



# NOVINKY IBM SPSS STATISTICS

Nové nástroje a funkce pro urychlení  
a zjednodušení analýz

IBM SPSS Statistics

IBM SPSS Software  
Business Analytics

## Hlavní rysy:

- Nové statistické procedury: Modely zrychleného času selhání, Alternativy lineární Metody nejmenších čtverců.
- Vylepšení procedur Lineárních smíšených modelů a v Zobecněných lineárních smíšených modelech.
- Nový typ grafu Violin Plot.
- Zlepšení praktické použitelnosti.

Silver  
Business  
Partner

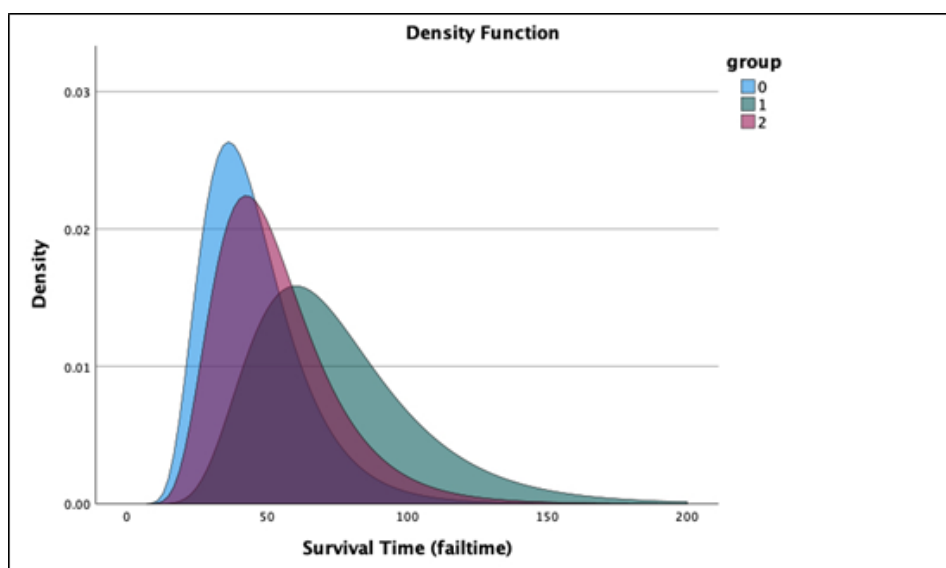


Competency  
Data Science &  
Business Analytics

## Modely zrychleného času selhání (Accelerated Failure Time, AFT)

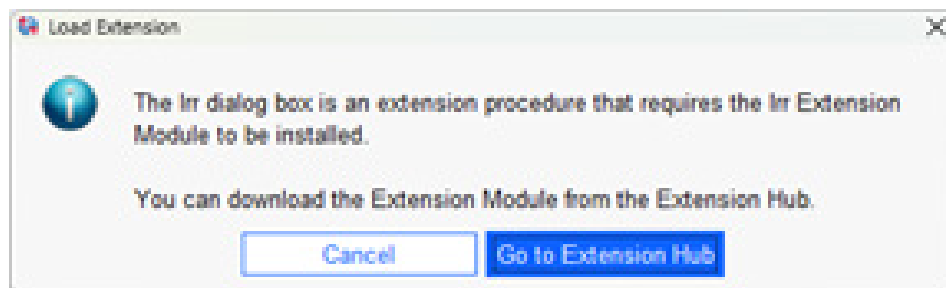
Verze IBM SPSS Statistics 29 přináší nový model do procedur analýzy přežití. Na rozdíl od stávajících tabulek přežití, Kaplan-Meierovy a Coxovy regrese, má nově přidáný model zrychleného času selhání parametrický charakter. To znamená, že se předpokládá, že závislá proměnná má specifické rozdělení. Parametrické modely se často považují za méně flexibilní než neparametrické modely, ale pokud závislá proměnná má identifikovatelné rozdělení, mohou být tyto druhy postupů velmi účinné. Zatímco modely proporcionálních rizik předpokládají, že účinkem kovariátu je znásobení rizika nějakou konstantou, model AFT předpokládá, že účinky kovariátu zkracují nebo prodlužují dobu přežití o nějakou konstantu. Tato schopnost může být užitečná pro výzkumníky, kteří zkoumají čas do selhání jako součást režimu preventivní údržby, zejména pokud je známo, že faktory jako míra využití zařízení urychlují nebo zpomalují čas do selhání.

