

# Slučování tabulek

Newsletter Statistica ACADEMY



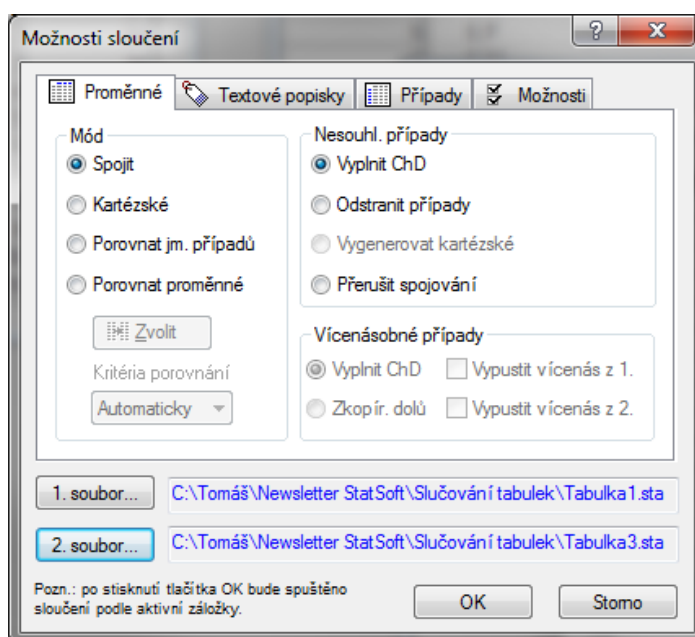
Téma: Příprava dat  
Typ článku: Návody

*Máte informace ve více tabulkách a chcete je sloučit dohromady? Pak je tento článek právě pro Vás. Vysvětlíme, jaké jsou možnosti softwaru Statistica v souvislosti se slučováním tabulek.*

Budeme se snažit popsat konkrétní příklady sloučení, aby bylo vidět, co software při kterých volbách a nastaveních dělá. Rádi bychom tímto dokumentem vytvořili jakéhosi poradce, který Vám pomůže. Bude po ruce pokaždé, když budete potřebovat sloučit soubory bez toho, abyste si museli pamatovat, které nastavení kdy použít.

## Sloučení dvou tabulek

První příklady se budou týkat slučování dvou tabulek. Poněvadž se jistě jedná o operaci nad daty, je tedy potřeba hledat tyto funkcionality v záložce **Data**, konkrétně volba **Sloučit (Merge)**.



## 1. Spojit (ostatní nastavení Defaultně)

Vyberete dva soubory, které chcete slučovat, vyberete mód **Spojit**. Software jednoduše slepí soubory dohromady tak, jak jsou připraveny (viz obrázek),

|    | 1 ID | 2 Gender | 3 Measurement |
|----|------|----------|---------------|
| 1  | 1    | M        | 171           |
| 2  | 2    | F        | 172           |
| 3  | 3    | F        | 163           |
| 4  | 4    | M        | 182           |
| 5  | 5    | F        | 174           |
| 6  | 6    | M        | 189           |
| 7  | 7    | M        | 192           |
| 8  | 8    | M        | 174           |
| 9  | 9    | F        | 175           |
| 10 | 10   | M        | 178           |

|    | 1 ID | 2 Numbers |
|----|------|-----------|
| 1  | 1    | 12        |
| 2  | 1    | 14        |
| 3  | 2    | 57        |
| 4  | 2    | 54        |
| 5  | 3    | 14        |
| 6  | 4    | 32        |
| 7  | 5    | 65        |
| 8  | 6    | 88        |
| 9  | 7    | 74        |
| 10 | 8    | 15        |
| 11 | 9    | 16        |
| 12 | 10   | 59        |
| 13 | 10   | 96        |

|    | 1 ID | 2 Gender | 3 Measurement | 4 ID | 5 Numbers |
|----|------|----------|---------------|------|-----------|
| 1  | 1    | M        | 171           | 1    | 12        |
| 2  | 2    | F        | 172           | 1    | 14        |
| 3  | 3    | F        | 163           | 2    | 57        |
| 4  | 4    | M        | 182           | 2    | 54        |
| 5  | 5    | F        | 174           | 3    | 14        |
| 6  | 6    | M        | 189           | 4    | 32        |
| 7  | 7    | M        | 192           | 5    | 65        |
| 8  | 8    | M        | 174           | 6    | 88        |
| 9  | 9    | F        | 175           | 7    | 74        |
| 10 | 10   | M        | 178           | 8    | 15        |
| 11 |      |          |               | 9    | 16        |
| 12 |      |          |               | 10   | 59        |
| 13 |      |          |               | 10   | 96        |

