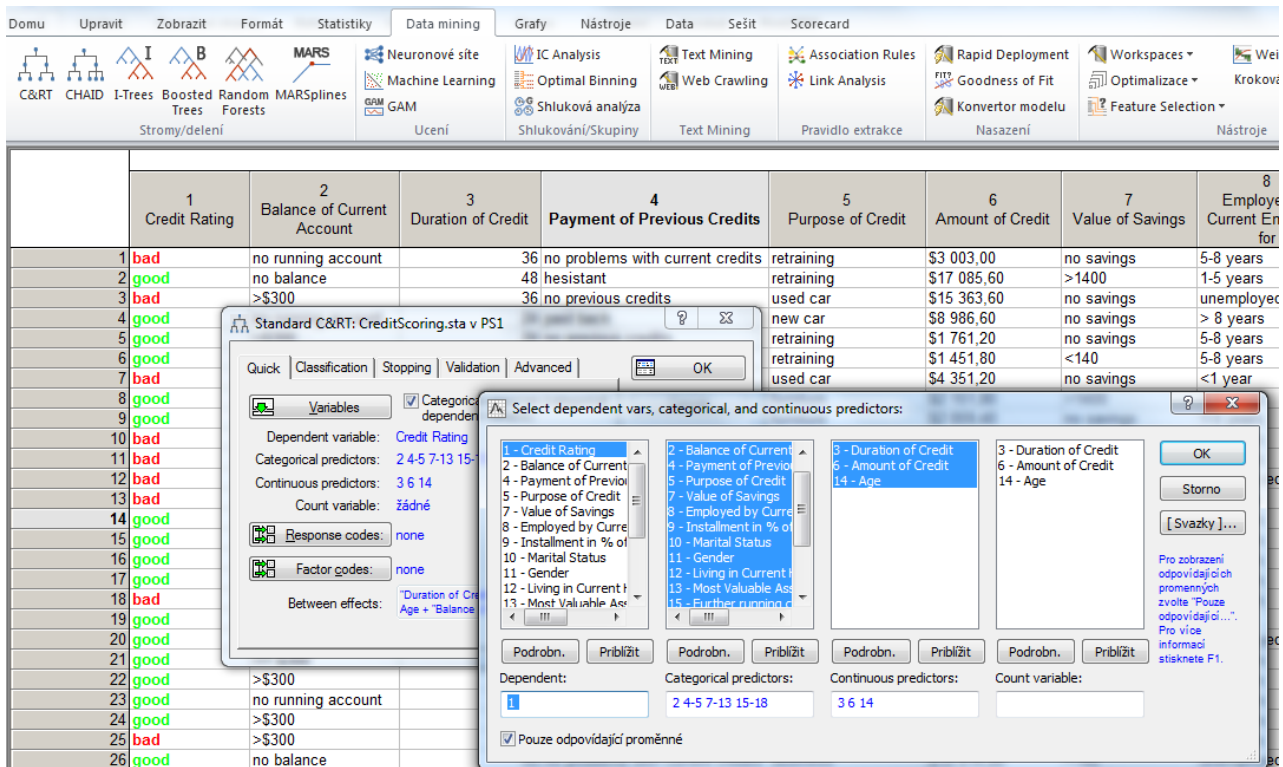


# Ovládání softwaru Statistica

Software Statistica je založen na přehledném "klikacím" rozhraní.

