

Textové popisky

Newsletter Statistica ACADEMY



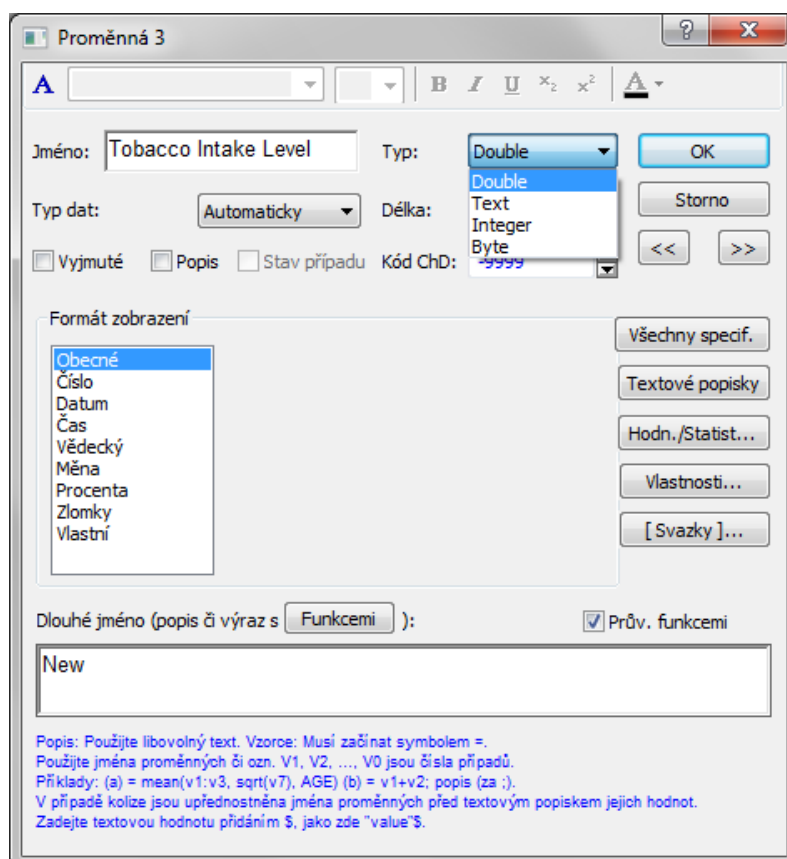
Téma: Možnosti softwaru, datová reprezentace
Typ článku: Tipy a triky

Máte ve svých datech kategorie ve formě textu? Víme, že někdy není úplně jasné, jak Statistica s těmito proměnnými pracuje, proto jsme se rozhodli vše vysvětlit tímto článkem.

Typ dat

Začněme tím, jak jsou data reprezentována. Každá buňka Vaší tabulky je nějakým způsobem uložena v paměti počítače, tedy každá buňka má nějaký datový typ. Buňky jedné proměnné (jednoho sloupce) mají vždy stejnou reprezentaci – stejný datový typ (to je proto, že proměnná obsahuje stejný typ informace pro daný případ - řádek). Datový typ se dá zkontrolovat či změnit v dialogu proměnné (vyvoláte například dvojklikem na záhlaví proměnné). Pokud si rozkliknete možnosti pro typ proměnné (jako na obrázku vpravo), vidíte, že zde máme 4 typy. 3 číselné *Byte* (může obsahovat pouze dvě hodnoty), *Integer* (jedná se o celá čísla), *Double* (desetinná čísla), *Text* (textový řetězec, maximálně dané délky).