**Poznámka:** Můžete si určit, které případy i proměnné pro každý soubor chcete – volby jsou v záložce **Možnosti**.

**Možnosti sloučení**

Proměnné | Textové popisky | Případy |  Možnosti

**Vstup**

Použít filtr pro 1. soubor     Data setříděna   

Použít filtr pro 2. souboru     Data setříděna   

**Výstup**

Vytvořit novou tabulku     Proměnné 1. souboru   

Zkopírovat formáty z 1. souboru     Proměnné 2. souboru   

Zkopírovat formáty z 2. souboru

Dodržet seřazení výstupu jako ve vstupu   

1. soubor... C:\Tomáš\Newsletter StatSoft\Slučování tabulek\Tabulka1.sta

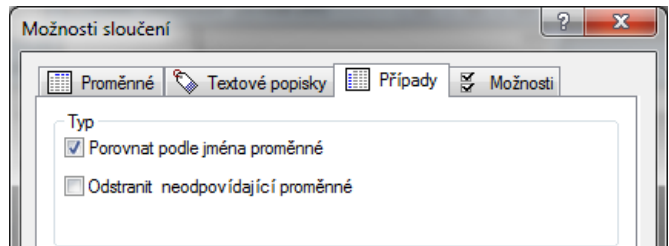
2. soubor... C:\Tomáš\Newsletter StatSoft\Slučování tabulek\Tabulka3.sta

Pozn.: po stisknutí tlačítka OK bude spuštěno sloučení podle aktivní záložky.

## 2. Spojit (porovnat podle jména proměnné)

Spojí tabulky, ale vezme v úvahu názvy proměnných, tedy pokud je v obou souborech stejný název proměnné, pak bude ve výsledku tato proměnná jen jednou (což je rozdíl oproti prvnímu příkladu).

Vše se nastaví jako v příkladu 1, jen v záložce **Případy** zaškrtneme políčko **Porovnat podle jména proměnné**.



Data: Tabulka1.sta\* (3s krát 10ř)

|    | 1 ID | 2 Gender | 3 Measurement |
|----|------|----------|---------------|
| 1  | 1    | M        | 171           |
| 2  | 2    | F        | 172           |
| 3  | 3    | F        | 163           |
| 4  | 4    | M        | 182           |
| 5  | 5    | F        | 174           |
| 6  | 6    | M        | 189           |
| 7  | 7    | M        | 192           |
| 8  | 8    | M        | 174           |
| 9  | 9    | F        | 175           |
| 10 | 10   | M        | 178           |

Data: Tabulka2.sta (3s krát 10ř)

|    | 1 ID | 2 Gender | 3 Measurement |
|----|------|----------|---------------|
| 11 | 11   | F        | 161           |
| 12 | 12   | M        | 152           |
| 13 | 13   | F        | 181           |
| 14 | 14   | M        | 135           |
| 15 | 15   | F        | 171           |
| 16 | 16   | M        | 191           |
| 17 | 17   | F        | 181           |
| 18 | 18   | F        | 151           |
| 19 | 19   | M        | 181           |
| 20 | 20   | M        | 201           |

Data: Tabulka14\* (3s krát 20ř)

|    | 1 ID | 2 Gender | 3 Measurement |
|----|------|----------|---------------|
| 1  | 1    | M        | 171           |
| 2  | 2    | F        | 172           |
| 3  | 3    | F        | 163           |
| 4  | 4    | M        | 182           |
| 5  | 5    | F        | 174           |
| 6  | 6    | M        | 189           |
| 7  | 7    | M        | 192           |
| 8  | 8    | M        | 174           |
| 9  | 9    | F        | 175           |
| 10 | 10   | M        | 178           |
| 11 | 11   | F        | 161           |
| 12 | 12   | M        | 152           |
| 13 | 13   | F        | 181           |
| 14 | 14   | M        | 135           |
| 15 | 15   | F        | 171           |
| 16 | 16   | M        | 191           |
| 17 | 17   | F        | 181           |
| 18 | 18   | F        | 151           |
| 19 | 19   | M        | 181           |
| 20 | 20   | M        | 201           |

Data: Tabulka1.sta\* (3s krát 10ř)

|    | 1 ID | 2 Gender | 3 Measurement |
|----|------|----------|---------------|
| 1  | 1    | M        | 171           |
| 2  | 2    | F        | 172           |
| 3  | 3    | F        | 163           |
| 4  | 4    | M        | 182           |
| 5  | 5    | F        | 174           |
| 6  | 6    | M        | 189           |
| 7  | 7    | M        | 192           |
| 8  | 8    | M        | 174           |
| 9  | 9    | F        | 175           |
| 10 | 10   | M        | 178           |