Nová procedura podporuje parametrické modely založené na Weibullově, exponenciálním, log-normálním a log-logistickém rozdělení. Procedura je součástí modulu Regression.



## Alternativy lineární Metody nejmenších čtverců: Lasso, Ridge a Elastic Net (Linear OLS Alternatives)

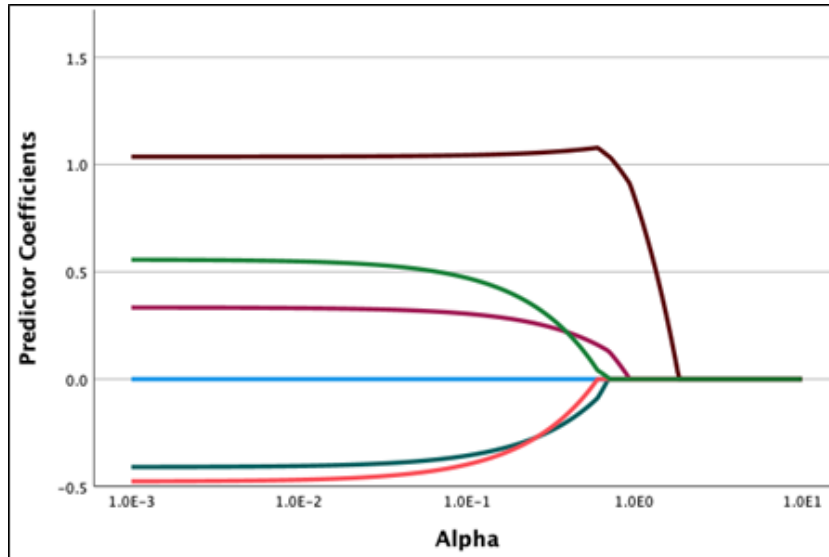
Verze 29 obsahuje také možnost instalace tří nových regresních procedur, které využívají různé formy regularizace. Procedury Lasso, Ridge a Elastic Net mohou být přidány do podnabídky Regression. Pomocí knihovny strojového učení Python Scikit-learn, lze tyto postupy rychle nainstalovat jako rozšíření.



Všechny tyto techniky jsou optimalizovány, aby se předešlo problémům s multikolinearitou, které jsou nezdědka spojeny s metodou nejmenších čtverců. Obecně tyto regulační techniky fungují tak, že penalizují velké modelové koeficienty.

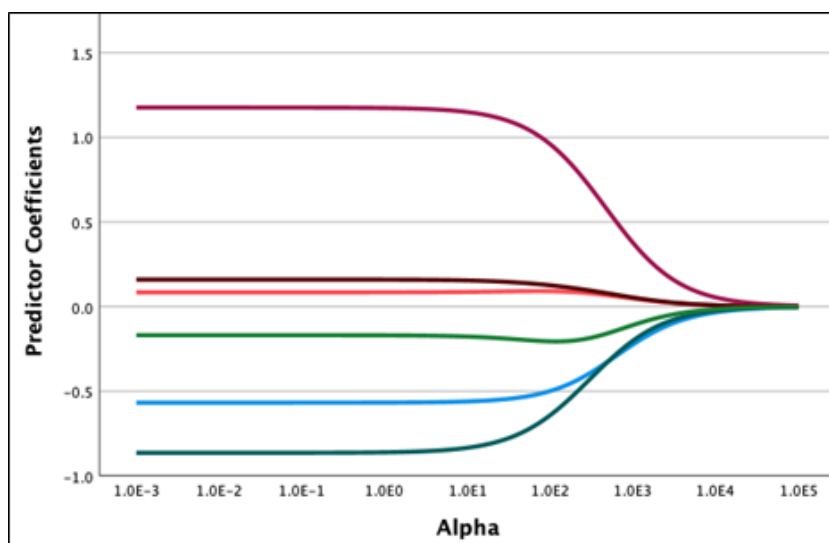
## Lasso

Procedura Lasso (Least Absolute Shrinkage and Selection Operator) často označována jako regularizace L1, funguje tak, že penalizuje nejméně důležité parametry a zmenšuje je k nule. Užitečná je při výběru proměnných, protože slabé proměnné se efektivně vynulují, čímž se zjednoduší konečný model.