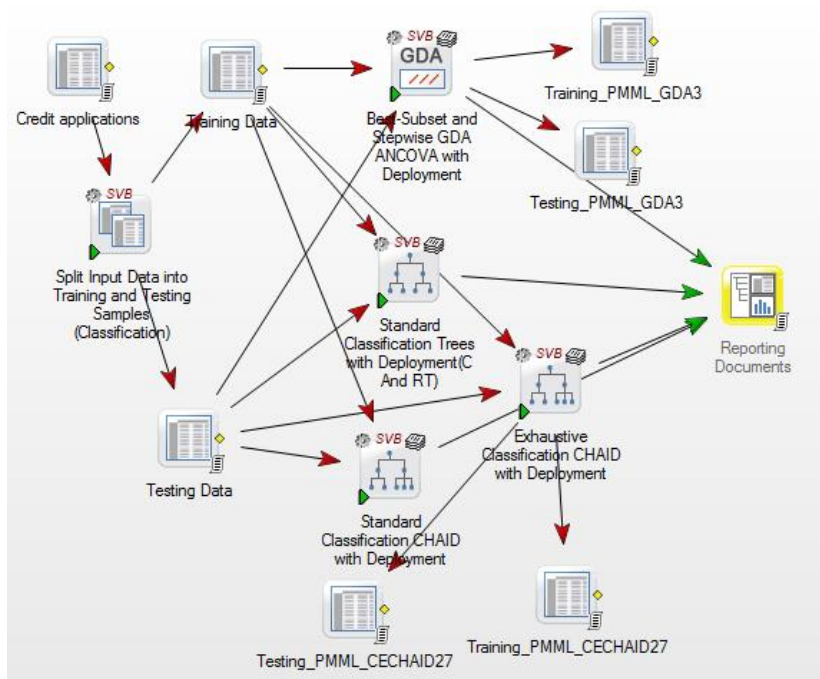


Pracovní prostředí je velmi podobné nástrojům MS Office. V dialozích jednotlivých analýz se snadno zorientujete.

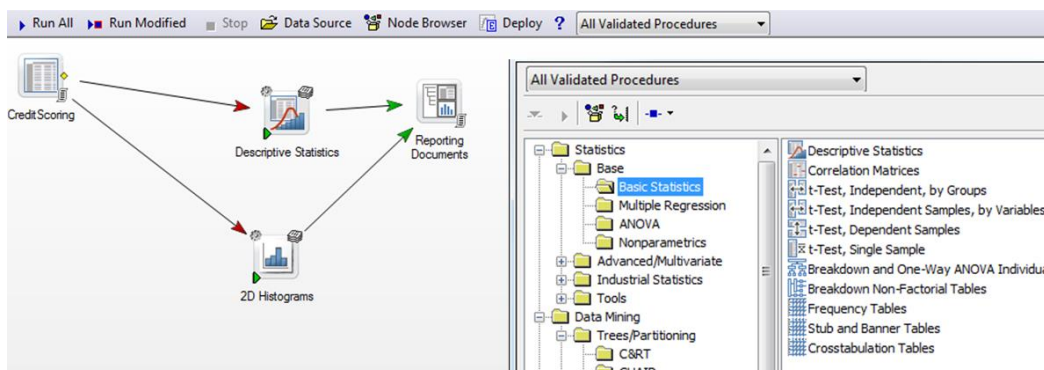


V programu Statistica jsou dva druhy rozhraní: *Klasické*, které je shodné se staršími verzemi softwaru Statistica a Pás karet (nová verze po vzoru pásu karet v produktech MS Office), tato zobrazení se přepínají přes menu *Zobrazit*.

Pro *dataminingové* uživatele (modul Data Miner) je v softwaru připravena možnost ovládání softwaru pomocí pracovní plochy (grafického interface - Workspaces):



V dialogu Node Browser si vyberete operaci (analýzu, transformaci, datový vstup), kterou chcete provést jako další krok a přetáhnete ji na plochu, šipkou spojíte s krokem předchozím. Dvojklikem na analýzu nastavíte konkrétní parametry výpočtu (proměnné apod.).



Celou sekvenci spustíte tlačítkem Run All. Samozřejmostí je možnost definovat si připojení do databáze společnosti, nastavit a uložit konkrétní dotazy. Příkaz Run All potom spustí celou sekvenci od načtení aktuální dat z DB, příptavy dat apod. Více o tomto tématu v kapitole [Ovládání modulu Data Miner](#) níže.

## Intuitivní ovládání softwaru Statistica

Pro provedení jakékoli analýzy musíme mít vždy nejdříve otevřena data. Data můžete do programu Statistica můžete načíst z různých datových zdrojů a databází, nativní datový formát Statistica má koncovku *sta*.