Data: Tabulka3.sta\* (2s ...)

|    | 1 ID | 2 Numbers |
|----|------|-----------|
| 1  | 1    | 12        |
| 2  | 1    | 14        |
| 3  | 2    | 57        |
| 4  | 2    | 54        |
| 5  | 3    | 14        |
| 6  | 4    | 32        |
| 7  | 5    | 65        |
| 8  | 6    | 88        |
| 9  | 7    | 74        |
| 10 | 8    | 15        |
| 11 | 9    | 16        |
| 12 | 10   | 59        |
| 13 | 10   | 96        |

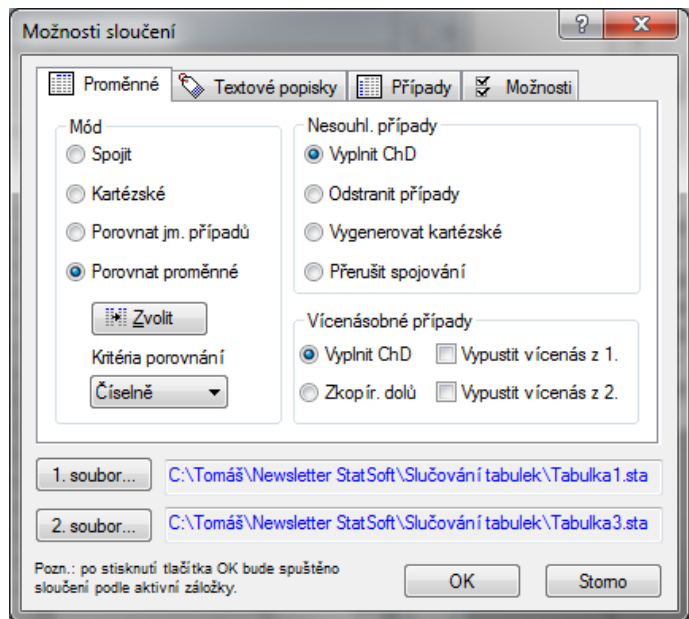
Data: Tabulka15\* (4s krát 23ř)

|    | 1 Gender | 2 Measurement | 3 ID | 4 Numbers |
|----|----------|---------------|------|-----------|
| 1  | M        | 171           | 1    | 12        |
| 2  | F        | 172           | 2    | 14        |
| 3  | F        | 163           | 3    | 57        |
| 4  | M        | 182           | 4    | 54        |
| 5  | F        | 174           | 5    | 14        |
| 6  | M        | 189           | 6    | 32        |
| 7  | M        | 192           | 7    | 65        |
| 8  | M        | 174           | 8    | 88        |
| 9  | F        | 175           | 9    | 74        |
| 10 | M        | 178           | 10   | 15        |
| 11 |          |               | 11   | 16        |
| 12 |          |               | 12   | 59        |
| 13 |          |               | 13   | 96        |
| 14 |          |               | 14   |           |
| 15 |          |               | 15   |           |
| 16 |          |               | 16   |           |
| 17 |          |               | 17   |           |
| 18 |          |               | 18   |           |
| 19 |          |               | 19   |           |
| 20 |          |               | 20   |           |
| 21 |          |               | 21   |           |
| 22 |          |               | 22   |           |
| 23 |          |               | 23   |           |

### 3. Porovnat proměnné

Pokud potřebujeme spojit informace podle nějaké proměnné – typicky to může být například ID, pak je to možné, jen je potřeba udělat několik nastavení: vybereme oba soubory, zvolíme mód **Porovnat proměnné**, vybereme proměnné, které identifikují stejné případy v obou souborech pod tlačítkem **Zvolit**.

Jako **Kritérium porovnání** dále vybereme typ, který odpovídá naší spojovací proměnné.



