

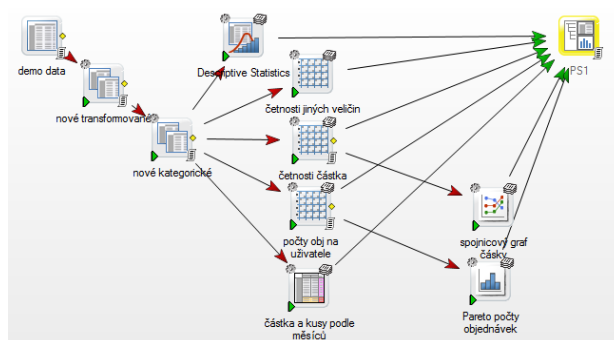


StatSoft

Prostředí workspace ve verzi 12

Od verze 12 se nám dostává do ruky, dá se říct, zcela nový vylepšený nástroj pro tvorbu nejen dataminingových analýz. Jednou z největších novinek verze 12 je upravené prostředí data miner workspace, které si teď popíšeme.

Workspace je jedno z prostředí, ve kterém lze provádět a spouštět analýzy. Jedná se vlastně o něco jako makro v grafické podobě. I člověk, který nezná programovací jazyk, je tedy schopen přenést svůj pracovní postup do formy, která je automatizovaná a je z ní vidět celý postup krok za krokem. Prostředí je založeno na budování pracovního postupu z jednotlivých bloků (uzlů), které představují jednotlivé funkcionality programu *STATISTICA*, a jejich propojení (šipek), která spojují uzly ve sledu, ve kterém mají být analýzy spouštěny. V tomto článku se tedy budeme pohybovat výhradně v prostředí uzlů a šipek.



Pracovní prostředí workspace otevřeme v programu buď přes *Soubor-Nový* nebo před *Data Mining-Workspaces*.

Porovnání rozdílů mezi „starým“ a „novým“ workspace

V této části bychom chtěli ukázat rozdíl mezi prostředím workspace ve verzi 10 a workspace ve verzi 12. Nejdříve bychom ale rádi odkázali na již existující studijní materiál. Workspace prostředí verze 10 bylo ukázáno na [příkladu zajímavé regresní úlohy](#), dále najdete podrobnější popis prostředí v [našem interním manuálu](#). Funkční příklad workspace ve verzi 12 najdete [zde](#).

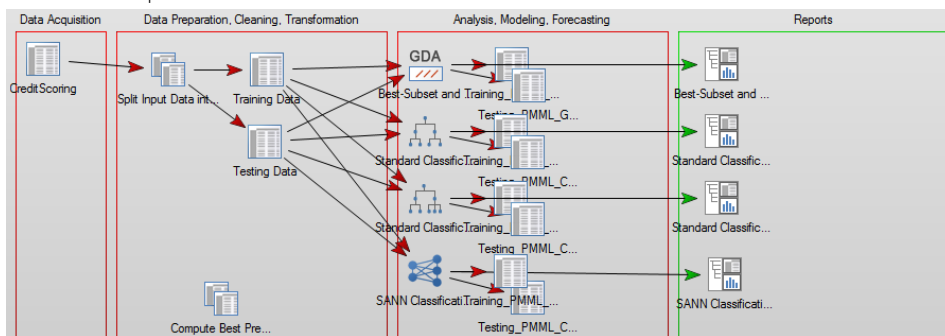
Ti, kteří začínají workspace prostředí používat až ve verzi 12, se prozatím verzi 10 zabývat nemusejí, nicméně s postupem času mohou vznikat případy, kdy se budou hodit i staré uzly a staré typy napojení a pak je dobré vědět, že něco takového je možné. Všechny funkcionality a již vytvořená pracovní prostředí, které fungovaly ve starších verzích, nadále fungují a všechny uzly starého workspace můžete používat i nyní.