### Pokročilá práce s datovými zdroji

Software Statistica podporuje import dat prakticky ze všech zdrojů (např. text, SAS, Excel atd.), včetně databází (SQL, Oracle, Teradata, přes OLEDB [ODBC]). Více se dočtete zde:

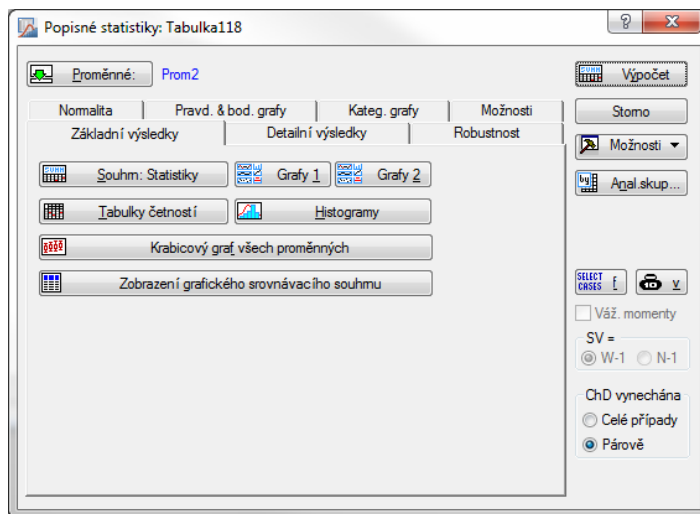
[http://www.statsoft.cz/file1/PDF/web/Pripojeni\\_do\\_DB.pdf](http://www.statsoft.cz/file1/PDF/web/Pripojeni_do_DB.pdf)

Je potřeba zmínit, že proměnné pro analýzy a grafy jsou vždy sloupcové – proměnné bychom tedy měli mít ve sloupcích. Řádkům se ve Statistice říká *případy*. Pro představu, máme pacienty a u každého měříme několik atributů. Pokud chceme dělat běžné analýzy bez dalších úprav, pak potřebujeme mít údaje o jednom pacientovi v jednom řádku.

### Filozofie dialogových oken

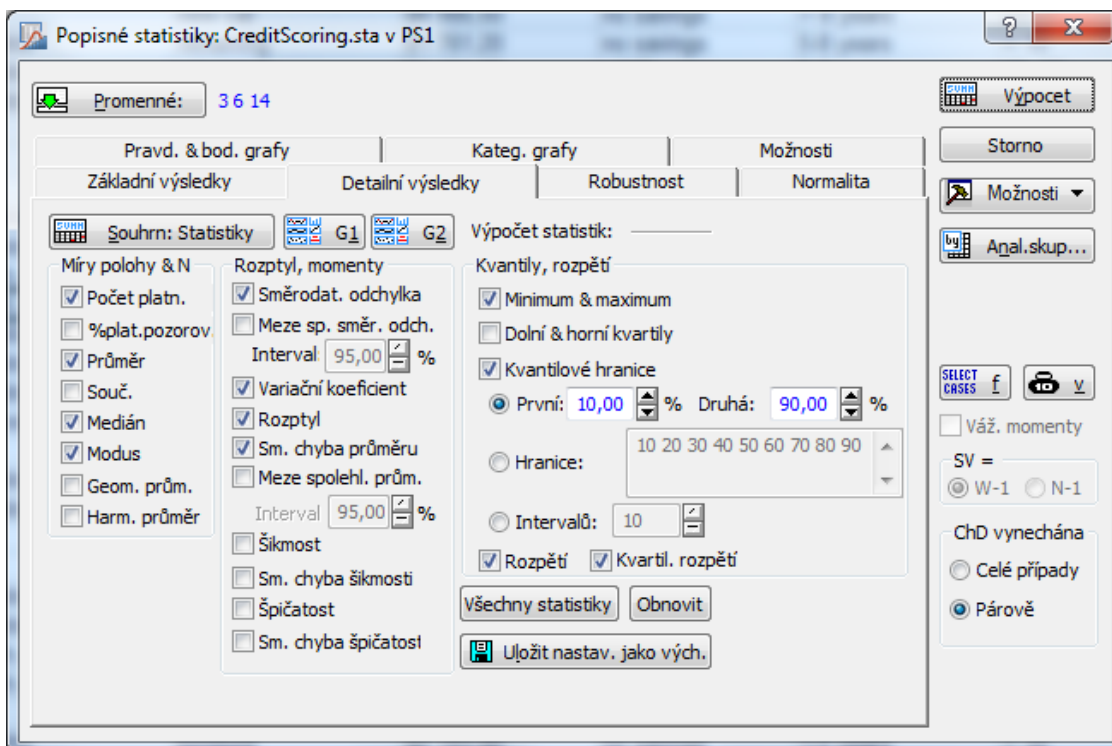
Když už máme datový soubor, můžeme na něm provést jakékoli analýzy či vykreslit grafy. Pojdme se podívat na logiku dialogových oken analýz.