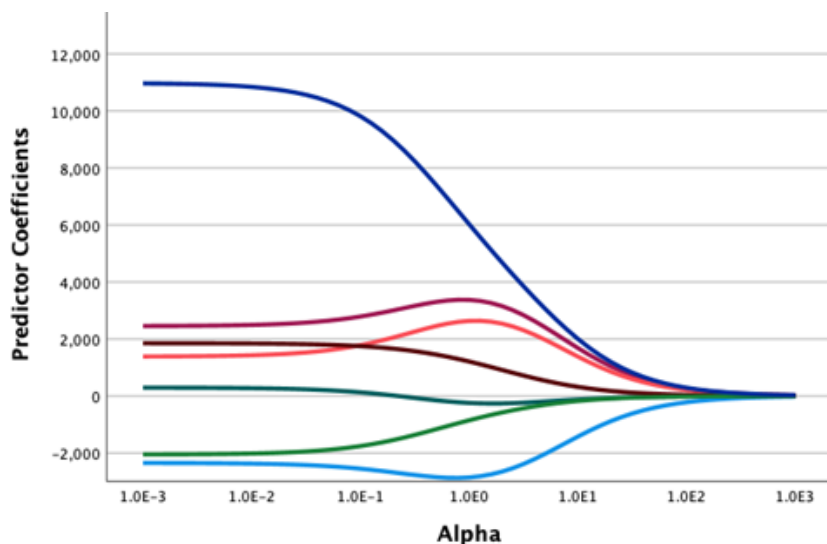
## Hřebenová regrese

Regularizace L2, známá jako Hřebenová regrese, má tendenci penalizovat koeficienty rovnoměrnějším způsobem než L1. Kromě vytváření jednoduších modelů se běžně používá při řešení otázek multikolinearity.



## Elastic Net

Elastic Net kombinuje Lasso (L1) a hřebenovou regresi (L2), což může vést k vyváženějšímu modelu, je-li každá jednotlivá metoda nějakým způsobem suboptimální.



## Míry Pseudo-R<sup>2</sup> v Lineárních smíšených modelech a v Zobecněných lineárních smíšených modelech

Výstup z lineárních smíšených modelů a zobecněných lineárních smíšených modelů nyní zahrnuje míry pseudo-R<sup>2</sup> a koeficient vnitrotřídní korelace - ICC. R<sup>2</sup> je běžně uváděná statistika označující podíl rozptylu vysvětleného lineárním modelem. Koeficient vnitrotřídní korelace (ICC) je statistika, která uvádí, jak velká část reziduálního rozptylu se vysvětluje skupinovým faktorem ve víceúrovňových/hierarchických modelech.

## Podpora vyšších verzí Python a R

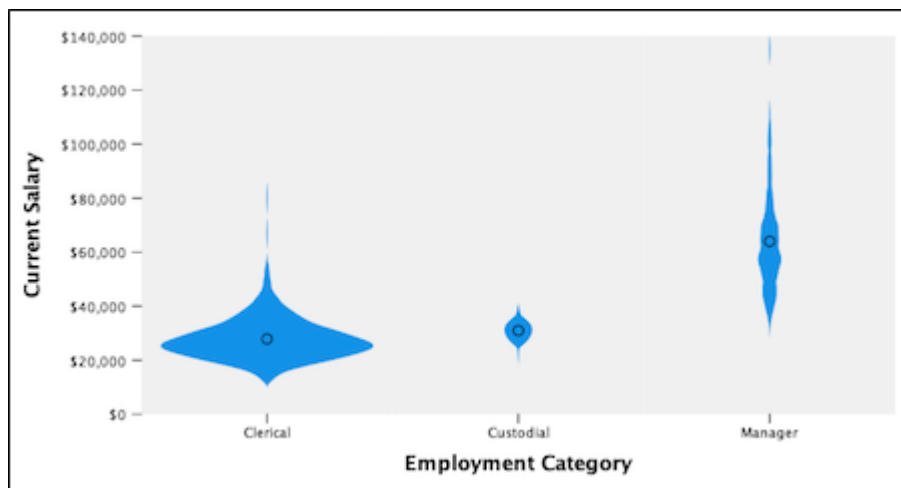
S verzí IBM SPSS Statistics 29 je nainstalován Python 3.10.4 a R 4.2.0

## Odfiltrované případy lze znovu zobrazit

Odfiltrované případy již nejsou skryty v editoru dat, když je vybrána podmnožina případů a nevybrané případy se nesmažou. Jedná se o návrat chování SPSS Statistics 27.0.1 a starších verzí.