Popíšeme si tedy hlavní rozdíly mezi starým a novým workspace:

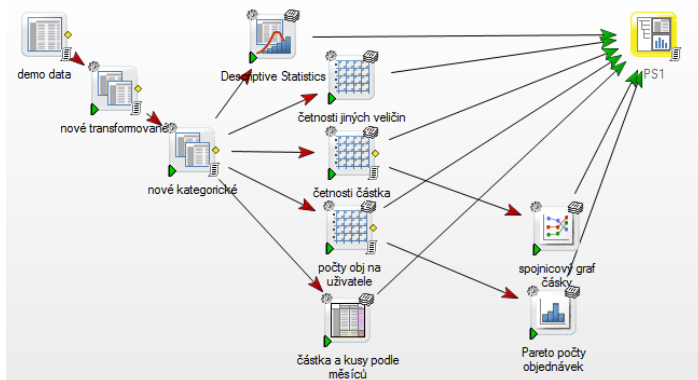
› Neexistující dělení workspace na sekce

Na první pohled na nové prostředí je vidět rozdíl ve vzhledu. U nových workspace už nenajdete sekce, které by oddělovaly jednotlivé fáze zpracování dat. Zmizely tedy omezení, která z tohoto dělení plynula, uzly lze nyní umísťovat a napojovat zcela libovolně. Výsledný výsledkový sešit je pouze jediný, u každé analýzy je pak ikona, která vede přímo k výsledkům z daného uzlu.

Příklad workspace ve verzi 10:



Jiný příklad workspace ve verzi 12:

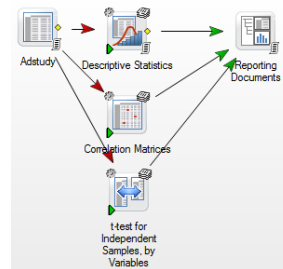
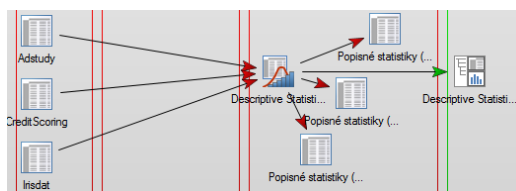


› **Podpora Pásu karet a zjednodušení ovládání prostředí**

Byly implementovány nové možnosti, jak zjednodušit práci ve workspace, toto bude detailněji popsáno níže v samostatné kapitole. Nyní je tedy používání workspace mnohem jednodušší a přívětivější.

› **Staré versus nové uzly**

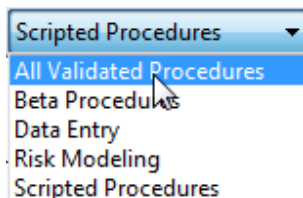
Ve verzi 12 jsme se dočkali nového typu uzlu, který bude popsán níže v samostatné kapitole. S postupem času by měly mít všechny staré uzly svou novou alternativu. Rozdíl mezi starými a novými uzly je zásadní. Je zde rozdíl nejen ve vzhledu (což bude popsáno níže) a v rozšířených možnostech použití (také bude níže), ale hlavně ve filozofii použití uzlu. Ve starých uzlech analýz nebylo možné vybrat proměnné (ty se nastavovaly přímo v uzlu s datovým souborem), byl tedy problém s napojením více analýz na jeden datový soubor (což se řešilo uzlem pro duplikování datového souboru), dále bylo možné napojit na jednu analýzu i více datových zdrojů zároveň, výstupem pak bylo také více výstupních souborů z jednoho uzlu. Naproti tomu, do nového uzlu lze najednou napojit jen jeden datový soubor a proměnné mohou být voleny přímo v uzlu analýzy, výstupním souborem je jen jeden výstupní datový soubor, důsledkem je, že už není potřeba duplikace datových souborů. Každá z těchto dvou konstrukcí může být případ od případu výhodnější než ta opačná. Hlavní rozdíly a zároveň výhody jsou zobrazeny na příkladech workspace výše na obrázcích. Co se týče možnosti využití, tak ve starých uzlech analýz byla pouze možnost provést předprogramovanou analýzu s nastavením několika parametrů a jejich zdrojový kód se dal měnit. Nové uzly mají funkcionality stejné jako kdybychom používali danou metodu v interaktivním módu (tedy klasicky klikáním přes menička), ale změny v nich již nejsou možné. Je ale dobře, že uživatelé mohou v nové verzi zároveň používat jak staré uzly s jejich stylem napojování, tak uzly nové. Výskyt kombinace obou typů postupů v jednom programu je unikátní a dává uživateli opravdu široké možnosti použití.