Následující příklad je situací, kdy k informacím v prvním souboru přidáme informace z druhého souboru pro odpovídající ID (ID tedy bylo vybráno jako proměnná k porovnání).

|    | 1 ID | 2 Gender | 3 Measurement |
|----|------|----------|---------------|
| 1  | 1    | M        | 171           |
| 2  | 2    | F        | 170           |
| 3  | 3    | F        | 163           |
| 4  | 4    | M        | 182           |
| 5  | 5    | F        | 174           |
| 6  | 6    | M        | 189           |
| 7  | 7    | M        | 192           |
| 8  | 8    | M        | 174           |
| 9  | 9    | F        | 175           |
| 10 | 10   | M        | 178           |

|    | 1 ID | 2 Numbers |
|----|------|-----------|
| 1  | 1    | 12        |
| 2  | 1    | 14        |
| 3  | 2    | 57        |
| 4  | 2    | 54        |
| 5  | 3    | 14        |
| 6  | 4    | 32        |
| 7  | 5    | 65        |
| 8  | 6    | 88        |
| 9  | 7    | 74        |
| 10 | 8    | 15        |
| 11 | 9    | 16        |
| 12 | 10   | 59        |
| 13 | 10   | 96        |

|    | 1 ID | 2 Gender | 3 Measurement | 4 Numbers |
|----|------|----------|---------------|-----------|
| 1  | 1    | M        | 171           | 12        |
| 2  | 1    | F        | 170           | 14        |
| 3  | 2    | F        | 163           | 57        |
| 4  | 2    | M        | 182           | 54        |
| 5  | 3    | F        | 174           | 14        |
| 6  | 4    | M        | 189           | 32        |
| 7  | 5    | F        | 174           | 65        |
| 8  | 6    | M        | 189           | 88        |
| 9  | 7    | M        | 192           | 74        |
| 10 | 8    | M        | 174           | 15        |
| 11 | 9    | F        | 175           | 16        |
| 12 | 10   | M        | 178           | 59        |
| 13 | 10   | M        | 178           | 96        |

#### 4. Porovnat proměnné (bez prázdných polí)

Pokud chceme udělat stejnou úlohu jako v příkladu 3, jen potřebujeme informaci z prvního souboru přepsat do všech polí s daným ID, musíme zaškrtnout **Zkopíř. Dolů**.

Možnosti sloučení

Proměnné | Textové popisky | Případy | Možnosti

Mód

- Spojit
- Kartézské
- Porovnat jm. případů
- Porovnat proměnné

Kritéria porovnání

Číselně

Nesouhl. případy

- Vyplnit ChD
- Odstranit případy
- Vygenerovat kartézské
- Přerušit spojování

Vícenásobné případy

- Vyplnit ChD
- Vypustit vícenás z 1.
- Zkopíř. dolů
- Vypustit vícenás z 2.

1. soubor... C:\Tomáš\Newsletter StatSoft\Slučování tabulek\Tabulka1.sta

2. soubor... C:\Tomáš\Newsletter StatSoft\Slučování tabulek\Tabulka3.sta

Pozn.: po stisknutí tlačítka OK bude spuštěno sloučení podle aktivní záložky.

Data: Tabulka1.sta\* (3s krát 10ř)

|    | 1 ID | 2 Gender | 3 Measurement |
|----|------|----------|---------------|
| 1  | 1    | M        | 171           |
| 2  | 2    | F        | 172           |
| 3  | 3    | F        | 185           |
| 4  | 4    | M        | 189           |
| 5  | 5    | F        | 174           |
| 6  | 6    | M        | 189           |
| 7  | 7    | M        | 192           |
| 8  | 8    | M        | 174           |
| 9  | 9    | F        | 175           |
| 10 | 10   | M        | 178           |

Data: Tabulka3.sta\* (2s ...)

|    | 1 ID | 2 Numbers |
|----|------|-----------|
| 1  | 1    | 12        |
| 2  | 1    | 14        |
| 3  | 2    | 57        |
| 4  | 2    | 54        |
| 5  | 3    | 14        |
| 6  | 4    | 32        |
| 7  | 5    | 65        |
| 8  | 6    | 88        |
| 9  | 7    | 74        |
| 10 | 8    | 15        |
| 11 | 9    | 16        |
| 12 | 10   | 59        |
| 13 | 10   | 96        |

Data: Tabulka29\* (4s krát 13ř)

|    | 1 ID | 2 Gender | 3 Measurement | 4 Numbers |
|----|------|----------|---------------|-----------|
| 1  | 1    | M        | 171           | 12        |
| 2  | 1    | M        | 174           | 14        |
| 3  | 2    | F        | 172           | 57        |
| 4  | 2    | F        | 172           | 54        |
| 5  | 3    | F        | 163           | 14        |
| 6  | 4    | M        | 182           | 32        |
| 7  | 5    | F        | 174           | 65        |
| 8  | 6    | M        | 189           | 88        |
| 9  | 7    | M        | 192           | 74        |
| 10 | 8    | M        | 174           | 15        |
| 11 | 9    | F        | 175           | 16        |
| 12 | 10   | M        | 178           | 59        |
| 13 | 10   | M        | 178           | 96        |

## 5. Porovnat proměnné (bez vícenásobných)

Pokud bychom chtěli oproti příkladu 4 naopak jen první výskyt ID z každého souboru, zvolili bychom **Vypustit vícenásobné**.