## Violin Plot

Violin Plot byly přidány do nabídky Grafboard Template Chooser. Tyto grafy jsou kombinací krabicových grafů a grafů hustoty jádra. Violin Plot zobrazují vrcholy dat a používají se pro vizualizaci rozdělení spojitých proměnných. Na rozdíl od krabicového grafu, který může zobrazovat pouze souhrnné statistiky, Violin Plot zobrazují souhrnné statistiky a hustotu každé proměnné.



## Vylepšení režimu Workbook

Pro uživatele pracující v oknech Workbook byla přidána dvě nová tlačítka na panelu nástrojů: Zobrazit/Skrýt všechna okna syntaxe a Vymazat veškerý výstup. Na stavový řádek bylo přidáno nové tlačítko, které umožňuje uživatelům rychle přepínat se mezi režimem (samostatná okna Výstup a Syntax) a režimem Workbook.

**Statistics**

Level of education	
N	Valid 850
	Missing 0

**Level of education**

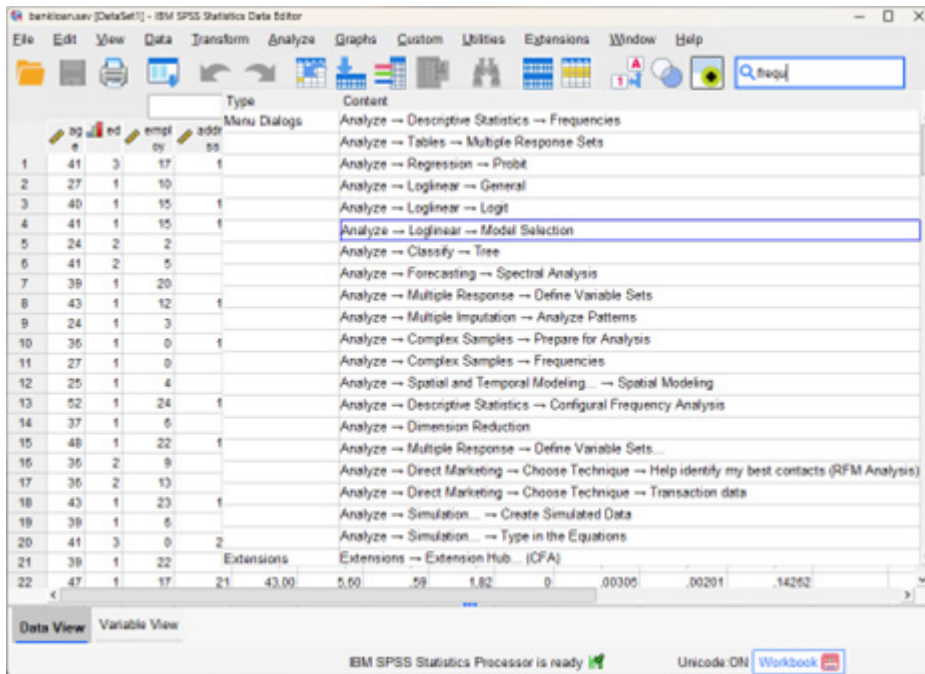
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Did not complete high school	460	54,1	54,1	54,1
High school degree	235	27,6	27,6	81,8
Some college	101	11,9	11,9	93,6
College degree	49	5,8	5,8	99,4
Post undergraduate degree	5	,6	,6	100,0
Total	850	100,0	100,0	

IBM SPSS Statistics Processor is ready

Unicode ON | Workbook

## Vylepšení vyhledávání

Funkce hledat nyní poskytuje možnosti pro zadávání výrazů přímo do panelu nástrojů, což uživatelům umožňuje zobrazit výsledky vyhledávání v rozbalovacím panelu.



## Přidáno ve verzi 28

- Nová statistická procedura meta-analýza
- Vylepšení procedur Power analýza a Ratio statistics
- Nový typ grafu Relationship map
- Zlepšení každodenní použitelnosti

### Meta-analýza (Meta analysis)

Novinkou verze 28 je nová statistická procedura Meta-analýza (meta analysis). Meta-analýza je vědecká metoda, která souhrnně analyzuje data z několika na sobě nezávislých studií. Cílem je identifikace a kvantifikace převažujících trendů nebo zjištění příčin rozdílných závěrů studií. IBM SPSS Statistics 28 představuje celkem 5 nových procedur. Mimo jiné obsahují testy heterogenity, meta-regresní analýzu, či několik nových typů grafů pro lepší vizualizaci výsledků.