Typicky na Vás vyskočí dialogové okno s mnoha tlačítky či záložkami. Pokusíme se tedy vysvětlit filozofii takového okna a vůbec práce s programem. V některých programech máte jediné tlačítko, a když na něj kliknete, vysype se na vás záplava výsledkových grafů a tabulek (je to sice jednoduché na naklikání, nicméně pak je potřeba se zorientovat v mnoha výstupech a protřídit si zpětně, co vlastně člověk potřebuje a co to vlastně znamená). Statistica se vydává úplně jinou cestou. Pokaždé si můžete přesně vybrat, který výstup potřebujete ve formě tlačítek či zaškrtnutí. Nemusí Vás tedy vůbec zajímat, že je zde nějaké tlačítko, kterému třeba nerozumíte nebo které nepotřebujete, prostě ho nepoužijete.



Také je zde mnoho tlačítek, které pomáhají v následujících analýzách, podívejme se například na dialog popisných statistik (*Statistiky*->*Popisné statistiky/tabulky*). Mějme situaci, kdy chceme spočítat popisné statistiky, tedy nějaké shrnující informace o proměnných a jejich rozdělení. Pokud nás už zajímají takovéto informace, velmi pravděpodobně nás následně budou zajímat například také histogramy nebo testy normality. Možnosti dalších analýz a grafických výstupů, které úzce souvisí se zkoumaným tématem, jsou tedy také přítomny a připraveny v dialogovém okně. Šetří tak práci uživateli (v našem příkladu například nemusíte histogram hledat v záložce grafy, nastavovat znovu

proměnné, vše je již připraveno v dialogu popisných statistik). Prostě je v dialogu vždy výčet metod, které by mohly být potřebné v souvislosti se zkoumaným tématem.



Pozn.: Samozřejmě takovéto možnosti jsou výhodné také z hlediska flexibility a použitelnosti softwaru, protože pokryjí požadavky většího množství uživatelů - každý potřebuje trochu něco jiného a takováto struktura dialogu to dokáže zajistit. Důsledkem tohoto je, že stejné výsledky dokážete získat více způsoby.

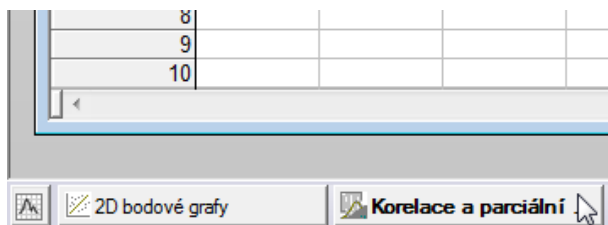
## Spuštění analýzy

Pro vytvoření výsledků či grafu je potřeba zvolit v dialogu jedinou věc a to, pro které proměnné se má analýza napočítat. Pokud byste proměnné nezvolili, program Vás sám vyzve, abyste tak učinili.

Tak schválně, zkusme si vytvořit nějaký výstup a uvidíme, jak je to jednoduché. Máme otevřen datový soubor a chceme vypočítat například základní popisné statistiky, jako tomu bylo již výše. Chceme analýzu, klikneme tedy na *Statistiky*, poté si uvědomíme, že popisné statistiky mohou být schovány jedině v *Základních statistikách*, zde už vybereme *Popisné statistiky* pomocí kliknutí na *OK*. Nic neřešíme a klikneme na *Výpočet*. Program nás vyzve k výběru proměnných, po vybrání proměnných na nás ihned vyskočí výsledek. Je to defaultně nastavený výstup (s parametry nastavenými tak, aby vyhovoval obecně nejpoužívanějším volbám), my jsme nemuseli dělat nic jiného než vybrat proměnné. Nicméně pokud nám z nějakého důvodu výsledek nevyhovuje, můžeme si analýzu maximalizovat z lišty analýz, kde se schovala (viz další odstavec) a změnit nastavení či vyvolat jiný výsledek.