Pomůcky pro snadnější ovládání prostředí workspace

› Tematická menu

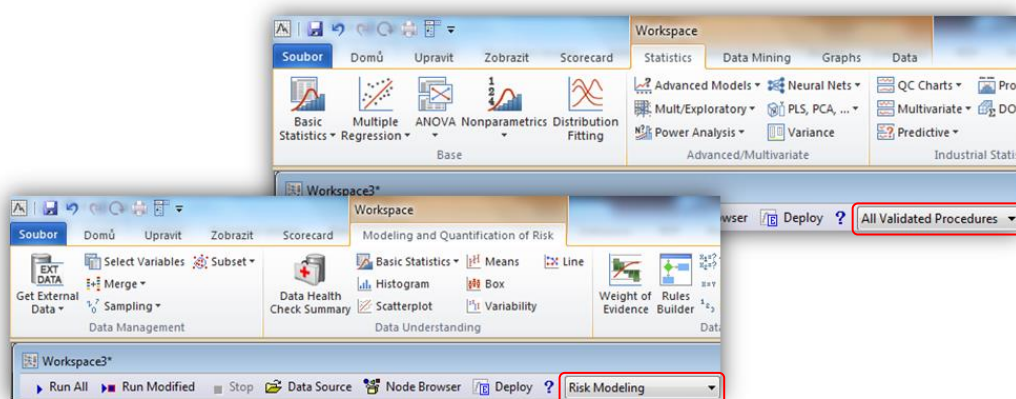
Po otevření workspace si můžeme všimnout nového menu, kde je jakýsi tématický výběr pro uzly, které můžeme vybírat. Tato volba následně ovlivňuje výběr uzlů, které máme k dispozici v meníčkách - jednak v tzv. Node Browseru, stejně jako v nabídkách Pásu karet (bude detailněji níže), nicméně neovlivňuje ty uzly, které máme již vybrány. Tyto tématické okruhy budou jistě s vývojem programu přibývat. Nyní zde máme následující možnosti:



All Validated Procedures – všechny otestované a validované procedury
Beta Procedures – uzly, které jsou teprve ve vývoji, nejsou perfektně vyladěny na všechny možnosti a mohou se u nich vyskytnout chyby nebo nekompatibilita
Scripted Procedures – všechny „staré“ uzly, tak jak je znáte z verze 10
Data Entry – speciální podvýběr uzlů pro oblast vkládání a záznamu dat
Risk Modeling – podvýběr uzlů hodících se pro oblast modelování riziku

› Možnost vybírat uzly přímo z prostředí Pásu karet

Nabídky v typu zobrazení Pásu karet jsou uzpůsobeny přímo podle vybraného tématického menu (viz předchozí kapitola). Propojení s workspace poznáte podle oranžového podbarvení záložek, podbarvené záložky jsou přímo propojeny s otevřeným workspace, pokud kliknete na nějakou metodu, přímo se ve workspace objeví daný uzel. Toto je tedy alternativní, rychlá a velmi příjemná možnost výběru uzlů. Níže vidíte, že nabídky se přizpůsobují podle toho, který okruh uzlů máte vybrán:



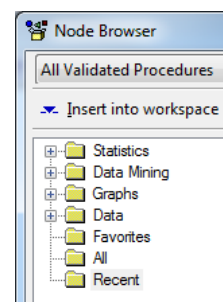
› Složky Recent a Favourites v Node Browser

Potřebný uzel lze klasicky vybírat také přes tlačítko *Node browser*, kde si můžete potřebný uzel najít. Zajímavé jsou zde tři složky:

All – kde je výčet všech přístupných uzlů (to je podmíněno výběrem tématického okruhu v rozevíracím menu)

Recent – zde najdete naposledy použité uzly. Toto je velmi přínosná varianta, protože většinou používáte jen omezenou skupinu uzlů, a zde najdete velmi pravděpodobně právě ty uzly, které potřebujete

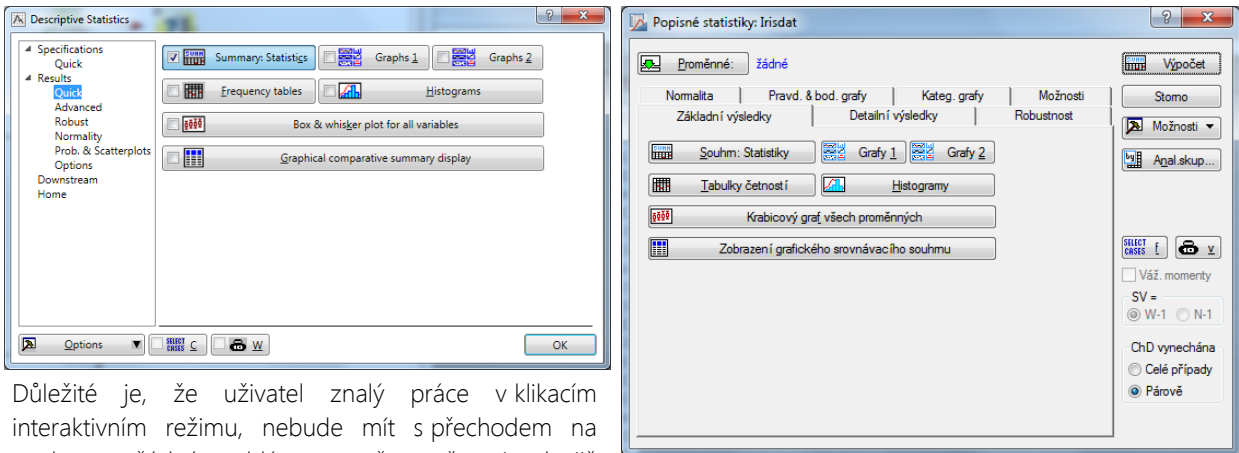
Favourites – zde si můžete vložit svou oblíbenou sadu uzlů – přetažením z jiných složek Node Browseru