The 'Možnosti sloučení' dialog box is shown with the following settings:

- Mód:** Porovnat proměnné (selected)
- Nesouhl. případy:** Vyplnit ChD (selected)
- Vícenásobné případy:** Vypustit vícenás z 1. (checked), Vypustit vícenás z 2. (checked)

The dialog also shows two source files:

1. soubor... C:\Tomáš\Newsletter StatSoft\Slučování tabulek\Tabulka1.sta
2. soubor... C:\Tomáš\Newsletter StatSoft\Slučování tabulek\Tabulka3.sta

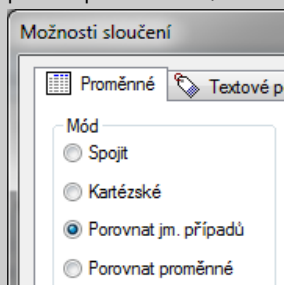
Below the dialog, three data tables are shown with arrows indicating the mapping of data from the source files to the merged result:

|    | 1 ID | 2 Gender | 3 Measurement |
|----|------|----------|---------------|
| 1  | 1    | M        | 171           |
| 2  | 2    | F        | 172           |
| 3  | 3    | F        | 163           |
| 4  | 4    | M        | 160           |
| 5  | 5    | F        | 174           |
| 6  | 6    | M        | 189           |
| 7  | 7    | M        | 192           |
| 8  | 8    | M        | 174           |
| 9  | 9    | F        | 175           |
| 10 | 10   | M        | 178           |

|    | 1 ID | 2 Numbers |
|----|------|-----------|
| 1  | 1    | 12        |
| 2  | 2    | 57        |
| 3  | 2    | 57        |
| 4  | 2    | 54        |
| 5  | 3    | 14        |
| 6  | 4    | 32        |
| 7  | 5    | 65        |
| 8  | 6    | 88        |
| 9  | 7    | 74        |
| 10 | 8    | 15        |
| 11 | 9    | 16        |
| 12 | 10   | 59        |
| 13 | 10   | 96        |

|    | 1 ID | 2 Gender | 3 Measurement | 4 Numbers |
|----|------|----------|---------------|-----------|
| 1  | 1    | M        | 171           | 12        |
| 2  | 2    | F        | 172           | 57        |
| 3  | 3    | F        | 163           | 14        |
| 4  | 4    | M        | 182           | 32        |
| 5  | 5    | F        | 174           | 65        |
| 6  | 6    | M        | 189           | 88        |
| 7  | 7    | M        | 192           | 74        |
| 8  | 8    | M        | 174           | 15        |
| 9  | 9    | F        | 175           | 16        |
| 10 | 10   | M        | 178           | 59        |

**Poznámka:** Pokud chceme sloučit soubory podle názvů případů (analogicky jako v příkladech 3, 4 a 5 podle proměnné) v souborech, pak zvolíme mód **Porovnat jm. případů**.



## 6. Kartézské sloučení

Poslední mód, který jsme si ještě neukázali je kartézské sloučení. Ve výsledném souboru se každý řádek z jednoho souboru bude vyskytovat s každým řádkem souboru druhého.

Tuto možnost můžete využít například při tvorbě schémat pro analýzy.

The screenshot shows the 'Možnosti sloučení' (Join Options) dialog box. Under the 'Mód' (Mode) section, 'Kartézské' (Cartesian) is selected. The 'Nesouhl. případy' (Inconsistent cases) section has 'Vyplnit ChD' (Fill with #N/A) selected. The 'Vícenásobné případy' (Multiple cases) section has 'Vyplnit ChD' (Fill with #N/A) selected. The dialog also shows file paths for two source files: 'C:\Tomáš\Newsletter StatSoft\Slučování tabulek\Tabulka1.sta' and 'C:\Tomáš\Newsletter StatSoft\Slučování tabulek\Tabulka4.sta'.

Below the dialog, three data tables are shown. The first table, 'Data: Tabulka1.sta\* (3s krát 10ř)', has columns '1 ID', '2 Gender', and '3 Measurement'. The second table, 'Data: Tabulka4.sta\* (...)', has columns '1 Variant' and '2 Measurement'. The third table is the result of the Cartesian join, with columns 'ID', 'Gender', 'Measurement', and 'Variant', showing 20 rows (10 from each source file).

