



StatSoft

Vzorové příklady použití prostředí Workspaces

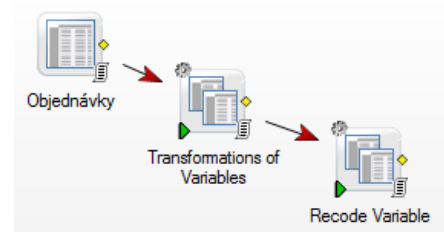
Pokud jste právě začali prostředí workspace používat nebo pokud si chcete udělat lepší představu, o čem tvorba úloh v tomto prostředí je, pak toto je článek přesně pro Vás.

V tomto článku najdete několik praktických ukázek jednoduchých struktur, které jsou často používané nebo ukazují hlavní principy. Všechny příklady je možné stáhnout a vyzkoušet si, což je také doporučený způsob práce s tímto článkem, jakmile máte možnost kliknout si na jednotlivé nastavení a výstupy uzlů, vše je najednou velmi rychle pochopitelné. Než začneme, bylo by vhodné poznamenat, že tento článek je vhodným doplněním informací, které byly sepsány [zde](#) (zde také najdete odkazy na další literaturu ohledně workspace).

Příklad 1 – Transformace a překódování

Běžně potřebujeme na vstupním souboru nejprve provést před analýzou nějaké transformace (vytvořit nové pomocné proměnné), nové proměnné tvoříme buď vzorcem nebo překódováním podle podmínek.

Poznámka: V uvedeném [příkladu](#) chceme přidat nové proměnné vycházející z datumu, který v souboru máme, tedy zde například vypočítáme, o který den v týdnu nebo měsíc v roce se jedná, atd. Druhý uzel následně kategorizuje nově vytvořenou proměnnou *částka*. Připomeňme, že mezivýsledky lze zkontrolovat po kliknutí na pravý dolní roh uzlu. Použité funkcionality programu jsou zde výrazy pro dávkovou transformaci (transformace) a překódovat. Jednotlivé části uzlu popisuje již v úvodu odkazovaný článek na str. 4.



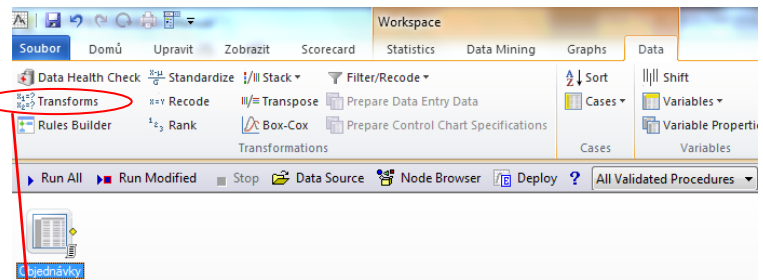
Postup: V tomto prvním příkladu ukážeme postup tvorby tohoto workspace detailně od začátku až do konce, v ostatních příkladech již podrobný popis nebude, nicméně každé nastavení je možné vyvolat v příkladovém workspace kliknutím na nastavení, případně výstupy jednotlivých uzlů.

1. Začínáme s prázdným workspace, který vytvoříme pomocí *Data Mining-Workspaces-All Procedures*.
2. Datový zdroj [Objednávky.sta](#) načteme buď hned po vytvoření workspace nebo až po otevření pomocí tlačítka *Data Source*.