#### Seznam procedur:

1. Meta-analysis of continuous outcomes: Raw data, precalculated effect size data
2. Meta-analysis of binary outcomes: Raw data, precalculated effect size data
3. Meta-regression

### Power analýza

Power analýza, která byla představena ve verzi 27 přichází s novým vylepšením procedury. Power analýza vám dovolí určit optimální rozsah výběru pro vaši studii nebo projekt. Pokud je získání dat drahé, je důležité vědět, kolik pozorování je třeba získat, aby se dosáhlo dostatečné pravděpodobnosti detekce určeného efektu. Novinkou verze 28 je možnost zadat několik samostatných hodnot požadované síly (power). Hodnoty mohou být zadány buď jednotlivě nebo je možné zadat celý rozsah, ve kterém se mají pohybovat. Tato novinka byla přidána do všech procedur Power analýzy.

### Ratio

Nejzásadnější změnou procedury v nové verzi je přidání nové statistiky PRB (price-related bias) a příslušný interval spolehlivosti. V hlavním okně procedury je možnost zahrnout počet případů, ze kterých byly příslušné statistiky počítány do výstupové tabulky.

### Relationship map

Jednou z hlavních novinek ve verzi 28 je nový typ grafu, tzv. Relationship map. Pomocí tohoto typu grafu lze vizuálně

zkoumat vzájemné vztahy mezi několika kategorizovanými proměnnými. Graf vizuálně znázorňuje vzájemné vazby pomocí uzlů a spojnic. Uzly představují proměnné a jejich kategorie, tloušťka spojnic mezi uzly reprezentuje sílu vztahu mezi nimi. Graf je možné najít na záložce Graphs -> Relationship Map.

### Další vylepšení procedur

V uživatelském rozhraní procedury GLM (General Linear Model) nyní navíc přibýlo pod volbou EM Means nové zaškrtačkové políčko Compare simple main effects. Procedura One-way ANOVA nyní podporuje nečíselné kategorizované proměnné.

### ZLEPŠENÍ KAŽDODENNÍ POUŽITELNOSTI

IBM SPSS Statistics 28 představuje nové funkce a rozšíření funkcí, jejichž cílem je zlepšit každodenní použitelnost a celkovou produktivitu.

#### Workbook

IBM SPSS Statistics 28 přichází s novým typem okna Workbook. Mimo jiné umožňuje toto okno zároveň spuštění syntaxe a zobrazení jejího výsledku. Další výhodou je možnost plné editace syntaxe, výstupů nebo i textu, přímo v tomto okně. Soubory tohoto formátu mají příponu .spwb. Pro aktivaci je nutné změnit nastavení pod volbou Edit > Options > Application Mode na Workbook

#### Side-panel table editor

Verze 28 představuje nové vylepšení editace tabulek. Editační okno nyní obsahuje boční panel, který zahrnuje všechny funkce z nabídky Cell Properties a Table Properties, které byly původně obsaženy pod volbou Format, což značně zkrátí a zjednoduší čas strávený editací tabulek.

#### Search

Funkce Search, která byla poprvé představena ve verzi 27, nyní prohledává všechna slova / výrazy v každém dialogovém okně uživatelského rozhraní.

#### High contrast

Do SPSS Statistics 28 byla přidána podpora pro režim Vysokého kontrastu. Pokud zvolíte režim Vysoký kontrast a otevřete SPSS Statistics, bude se automaticky zobrazovat v tomto režimu. Chcete-li použít vysoký kontrast pro počítače s Mac OS, je nutné vybrat položku Systém Preferences > Accessibility a tažením posuvníku Kontrastu displeje upravíte kontrast.

## Přidáno ve verzi 27

- nové procedury: Power Analysis,
- změny v modulu Base: příprava dat a bootstrapové funkce,
- vylepšení procedur: Effect size, Weighted Kappa a MATRIX a kvantilová regrese.

### Změny v modulu Base

Od verze IBM SPSS Statistics 27 jsou do modulu Base zahrnuty moduly Data Preparation a Bootstrapping. Daná změna poskytuje všem zákazníkům Statistics 27 přístup k pokročilým funkcím z těchto dvou modulů.

#### Power analysis

Power analýza hraje klíčovou roli při plánování experimentů. SPSS Statistics 27 nově nabízí 11 procedur pro power analýzu, které Vám pomohou určit optimální rozsah výběru pro Vaši studii. Procedury umí spočítat nejen potřebný rozsah výběru pro požadovanou sílu, ale i sílu testu pro daný rozsah. K dispozici jsou i grafické možnosti, např. zobrazení síly jako funkce tzv. effect size nebo 3-dimenzionální grafy zobrazující sílu jako funkci effect size a rozsahu výběru.