**Poznámka:** V tomto příkladu jsme použili formátování, k jednoduššímu vysvětlení toho, co dělá mód **Kartézské**. Aby bylo zachováno ve výsledku formátování, stačí zaškrtnout políčka v záložce **Možnosti**.

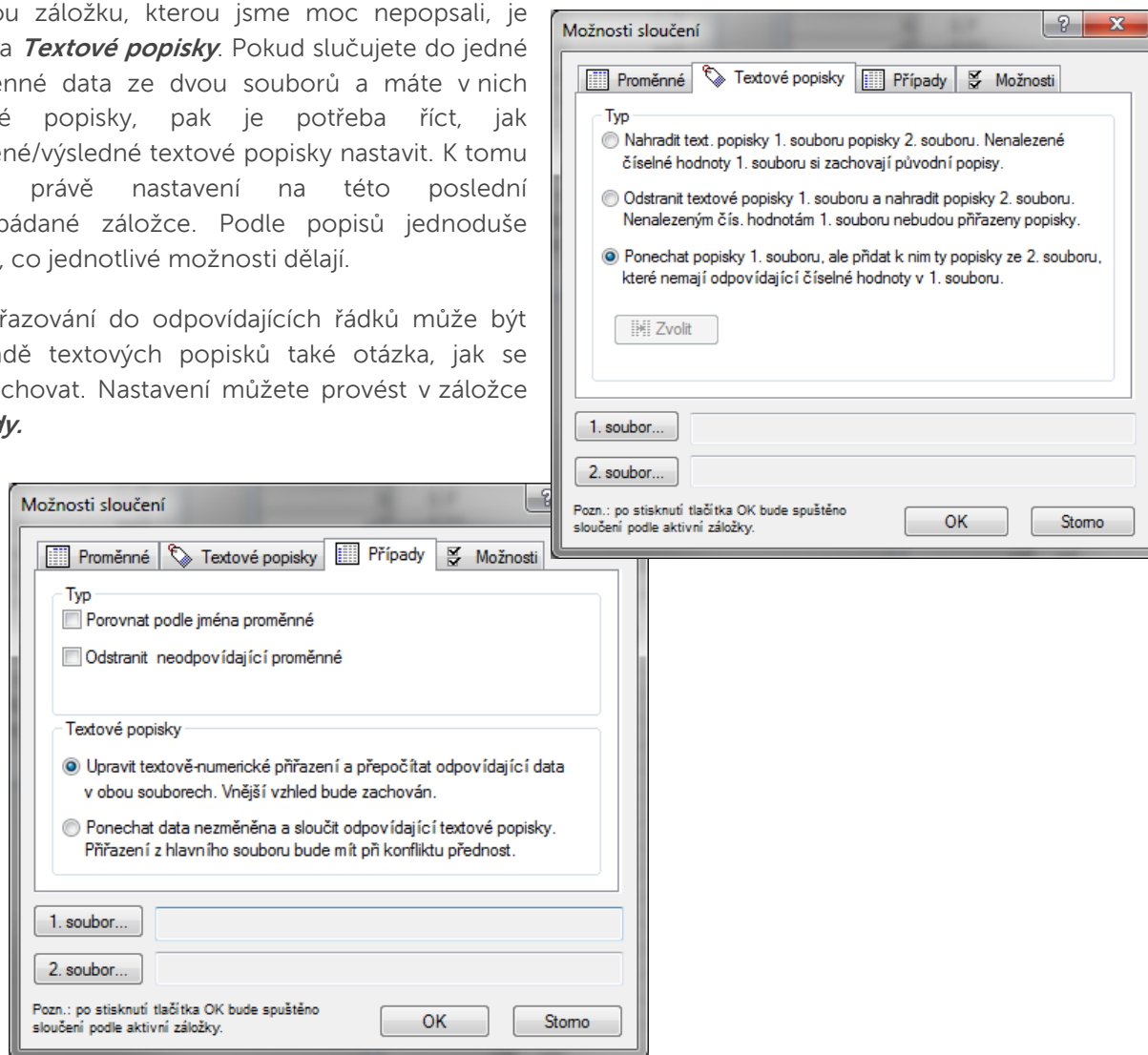
The screenshot shows the 'Možnosti sloučení' dialog box, 'Možnosti' (Options) tab. The 'Vstup' (Input) section has 'Použít filtr pro 1. soubor' and 'Použít filtr pro 2. soubor' unchecked. The 'Výstup' (Output) section has 'Vytvořit novou tabulku' checked, 'Zkopírovat formáty z 1. souboru' checked, and 'Zkopírovat formáty z 2. souboru' checked. The 'Dodržet seřazení výstupu jako ve vstupu' (Preserve output order as in input) checkbox is unchecked.