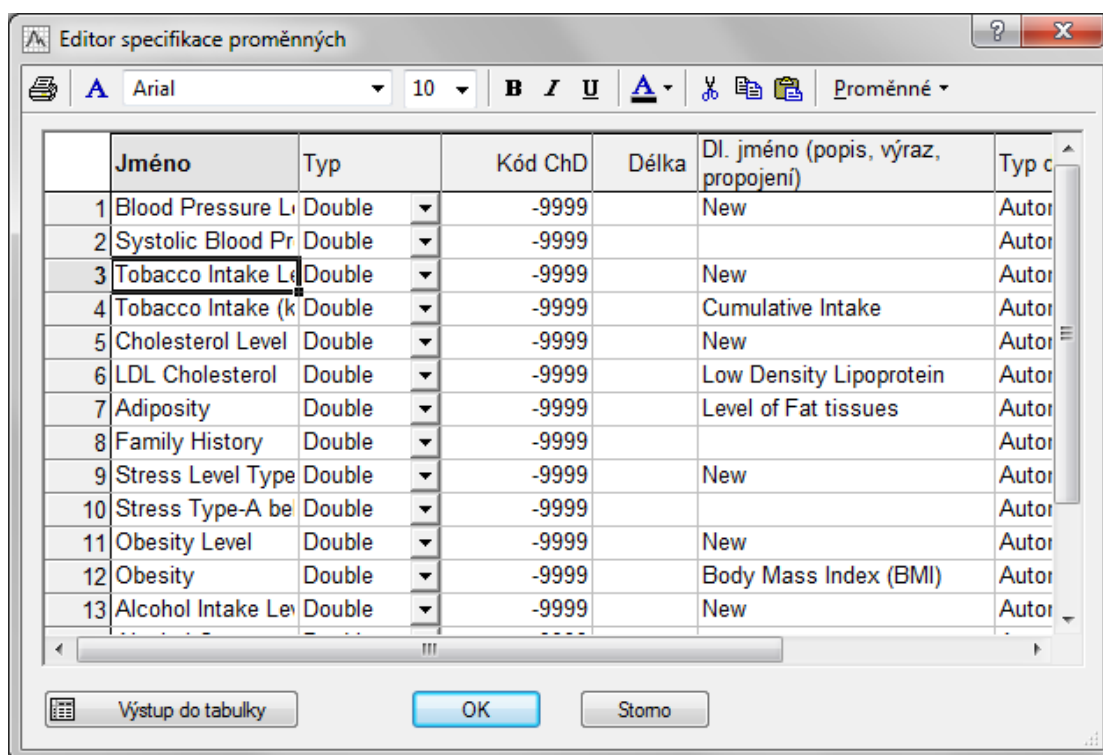
Pro běžné použití většinou



nepotřebujeme typy Integer a Byte, vystačíme si s Double. Tyto typy se používají, pokud potřebujeme šetřit prostor na disku, v paměti totiž zaberou méně místa.

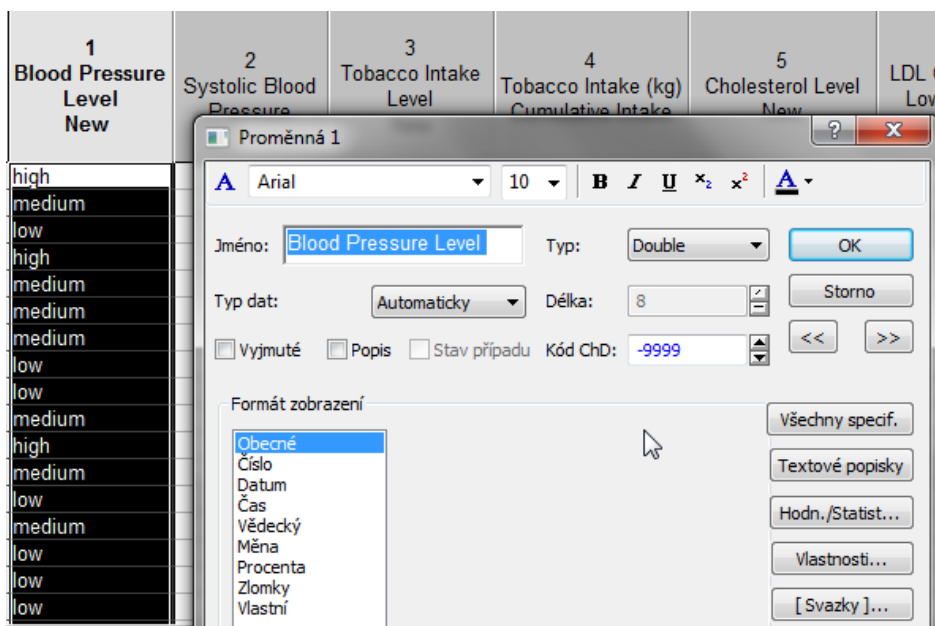
Hlavní informací je, že z těchto typů poznáme, jestli se jedná o číslo nebo text. Pokud načtete data a proměnná, která měla být číslo je typu text, pak je jistě někde chyba (typicky může jít o načtení nějakého textu, který do proměnné nepatří – například komentář na konci souboru, nebo špatně nadefinované desetinné čárky – Statistica bere desetinnou čárku podle nastavení ve Windows).

Poznámka: Pokud chcete vidět najednou všechny typy u všech proměnných najednou (může se hodit, pokud například chcete zkontrolovat, jestli se Vám data načtela ve správném formátu, typicky chcete vědět, jestli program rozpoznal čísla jako čísla). Tyto informace uvidíte v **Editoru specifikace proměnných** (při otevření kterékoli proměnné získáte tento dialog po stisknutí tlačítka **Všechny proměnné**).