#### Weighted Kappa

Nově zařazená statistická metoda Weighted Kappa se používá pro vyjádření shody mezi dvěma hodnotiteli, kteří hodnotí na ordinální škále, a umožňuje použít váhy pro různou míru neshody.

#### Effect size

Tzv. effect sizes zobrazují míru porušení nulové hypotézy. O výpočet effect size jsou nově obohaceny dialogy t-testů a ANOVA.

#### MATRIX a kvantilová regrese

IBM SPSS Statistics 27 přináší také vylepšení procedury MATRIX o nové funkce a rozšířila se nabídka procedury kvantilové regrese o možnost efektivněji zadávat mřížku kvantilů.

#### Další změny

Některé další úpravy jsou například:

- nová funkce Search v nabídce Edit,

- bublinový graf jako čtyřdimenzionální scatterplot,
- kompatibilita programu s vyšší verzí Pythonu 3.8.2.

## Přidáno ve verzi 26

- nové Bayesovské metody,
- vylepšená integrace s produkty třetích stran (Microsoft Graphic Object),
- možnost úpravy grafů již při jejich tvorbě,
- přidání klávesových zkratk do syntax editoru.

### Bayesovská statistika

Bayesovská statistika je založena na odlišném principu než klasická (frekvencistická) statistika a může sloužit jako alternativa, nebo další možnost k analýze našich dat. Některé procedury využívající bayesovský přístup v IBM SPSS Statistics jmenovitě jsou:

- **T-test (jednovýběrový, párový, dvouvýběrový),**
- **Pearsonův korelační koeficient,**
- **lineární regrese,**
- **analýza rozptylu jednoduchého třídění.**

### Vylepšení grafiky

Díky nové možnosti exportu, Microsoft Graphic Object, se mnohem lépe manipuluje s grafy v programech MS Office. Tyto grafy se pak chovají jako by byly vytvořeny přímo v prostředí MS Office a jejich následná úprava je výrazně jednodušší.

Dalším vylepšením prošel tzv. Chart Builder, kde přibýly další možnosti editace grafu již při jeho tvorbě a celkově byla vylepšena vizualizace grafických výstupů. Jmenovitě můžeme vybírat barvy kategorií pro kategorizované proměnné, slučovat méně četné kategorie do jedné nebo upravovat samotnou velikost grafu.

### Syntax editor

Změny v syntax editoru jsou mířeny hlavně na uživatele, kteří s ním běžně pracují. Zápis syntaxe se dočkal úpravy ve formě klávesových zkratk, které zjednodušují psaní, formátování i editaci. Pomocí klávesových zkratk lze řádky rychleji spojovat, duplikovat, odstraňovat, přesunovat atp. Asi nejvýznamnější úpravou je možnost kopírování dat z datového editoru či MS Excelu do více řádků najednou. Tyto novinky jsou dostupné v modulu Base.



## Další změny

Některé další úpravy jsou například:

- předělání dialogového okna pro spojování souborů,
- zjednodušení panelu nástrojů,
- možnost otevírat a ukládat soubory Stata v14.

## Přidáno ve verzi 25

Bayesovská statistika je založena na odlišném principu než klasická (frekventistická) statistika. Může sloužit jako další možnost k analýze našich dat. Některé procedury využívající bayesovský přístup v IBM SPSS Statistics jmenovitě jsou:

- **T-test (jednovýběrový, párový, dvouvýběrový),**
- **Pearsonův korelační koeficient,**
- **lineární regrese,**
- **analýza rozptylu jednoduchého třídění.**

Díky nové možnosti exportu, Microsoft Graphic Object, se mnohem lépe pracuje s grafy v programech MS Office. Tyto grafy se pak chovají jako by byly vytvořeny přímo v prostředí MS Office a jejich následná úprava je výrazně jednodušší.

Dalším vylepšením prošel tzv. Chart Builder, kde přibyly další možnosti editace grafu již při jeho tvorbě a celkově byla vylepšena vizualizace grafických výstupů. Jmenovitě můžeme vybírat barvy kategorií pro kategorizované proměnné, slučovat méně četné kategorie do jedné nebo upravovat samotnou velikost grafu.