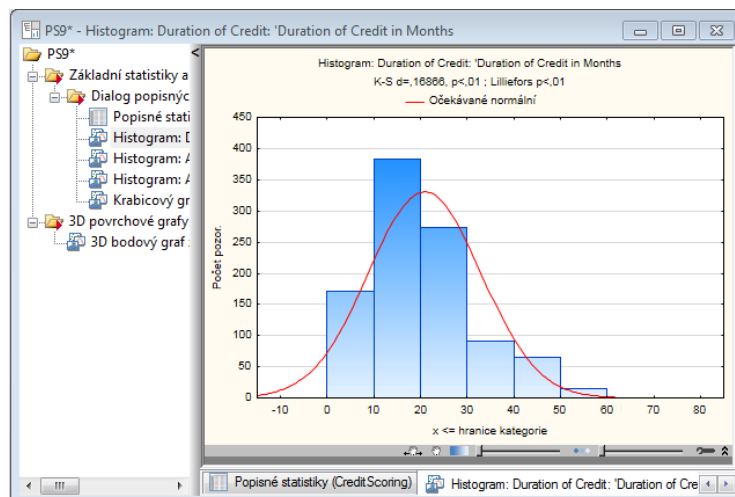
## Minimalizované analýzy

Pokud kliknete na něco, co jste nepotřebovali nebo chcete pokračovat v analýze, stačí kliknout na minimalizovanou analýzu dole v takzvaném *Panelu analýz*. Každá analýza (nebo graf) se po výpočtu minimalizuje na lištu. Nemusíte se tedy bát, že byste něco zkazili a museli následně všechno nastavovat znovu, minimalizovaná analýza zůstane k dispozici, dokud ji sami nezavřete nebo dokud nezavřete datový soubor, nad kterým analýza pracovala.



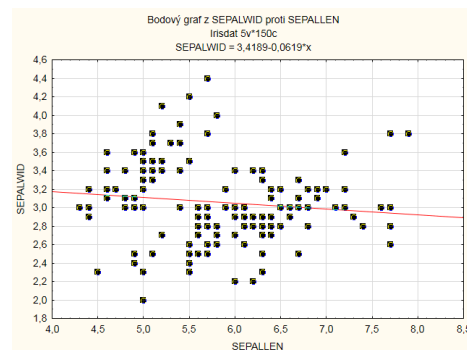
## Výstupy

Všechny výsledky se objevují (defaultně, samozřejmě můžeme nastavit i jiný typ výstupu: přes záložku *Nástroje->Možnosti*) v souboru typu *sešitu* Statistica (více o typech souborů [zde](#)). Tento formát slouží jako „kontejner“ pro všechny výsledky, můžete se v něm proklikávat, přesouvat, přejmenovávat a mazat soubory (stačí klávesou *Delete*). Také můžete soubory exportovat do protokolů či je osamostatňovat ze sešitu Statistica do samostatných oken.



## Grafy

Jakýkoli výsledný graf je plně interaktivní a dodatečně nastavitelný objekt. Chcete například změnit velikost bodů? Žádný problém, prostě intuitivně uděláte to, co Vás napadne v tomto kontextu jako první, tedy poklikáte na body v grafu. Otevře se Vám dialog *Možnosti grafu* právě na místě, které se týká bodů, na které jste klikli. Není tedy potřeba nic dalšího hledat. Nelíbí se Vám nadpis, kliknete na něj, zmáčknete klávesu *Delete* a nadpis je pryč. Takto jednoduše můžete dostat graf velmi rychle do podoby, která Vám vyhovuje.



Existuje také mnoho vychytávek, které Vám pomohou se na data podívat detailněji a lépe (více o možnostech grafů se dozvíte v [tomto článku](#) nebo z ilustračních videí [zde](#)).