## Shrnutí funkcionality „Sloučit“

Jak jste mohli vidět, funkcionalita je celkem silná a dovolí Vám spojit tabulky nejrůznějším způsobem. To má za následek to, že možností nastavení je celkem dost a nemusí být úplně jednoduché vše nastavit správně napoprvé. A proto jsme také vytvořili tento návod s příklady. Ukázali jsme si nejběžnější využití a také postupně vysvětlili téměř všechny možnosti a nastavení funkcionality pro spojování dat ze dvou tabulek.

Jedinou záložku, kterou jsme moc nepopsali, je záložka **Textové popisky**. Pokud slučujete do jedné proměnné data ze dvou souborů a máte v nich textové popisky, pak je potřeba říct, jak sloučené/výsledné textové popisky nastavit. K tomu slouží právě nastavení na této poslední neprobádané záložce. Podle popisů jednoduše zjistíte, co jednotlivé možnosti dělají.

Při přiřazování do odpovídajících řádků může být v případě textových popisků také otázka, jak se k nim chovat. Nastavení můžete provést v záložce **Případy**.





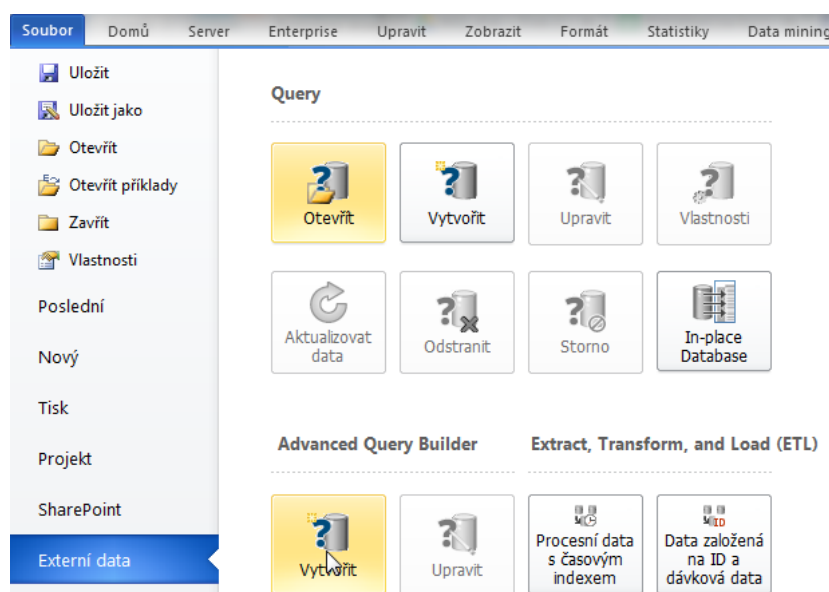
## Další možnosti slučování

Další možností, jak slučovat tabulky, je forma databázového dotazu.

Lze využít prostředí **Advance Query Builder** nebo **Query Builder** na uměle vytvořenou databázi ze souborů Statistica Spreadsheet (koncovka *sta*). Detaily ohledně ovládání těchto modulů byly zveřejněny v [newsletteru 28/05/2014](#).

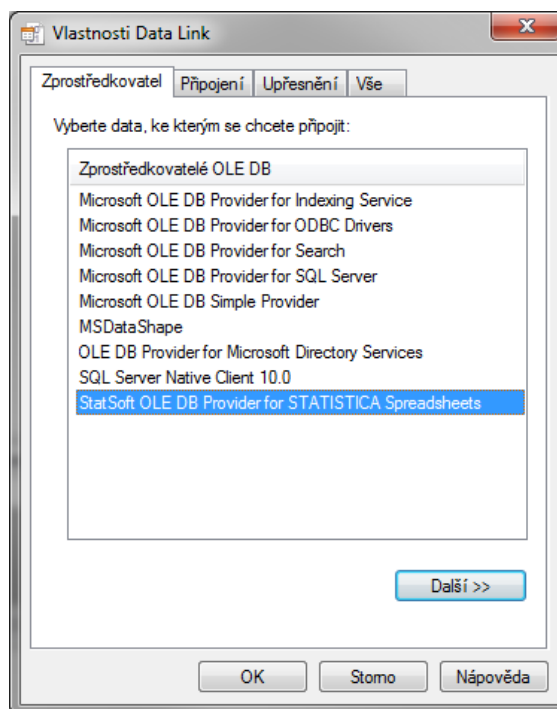
Ukážeme, jak si takovou databázi z vlastních tabulek Statistica vytvořit.