Nový typ uzlů

› Kompatibilita s interaktivním rozhraním

Po rozkliknutí uzlu (například klasických popisných statistik) jistě ihned poznáte, v čem je největší výhoda nových uzlů, v jejich menu jsou shodné možnosti, které najdete v interaktivním režimu při klasickém klikání na analýzy bez použití workspace (viz například popisné statistiky):



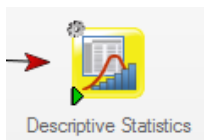
Důležité je, že uživatel znalý práce v klikacím interaktivním režimu, nebude mít s přechodem na workspace žádný problém, protože možnosti uzlu již zná a umí používat.

Poznámka: nové uzly nejsou zatím přítomny pro všechny analýzy, nicméně na rozšiřování počtu uzlů nového typu se neustále pracuje.

› Výběr proměnných přímo v uzlu

Nově jde přímo v uzlech vybírat proměnné, na které budou aplikovány analýzy. Vše potřebné pro výstup se tedy nastaví přímo v uzlu.

› Ikona uzlu



Žluté podbarvení nového typu uzlu značí, že uzel ještě nebyl použit nebo u něj došlo ke změně nastavení

V každém rohu úspěšně spuštěného uzlu mohou být následující možnosti:



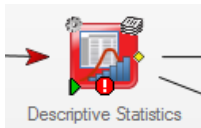
Levý horní roh (ozubené kolečko) – zde je možné nastavit (po kliknutí levým tlačítkem na ozubené kolečko) parametry výpočtu (nastavení je vyvoláno i dvojklikem doprostřed ikony uzlu)

Levý dolní roh (zelená šipka) – spustí daný uzel a případně i předcházející neprovedené uzly potřebné ke spuštění daného uzlu

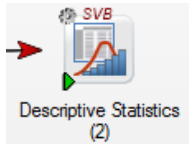
Pravý horní roh (stol listů) – otevře výsledkový sešit v místě, kde jsou zobrazeny výsledky, které vznikly spuštěním tohoto uzlu

Pravý dolní roh (jeden list) – jedná se o výstupní datovou tabulku, která může být použita pro další analýzy

Střed pravé strany ikony (žlutý kosočtverec) – kliknutím levým tlačítkem a přetáhnutím do jiného uzlu vytvoříte šipku, spojnicí mezi uzly



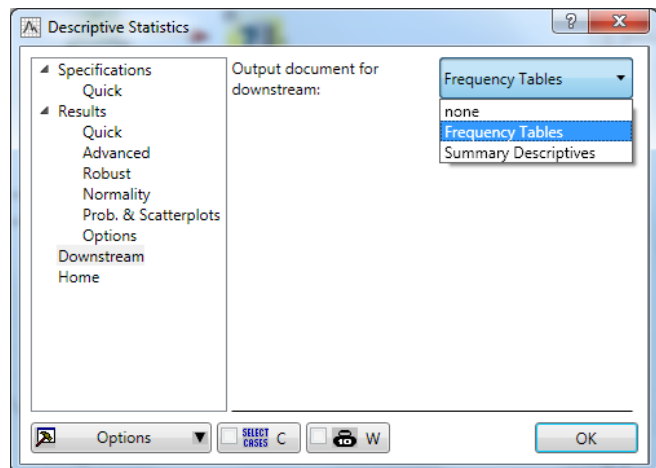
Červeně podbarvený uzel – u uzlu došlo k nějaké chybě, pro znovuvyvolání chybové hlášky klikněte na červený výkřičník uprostřed spodní hrany ikony