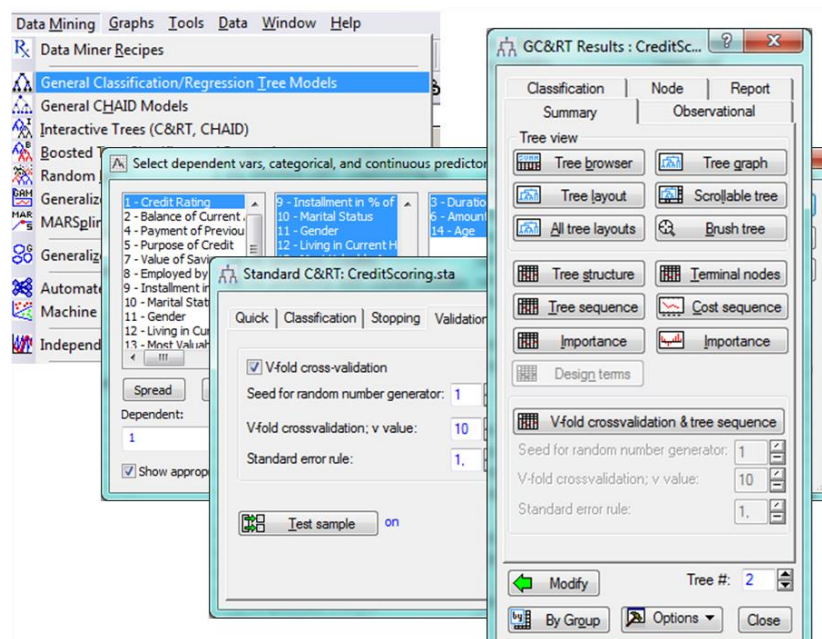


## Možnosti ovládání modulu Statistica Data Miner

Data miningový modul [Statistica Data Miner](#) obsahuje 3 možnosti ovládání:

- I. **Interaktivní nabídky** – pod tímto názvem si představme klasické menu, ve kterém si zvolím příslušnou metodu, kterou chci na konkrétní úlohu použít, definuji nezbytné parametry a dostanu výsledný panel této úlohy s výsledky analýzy, viz následující obr.:

*Jednotlivá okna dataminingové analýzy*



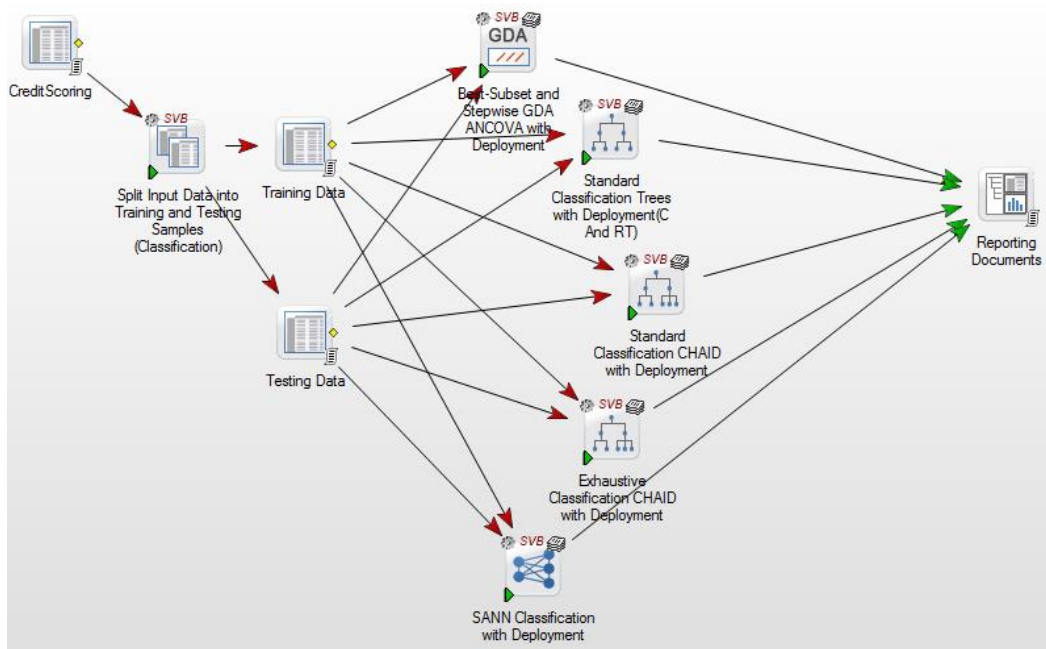
## II. Data Miner Workspaces

Druhou možností je *Data Miner Workspaces* – grafický interface, ve kterém si lze definovat vlastní konstrukci celé analýzy od datového zdroje až po nasazení modelu. Jednotkové ikony (analytické uzly) reprezentují různé činnosti a metody, které se během celého modelovacího cyklu uplatní. *Data Miner Workspaces* funguje na principu „drag&drop“ díky kterému lze jednoduše propojovat data, analýzy a výsledky jednoduše pomocí přetažení ikon a propojení pomocí šipek. V jeden okamžik lze otevřít libovolné množství pracovních prostředí a například jednoduše pomocí myši přetahovat jednotlivé uzly mezi těmito prostředími.