Co jsou textové popisky

Aby to ale nebylo tak jednoduché (tedy pokud je proměnná typu Text, tak je to text), tak je zde další možnost a tou je numerická proměnná s textovými popisky. Proměnná je tedy typu například *Double*, na obrazovce ale vidíte text. Abychom si to ukázali, otevřeme si například soubor *HeartDisease.sta* z příkladů softwaru Statistica (*Soubor – Otevřít příklady - Datasets*). Vidíme, že typ je Double, ale v datech vidíme hodnoty *high, medium, low*.

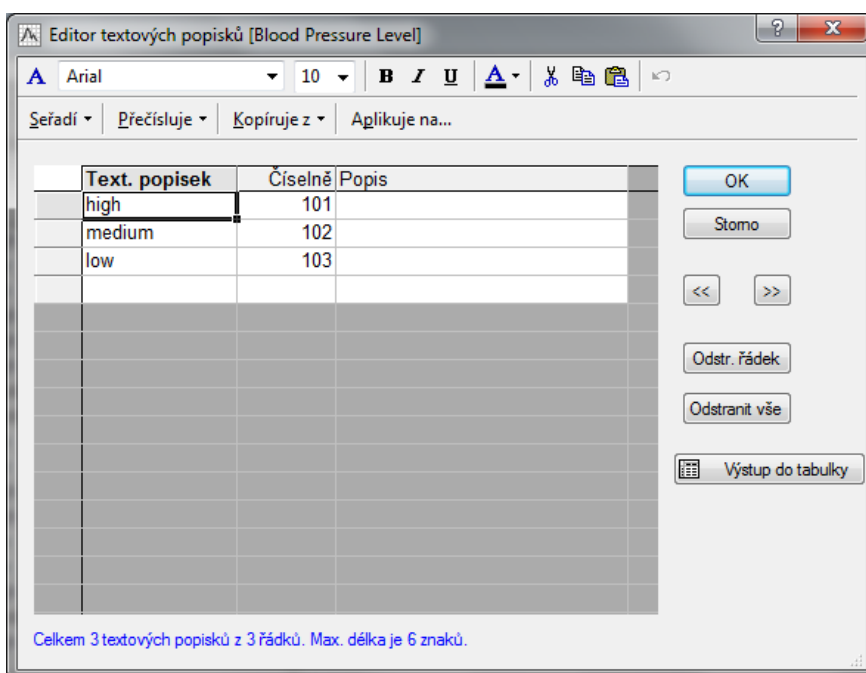


Klikněme nyní na tlačítko **Textové popisky**.

Hodnoty *high*, *medium*, *low* jsou textové popisky, které jsou přiřazeny k hodnotám 101, 102 a 103.

Jinak řečeno, proměnná je číselná, protože obsahuje hodnoty 101, 102 a 103, ale pokaždé, co se tato hodnota vyskytne, tak je zobrazena jako odpovídající textový popisek. Zkuste si sami napsat někam do této proměnné číslo 101 – ihned se Vám přepíše na *high*, stejně dobře ale můžete napsat přímo *high*.

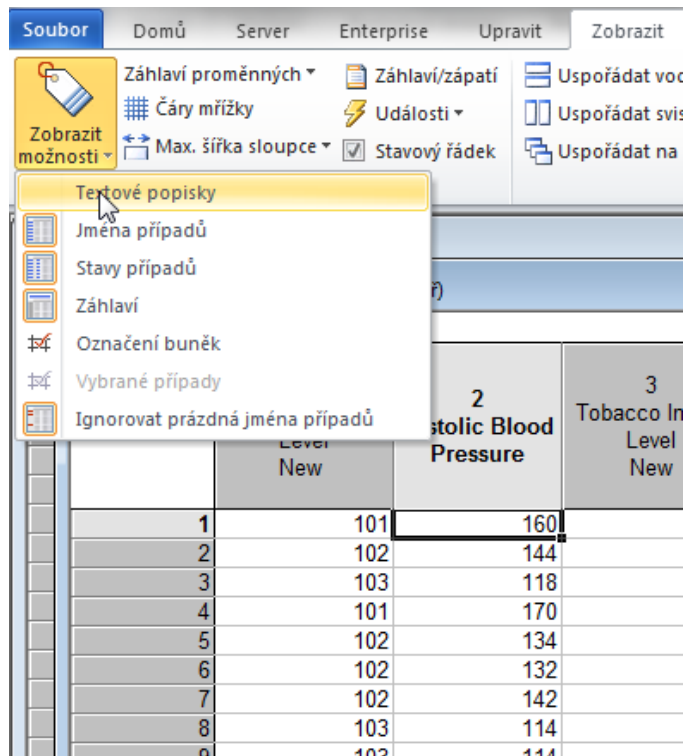
Pokud byste napsali jiné číslo než uvedené v popiscích, pak bude v datech uvedeno jako číslo (tedy bez popisku).