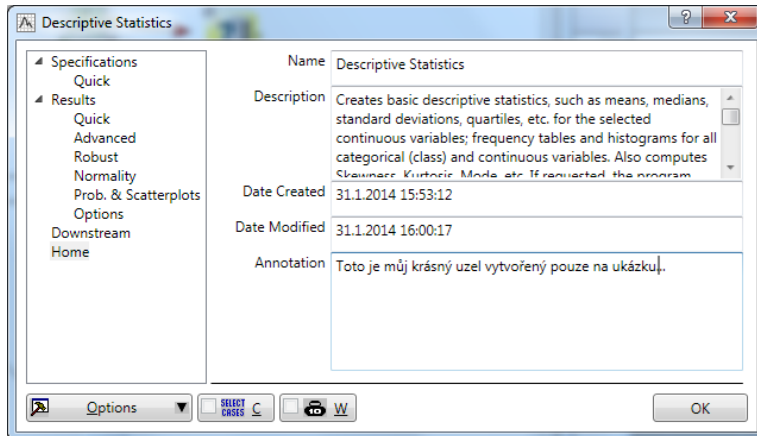
Nápis SVB upozorňuje, že se jedná o starý skriptovaný uzel

› Sekce uzlu s názvem *Downstream*

Každý uzel nového typu má kromě klasických voleb přístupných z interaktivního módu i dvě záložky navíc. Nyní si popíšeme první z nich a tou je velmi důležitá záložka *Downstream*. Zde si může uživatel zvolit, který dokument se použije jako potenciální vstup do další analýzy (tedy, co bude použito jako soubor v pravém dolním rohu ikony uzlu). Někdy pomocí uzlu získáváme výstupní tabulky, které nepotřebujeme použít, jindy můžeme z jednoho uzlu získat více tabulkových výstupů, sekce *Downstream* je určena pro výběr jednoho výstupu použitelného dále.



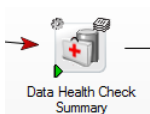
› Sekce uzlu s názvem *Home*:



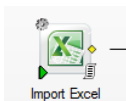
Druhou novou záložkou je sekce *Home*, ve které je defaultní popis uzlu, čas, kdy byl uzel vytvořen a místo pro vlastní komentář. Můžete si zde tedy například vypsat, proč jste daný uzel použili a co přesně dělá. Takováto informace může být velmi užitečná, pokud bude Váš projekt spouštět také někdo jiný, případně může sloužit popis jako pojistka, kdybychom si v budoucnu nemohli vzpomenout, proč jste zrovna tento daný uzel zařadili.

Výběr důležitých uzlů s novými funkcionalitami

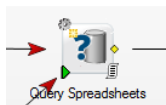
Bude následovat výčet nejzajímavějších nových uzlů a také speciálních uzlů vzniklých jinou potřebou workspace oproti klikacímu prostředí.



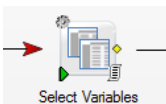
Data Health Check Summary – Velmi užitečný uzel pro prvotní analýzu dat. Uzel vytvoří souhrnný report o kvalitě datového souboru – tj., řekne, které proměnné jsou řídké, které kategorie mají nízký výskyt, která pozorování jsou odlehlá, kde jsou chybějící hodnoty, které proměnné jsou invariantní, zda jsou zde proměnné, které spolu hodně korelují, atd.



Import Excel - Uzel pro načtení Excelovského datového souboru. Je v něm potřeba nastavit cestu k souboru a oblast dat, stejně jako při standardním načítání Excelu. Lze nastavit i typ proměnných.



Query spreadsheets - Jedná se jednoduše o Query (tedy funkcionalitu pro tvorbu dotazů nad databází), nicméně databáze se v tomto případě skládá z datových souborů napojených (vstupujících) do *Query Spreadsheets*. Vhodné například pro spojování souborů z různých datových zdrojů.



Select variables - Staré typy uzlů potřebují mít před spuštěním předem vybrány proměnné, které do analýzy budou vstupovat. Pokud je datový soubor vstupující do analýzy výstupem z uzlu nového typu, nelze v něm vybrat proměnné, uzel *Select Variables* je tedy propojovacím článkem mezi novým a starým typem uzlu.

Další materiály

Nakonec bychom odkázali na [článek](#), ve kterém jsou uvedeny konkrétní jednoduché příklady workspace, které mohou jistě velmi pomoci při práci v tomto prostředí.