Změny v syntaxovém editoru jsou mířeny hlavně na uživatele, kteří s ním běžně pracují. Zápis syntaxe se dočkal úpravy ve formě klávesových zkratk, které zjednodušují psaní, formátování i editaci. Pomocí klávesových zkratk lze řádky rychleji spojovat, duplikovat, odstraňovat, přesunovat atp. Významnou úpravou je možnost kopírování dat z datového editoru či MS Excelu do více řádků najednou. Tyto novinky jsou dostupné v modulu Base.

**Některé další úpravy jsou například:**

- předělání dialogového okna pro spojování souborů,
- zjednodušení panelu nástrojů,
- možnost otevírat a ukládat soubory Stata v14.

## Přidáno ve verzi 24

- **rozšiřující doplňky z webu,**
- **snadnější tvorba a sdílení vlastních doplňků,**
- **nový průvodce načítáním dat,**
- **vylepšení v modulu IBM SPSS Custom Tables.**

V IBM SPSS Statistics 24 naleznete nový způsob práce se zdrojem programových doplňků od třetích stran na **Statistics Extension Hub** a nástroj **Custom Dialog Builder for Extensions** pro snadnou práci s doplňky.

The Extensions Hub je nové prostředí pro správu a získání doplňků, kde najdete, stáhnete, aktualizujete, případně odinstalujete doplněk založený na syntaxi IBM SPSS, která může přistupovat k programu Python nebo R. Prostřednictvím IBM SPSS Statistics Custom Dialog Builder for Extensions je správa, vytváření a sdílení doplňků založených na R a Python velmi jednoduchá - snadno a rychle vytvoříte doplněk v grafickém rozhraní pro tvorbu dialogových oken, doplněk můžete sdílet, instalovat a umístit do nabídek programu.

Způsob **načítání a zápisu dat** z externích zdrojů ve verzi 24 obsahuje lepší rozpoznávání typu proměnných, vynechání skrytých sloupců a řádků v MS Excel, odstranění mezer ze začátků a konců textových řetězců, export popisků hodnot do databáze a mnoho dalšího.

**Modul IBM SPSS Custom Tables** obsahuje několik vylepšení. Součástí je např. designová váha (effective weights) pro výběry s nestejnou pravděpodobností, zobrazení intervalů spolehlivosti a směrodatných odchylek statistik, zobrazení výsledků signifikantních testů přímo v těle tabulky a zobrazení signifikancí pro testy sloupcových průměrů a porovrcí.

## Přidáno v IBM SPSS Statistics 23

### Nové algoritmy

- modelování vztahů mezi velkým počtem časových řad pomocí časového kauzálního modelování (Temporal Casual Modeling)

### Geoprostorové analýzy

- modely vztahů napříč časem a prostorem pomocí časoprostorové předpovědi (Spatial Temporal Prediction)
- zobecněná prostorová asociační pravidla zahrnující geoprostorové atributy pro předpověď výskytu událostí (Generalized Spatial Association Rule)

## Zvýšená efektivita

- vylepšená kategorická analýza hlavních komponent obsahující neparametrický bootstrapping a nové rotace pro vylepšení konvergence
- vylepšené možnosti vytváření vlastních skriptů pomocí jazyka R ve vývojovém prostředí zabudovaném v SPSS pro snadnější vytváření a ladění skriptů
- nová generace webových výstupů s vyšší interaktivitou, lepší funkcí a podporou webových serverů
- načítání a ukládání souborů Stata 9-13
- lepší možnosti importovat, číst a zapisovat soubory Stata 9-13
- rychlejší zápis do databáze na základě dávkového nahrávání (bulk export/insert)
- podpora přístupu k softwaru pomocí identifikačních karet

## Přidáno v IBM SPSS Statistics 22

- interaktivní zobrazení výstupu na chytrých zařízeních
- příprava výstupů v prezentační kvalitě
- efektivnější modelování pomocí Monte Carlo simulací, podpora automatického lineárního modelování (Automatic Linear Modeling), simulace textových proměnných a teplotních grafů
- rychlejší nalezení informací na úvodní stránce
- zvýšení výkonu a škálovatelnosti pomocí IBM SPSS Server a SQL pushback
- instalace Python plugin v rámci hlavní instalace SPSS Statistics
- vyhledávání, stahování a instalace dostupných skriptů přímo z SPSS Statistics., možnost vyhledávání aktualizace již nainstalovaných skriptů a stahování a instalování jejich aktualizované verze
- zjednodušená metoda specifikace uživatelských odhadů v SPSS Amos
- zdokonalená podpora přihlašování pro podnikové normy (Enterprise Standard) v normách platformy (Platform Standards)
- možnost číst/zapisovat šifrované datové soubory SPSS Statistics prostřednictvím jiných aplikací pomocí vstupně-výstupní DLL knihovny
- výstup neparametrických testů formou standardních pivotních tabulek