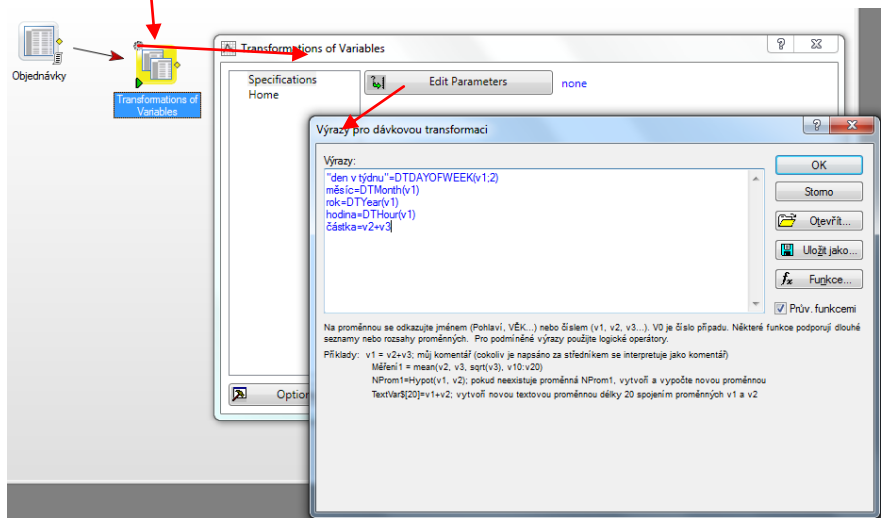


3. Nyní máme na pracovní ploše jeden uzel s datovým souborem, pokud klikneme na pravý dolní roh, objeví se nám datový soubor. Jedná se o soubor s objednávkami v internetovém obchodě, jsou zde časy vytvoření objednávky a poté částky, které byly anebo nebyly z dané objednávky zaplacený.

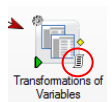
4. Nyní chceme přidat uzel pro vytvoření transformovaných proměnných. Klikneme tedy v záložce *Data* (v rozhraní *Pás karet*) na *Transforms*, což napojí na náš uzel uzel transformací (pokud by se nenapojil na správné místo, vytvoříme mezi oběma uzly šipku potáhnutím levým tlačítkem ze žlutého kosočtverce uprostřed pravé hrany uzlu s daty do uzlu transformací).



5. Otevřeme uzel transformací kliknutím na ozubené kolečko v ikoně transformací. Objeví se nám dialog uzlu, vybereme *Edit Parameters* a vložíme požadované transformace (chceme získat několik proměnných z proměnné *datum nákupu*, tvoříme nové proměnné *den v týdnu*, *měsíc*, *rok* a *hodina*, a poté proměnnou *částka*, což je součet obou částí objednávky):

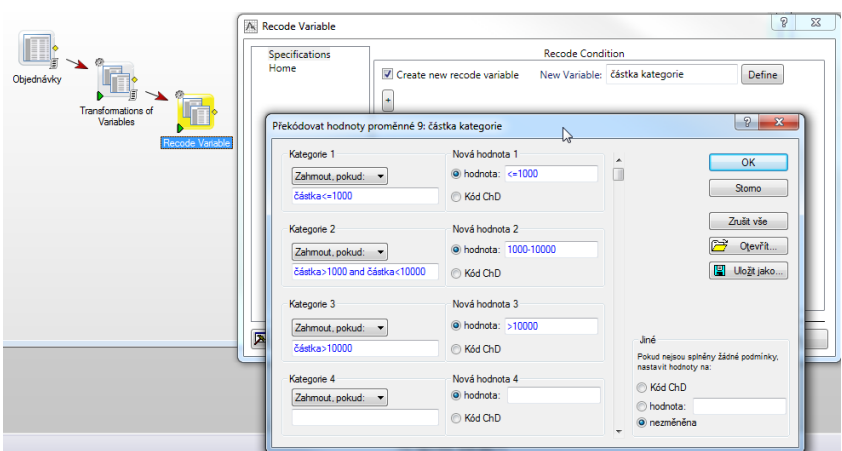


6. Dáme *OK-Ano-OK* a spustíme uzel transformací zelenou šipkou v ikoně uzlu vlevo dole. Tím se vytvoří výstupní soubor, který lze skontrolovat kliknutím na pravý dolní roh, ikona listu papíru, v němž již máme nové proměnné.



7. Přidáme a napojíme uzel překódování stejným způsobem jako předtím uzel transformací (uzel najdeme v *Data-Recode*).

8. V uzlu nastavíme překódování, v tomto případě jsou to 3 kategorie (s částkou pod 1000, mezi 1001-10000 a více než 10000). V tomto uzlu lze překódovat stávající proměnné, stejně tak vytvořit nové, proměnných lze překódovat více najednou.