Možnosti nastavení uzlů analýz jsou ve většině případů totožné (vývoj softwaru směřuje do stavu, kdy budou mít všechny uzly shodné možnosti jako interaktivní rozhraní) s možnostmi v interaktivním prostředí, jak se můžete přesvědčit z následujícího obrázku. Tato shoda umožňuje velmi rychlou intuitivní adaptaci na workspace prostředí. Možnosti uzlu analýzy obsahují oproti interaktivní variantě navíc výběr toho, jaký výstup bude přístupný pro možné napojení dalších uzlů, také je zde možnost vložit poznámky (například o tom, proč byl zrovna tento uzel použit).

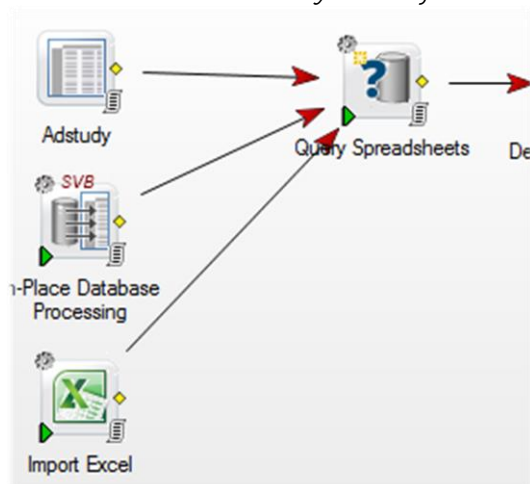
Statistica Data Miner také umožňuje snadnou přípravu vlastních uzlů a jejich kompletní integraci do systému. Výsledkem je systém odpovídající specifickým požadavkům uživatele. Součástí je také standardní vývojové prostředí, které umožňuje spravovat optimalizované analytické objekty (uzly) za použití skriptů v jazyku Statistica Visual Basic. Výsledky lze přehledně organizovat v protokolech, tabulkách, pracovních sešitech apod.

*Data Miner Workspaces (aplikace pěti různých modelů na jednom souboru)*



Data Miner Workspaces (načtení 3 souboru z různých zdrojů- soubor .sta, .xml a databáze a následné vytvoření dotazu nad těmito třemi datovými soubory):

*Načtení 3 souboru z různých zdrojů*



Náhled na data v konkrétním uzlu

	Total Talk Predictive	Total Agent Handle Time
1	0:00:00	2:28:33
2	0:13:35	2:30:31
3	0:00:00	0:42:56
4	0:53:52	3:32:39
5	0:00:00	1:31:58
6	1:27:27	3:04:04
7	0:00:00	0:30:22
8	0:00:00	0:00:00
9	1:14:58	3:15:52

Kliknutí na ikonu uzlu vpravo dole otevře datový soubor, který je výstupním souborem tohoto uzlu.

### III. Data Miner Recipes

Wizard, který Vás provede celou data miningovou analýzou. Tento výukový modul pro data mining je vhodný pro začínající specialisty. Krok po kroku provede uživatele celou tvorbou modelu až po jeho nasazení:

*Data Miner Recipes – ukázka jednoho kroku*

Model ID	Name	Training error (%)	Testing error (%)	Select for evaluation
2	Boosted trees	20.22	27.17	TRUE
3	Neural network	22.86	27.73	TRUE
1	C&RT	30.02	32.77	TRUE



## **Uložení rozdělené práce a výsledků analýz**

Software Statistica také obsahuje pokročilé možnosti uložení výsledků analýz či rozdělané práce:

- ✓ Pracovní sešity
- ✓ Projekt
- ✓ Samostatná okna
- ✓ Protokol
- ✓ Microsoft Word
- ✓ Výstup na web
- ✓ PDF

Více o tomto tématu se dočtete zde:

[http://www.statsoft.cz/file1/PDF/web/Priprava\\_dat\\_v\\_softwaru\\_Statistica.pdf](http://www.statsoft.cz/file1/PDF/web/Priprava_dat_v_softwaru_Statistica.pdf)