## Přidáno v IBM SPSS Statistics 21

- Monte Carlo simulace pro vytváření lepších modelů pro nejisté vstupy
- porovnání dvou datových souborů nebo datových matic pro zajištění kompatibility hodnot a informací o proměnných
- zabezpečení citlivých datových a výstupních souborů před prohlížením nekompetentními osobami pomocí šifrování a hesla
- vylepšené a rychlejší spojování dat
- zdokonalení pivotních tabulek novými funkcemi jako je např. snadná navigace a třídění
- jednodušší specifikace modelu v IBM SPSS Amos, nástroj pro strukturální modelování vztahů (SEM)
- programování v SPSS Statistics pomocí Java pluginu
- možnost importu dat z IBM® Cognos® Business Intelligence do SPSS Statistics pro podrobnější analýzy
- export výstupů do Microsoft Excel 2007/2010
- lepší škálovatelnost a vyšší výkon pomocí „load balancing“ v IBM SPSS Collaboration and Deployment Services
- systém jednotného přihlášení (single sign-on) mezi klientem SPSS Statistics a serverem
- pro prostředí s omezeným zabezpečením: možnost spouštět SPSS Statistics Server jako „non-root“ uživatel na operačních systémech UNIX/LINUX
- komprese datových souborů pro uchování místa na disku
- kompatibilita novějších verzí klienta se staršími verzemi serveru a naopak (např. klient verze 21 a server verze 20 nebo klient verze 20 a server verze 21)

## Přidáno v IBM SPSS Statistics 20

- mapové šablony a podpora ESRI souborů v IBM SPSS Statistics Base
- rychlejší tvorba a práce s pivotními tabulkami
- GLMM procedura v IBM SPSS Advanced Statistics i pro ordinální závislé proměnné
- specifikace modelů v IBM SPSS Amos formou jednoduchého programového kódu
- spouštění dávek IBM SPSS Statistics Server v offline modu bez nutnosti připojení klienta na server
- komprese dočasných souborů vznikajících při třídění pomocí IBM SPSS Statistics Server pro ušetření místa na disku zejména při třídění velkých souborů

## Přidáno v IBM SPSS Statistics 19

- nová procedura Automatic Linear Models (ALM) poskytuje uživatelům, kteří nejsou odborníky přímo na statistiku, nástroj pro automatickou tvorbu lineárních modelů a predikcí číselných závislých proměnných
- nová procedura Generalized Linear Mixed Models (GLMM), součást modulu IBM SPSS Advanced Statistics, umožňuje tvorbu přesnějších modelů pro nelineární vztahy založené na hierarchických datech nebo na opakovaných měřeních
- několik novinek v IBM® SPSS® Direct Marketing
- rychleji se zobrazující tabulky v IBM® SPSS® Statistics Base
- více než tucet vylepšení editoru syntaxe pro snadnější ovládání (dostupné ve všech produktech IBM SPSS Statistics)
- nový „Statistics portal“ poskytující interaktivní online přístup k analýzám jak interním, tak i externím uživatelům (vyžaduje IBM SPSS Statistics Server a IBM SPSS Collaboration and Deployment Services)
- kompilace transformací v IBM SPSS Statistics Server pro urychlení výpočtů při velkých objemech datových manipulací
- analytici používající IBM SPSS Statistics Base mohou skórovat nová data, využívat předem vytvořené modely a přímo přistupovat k datům na Salesforce.com
- pivotace sloupců a kontingenčních tabulek v IBM SPSS Statistics Base a IBM® SPSS® Custom Tables
- zpracování malých datových souborů nebo řídce zastoupených jevů na OS Linux a Mac pomocí IBM® SPSS® Exact Tests
- podpora IBM SPSS Statistics Base na systému IBM System z (vyžaduje SuSE Linux)
- přehlednější zobrazení velkých pivotních tabulek
- rychlejší výpočet procedur Frequencies, Descriptives a Crosstabs při použití IBM SPSS Statistics Base Server
- podpora 64-bit hardwaru pro Windows a Mac
- podpora operačního systému Sierra® na Mac OS X 10.12