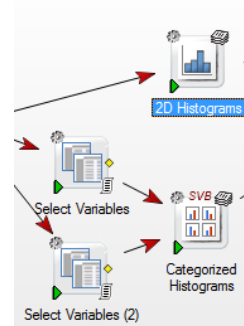
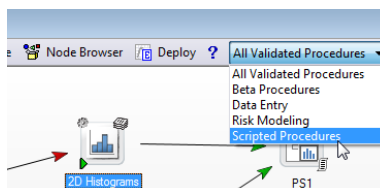
9. Spustíme uzel *Recode Variable* zelenou šipkou, vzniklý výstupní soubor (dostupný v pravém dolním rohu ikony

má nyní všechny potřebné proměnné a může být použit pro další analýzy).

Příklad 2 – Grafy a použití obou typů uzlů

Grafické výstupy jsou jistě velmi důležitou funkcionalitou, kterou jistě budete chtít používat. Tento **příklad** navazuje na příklad 1 vytvořením grafů na transformovaných datech z prvního příkladu.

Poznámka: Jak jste si jistě všimli, jsou zde použity oba typy uzlů, tedy jak nový, tak starý skriptovaný (SVB, či také starý typ uzlu). Rozdíl mezi starým a novým uzlem ukazuje již zmiňovaný článek v úvodu. Staré typy uzlů můžete do workspace zařazovat, pokud máte vybrán typ uzlů: *Scripted Procedures*. Pás karet sice vypadá stejně, ale jsou vybírány pouze skriptované uzly. Výsledné grafy z jednotlivých uzlů najdete po kliknutí na pravý horní roh uzlu vytvářejícím graf.

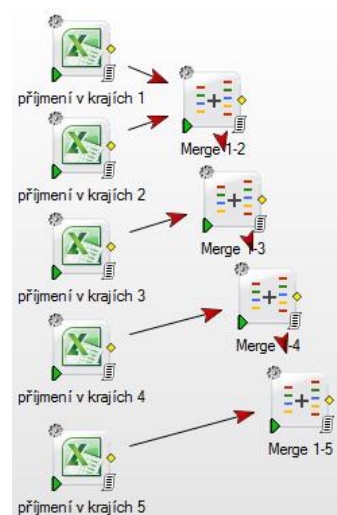


Příklad 3 – Spojování souborů stejné struktury

Typickým příkladem, kdy potřebujeme spojit soubory stejné struktury, je spojení jednotlivých listů Excelu do jednoho souboru. Pomůže Vám **následující workspace** (tento příklad byl použit v **tomto** článku).

Poznámka: Workspace načítá jednotlivé listy Excel souboru a pomocí uzlu *Merge variables* je po dvojicích spojuje (uzly typu *Merge* očekávají jako vstup dva soubory, existují i jiné varianty spojování, uzlů *Merge* je hned několik druhů, podle toho, jak chcete soubory spojovat. Zde jsme použili spojování podle proměnných).

Pro spuštění workspace na Vašem počítači je potřeba nastavit cesty k **souboru** v každém z Excel uzlů.

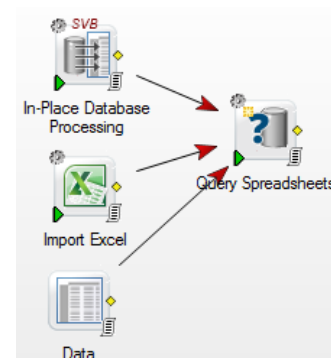
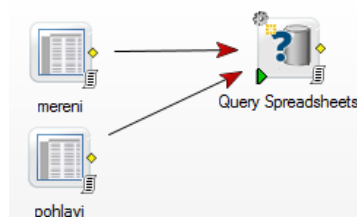


Příklad 4 – Spojování dat z různých zdrojů

Pokud potřebujeme spojovat tabulky s různými informacemi z různých zdrojů (například v jednom souboru máme jednu informaci o jednotce a v druhém jinou), pak použijeme uzel *Query Spreadsheets*, který funguje stejně, jako bychom se dotazovali do databáze.

Poznámka: na obrázku je napojení jak tabulky z databáze, tak Excelu, tak klasické tabulky STATISTICA.