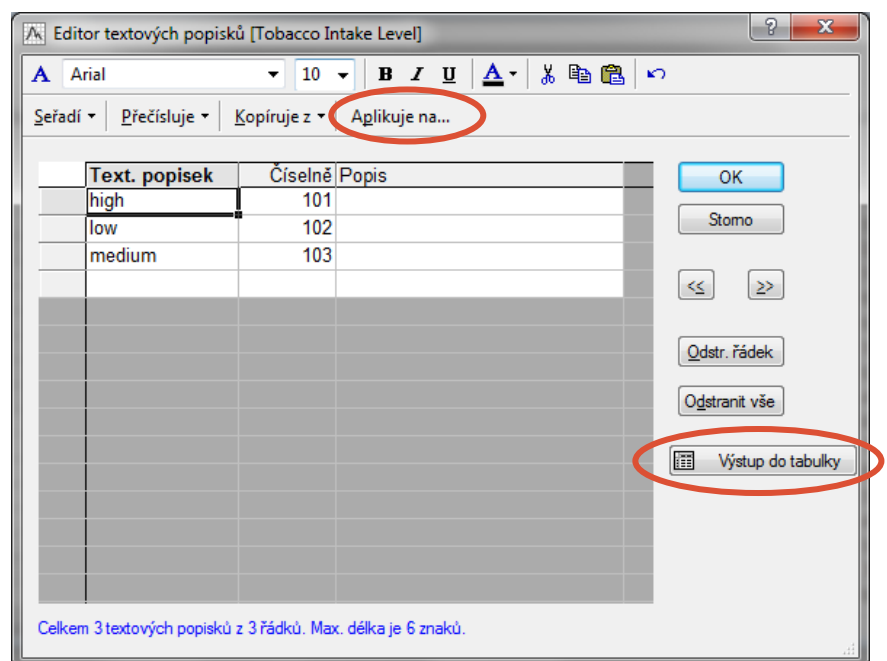
Textové popisky tedy umožňují mít v jedné proměnné číselnou i textovou informaci o hodnotách proměnné, což v jiných softwarech nemusí být běžné (a proto jsme se rozhodli také Vám tuto možnost podrobně ukázat). Toho se dá využít například k lepšímu přehledu v souboru, proměnná může být například nějaký číselník, pak je možné díky textovým popiskům vidět hodnoty, ale zároveň mít v souboru i kódy. Dále může textový popisek pomoci, pokud chceme data seřadit, ale toto seřazení není podle abecedy. Můžeme totiž řadit podle číselných hodnot. Podobně u grafů, kde jsou zobrazeny kategorie, můžete pořadí kategorií ovlivnit textovými popisky.

Pokud číselné hodnoty v editoru textových popisků přepíšete na jiná čísla, pak výstup bude pořád stejný, jen na pozadí bude jiná číselná reprezentace, stejně tak můžete do seznamu textových popisků přidat i další hodnoty, které se v datech třeba ani nevyskytují (to dává smysl například u proměnných typu měsíc v roce, kdy se očekává, že tato hodnota může nastat, ale zrovna nenastala).

Pokud byste se chtěli nezobrazovat popisky nebo si na chvíli přepnout vzhled na číselnou reprezentaci, pak je možné využít následující možnost: **Zobrazit – Zobrazit možnosti – Textové popisky**.



Někdo může využít například funkci pro kopírování textových popisků do jiných proměnných nebo export tabulky textových popisků do samostatné nové tabulky.



Textové popisky každé proměnné jsou uloženy spolu se souborem tabulky Statistica.

Závěr

Cílem tohoto článku bylo poukázat na fakt, že pokud vidíme textovou hodnotu v datech, může to znamenat buď textovou proměnnou nebo numerickou proměnnou s textovými popisky.

Pro použití v analýzách bude software předpokládat, že proměnná s textovými popisky je kategorického typu. V případě, že vyberete proměnnou s textovými popisky pro analýzu, kde se předpokládají spojitě proměnné, pak software vyhodí varovnou hlášku. Pokud i tak použijete proměnnou pro analýzu, software spočítá hodnoty z číselných hodnot. Typickým příkladem je, že někdo spočítá průměr pro kategorie. Pak mu typicky vyjde například 101,25, což jistě není výsledek, který by byl jakkoli opodstatněný či smysluplný.

A kde se vzaly hodnoty 101, 102, ...? Pokud načítáme data z jiného než sta souboru a některá data jsou textového charakteru, pak software doplní k těmto textovým hodnotám číselné hodnoty počítaje 101, 102, ... podle výskytu dané textové hodnoty v souboru. 101 tedy dostane první textová hodnota, která se v daném sloupci vyskytla.