1. Otevřeme funkcionalitu **Advanced Query Builder** : **Soubor** → **Externí data** → **Advanced Query builder** → **Vytvořit**



2. **Action** → **Connect** → **StatSoft OLE DB Provider for STATISTICA Spreadsheets**

3. V záložce **Připojení** je potřeba vybrat složku, ve které jsou soubory, které chceme spojovat. Případně vybrat přímo soubory.
4. Po potvrzení **OK** budou k dispozici pro tvorbu dotazu vybrané tabulky.



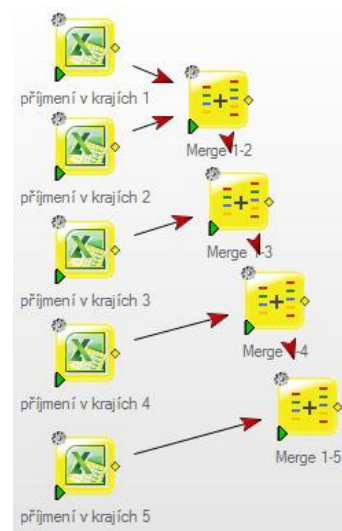
Výhody oproti funkcionalitě **Sloučit**.

- › Možnosti agregace.
- › Možnost spojení více tabulek najednou.
- › Využití všech funkcionalit SQL jazyka.
- › Dostupnost náhledu na výsledná data ještě před aplikací dotazu.
- › Možnost přímého seřazení

## Uzly v prostředí Workspace

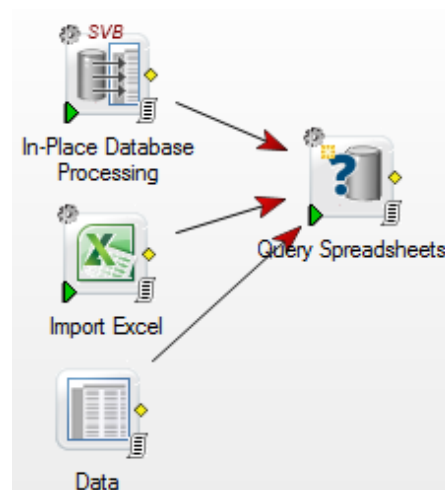
Abychom dodali kompletní obrázek o možnostech slučování, je potřeba zmínit možnosti v prostředí Workspace.

Je zde samozřejmě funkcionalita sloučit, jak jsme ji představili v prvních šesti příkladech:



Je zde ale i unikátní a velmi užitečný uzel, který se jmenuje **Query Spreadsheets**. Je jím možné vytvářet dotazy pomocí Query nástrojů zmíněných výše. Je možné udělat dotaz nad všemi tabulkami, které směřují do tohoto uzlu.

Toto využití je velmi výhodné, pokud máme data z různých zdrojů, stejně jako na obrázku:



Více se o prostředí Workspace můžete dozvědět v našich newsletterových článcích:

[http://www.statsoft.cz/file1/PDF/newsletter/2013\\_04\\_02\\_StatSoft\\_Data\\_miner\\_workspaces.pdf](http://www.statsoft.cz/file1/PDF/newsletter/2013_04_02_StatSoft_Data_miner_workspaces.pdf)

[http://www.statsoft.cz/file1/PDF/newsletter/2013\\_12\\_18\\_StatSoft\\_Odkud\\_tak\\_asi\\_je.pdf](http://www.statsoft.cz/file1/PDF/newsletter/2013_12_18_StatSoft_Odkud_tak_asi_je.pdf)

[http://www.statsoft.cz/file1/PDF/newsletter/2014\\_02\\_26\\_StatSoft\\_Priklady\\_workspace.pdf](http://www.statsoft.cz/file1/PDF/newsletter/2014_02_26_StatSoft_Priklady_workspace.pdf)

[http://www.statsoft.cz/file1/PDF/newsletter/2014\\_02\\_26\\_StatSoft\\_Prostredi\\_workspace\\_ve\\_verzi\\_12.pdf](http://www.statsoft.cz/file1/PDF/newsletter/2014_02_26_StatSoft_Prostredi_workspace_ve_verzi_12.pdf)

[http://www.statsoft.cz/file1/PDF/newsletter/2014\\_12\\_19\\_StatSoft\\_Copak\\_nam\\_to\\_tady\\_vyrostlo.pdf](http://www.statsoft.cz/file1/PDF/newsletter/2014_12_19_StatSoft_Copak_nam_to_tady_vyrostlo.pdf)