Protože ne každý se může napojit na databázi, ve **workspace ke stažení** jsou jen dva .sta soubory. Úkolem uzlu *Query Spreadsheets*

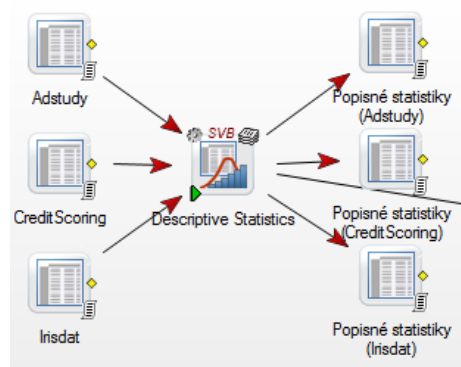


v tomto příkladu je přidat k mírám z prvního souboru informace o pohlaví ze souboru druhého. Soubory jsou spojeny podle identifikátoru, který je v obou souborech.

Příklad 5 – stejná analýza na více souborech

Pokud chceme využít stejný typ analýzy zároveň na více souborech, je potřeba použít skriptovaný (SVB, či také starý typ uzlu):

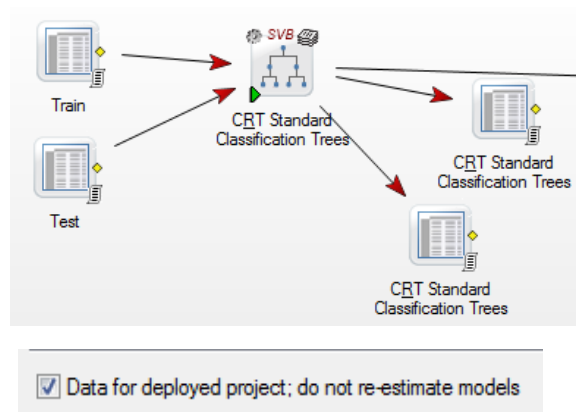
Poznámky: protože používáme starý typ uzlu, musíme mít předem zvoleny proměnné, které chceme v analýze zpracovávat. A to se provede buď přímo v uzlu datového souboru (po dvojkliku na něj). Případně lze použít uzel **Select variables**, který se napojuje mezi uzel souboru a uzel analýzy. V našem příkladu mají být proměnné pro uzel **Descriptive Statistics** spojitě, při volbě proměnných tedy volíme proměnné v sekci spojitě závislé (**Dependent, Continuous**). Nastavení si můžete skontrolovat přímo ve **workspace** s příkladem na obrázku. Příklad počítá stejné popisné statistiky pro všechny soubory.



Příklad 6 – Tvorba dataminingového modelu

Tímto příkladem přecházíme k dataminingovým modelům. Je zde ukázán uzel pro vybudování modelu (jedním z výstupů ve výsledkovém sešitě je i PMML skript, který může následně sloužit pro skórování na nových datech). Je zde také ukázáno, jak zároveň při tvorbě modelu udělat přímo i predikci na testovacích datech, které model nepoužil.

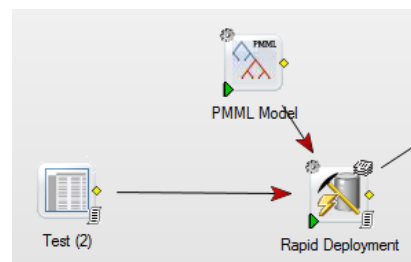
Poznámka: jelikož je model ve formě starého uzlu, je potřeba nastavit proměnné již přímo v uzlu s datovým souborem. Aby mohla být pro testovací data provedena predikce, je potřeba při volbě proměnných v uzlu s testovacími daty zaškrtnout volbu: *Data for deployed project...*



Příklad 7 – Nasazení PMML modelu

Máme vytvořen model a potřebujeme předpovědět hodnoty na novém souboru na základě modelu. Tento model již nevytváříme, pouze ho aplikujeme na nová data.

Poznámka: aby byl **ukázkový workspace** funkční, je potřeba nahrát **PMML soubor** do uzlu **PMML Model**. Na uzel **Rapid Deployment** se napojují jednak modely ve formě PMML, tak soubor, který chceme oskórovat (na kterém chceme provést predikci).



Předdefinovaná schémata

V programu existují také předdefinované často používané analýzy a postupy, které můžete jednoduše využít pro svůj data miningový projekt. Tyto předdefinované workspace najdete v menu *Data Mining-Workspaces*.

