UNIVERZITA OBRANY

Fakulta ekonomiky a managementu

Statistické zpracování dat v aplikaci STAT1

Výuková pomůcka pro předmět Statistika

Jiří Neubauer, Marek Sedlačík

3. 11. 2012

Obsah

Popis STAT1

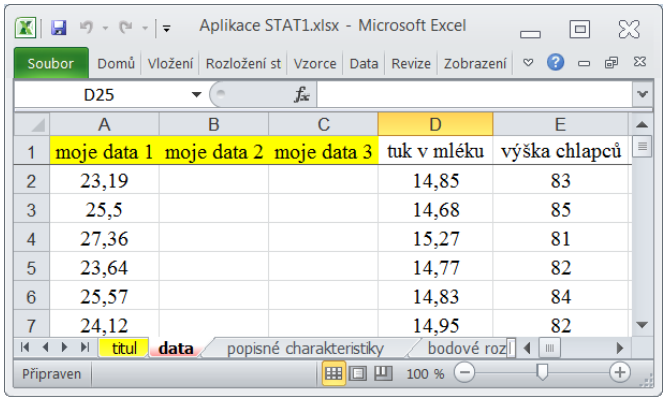
Aplikace STAT1 pracuje pod Microsoft Office Excel a je určena pro základní zpracování dat prostřednictvím exploratorní analýzy dat, metod jednorozměrné induktivní statistiky, dále jsou zde implementovány dvouvýběrové testy a chí-kvadrát test nezávislosti v kontingenční tabulce. Aplikace poskytuje řadu užitečných výstupů v podobě tabulek, grafů a statistických závěrů.

Uživatel může pomocí tohoto nástroje zpracovávat vlastní datové soubory, případně lze využít již vložených dat. Ovládání se provádí pomocí nabízených menu nebo pomocí parametrů, které jsou označeny červeně. Oporu lze najít rovněž v knize Neubauer, J., Sedlačík, M. a O. Kříž. Základy statistiky: Aplikace v technických a ekonomických oborech. Praha: Grada, 2012. ISBN 978-80-247-4273-1.

Použití

Vložení dat

Pro vložení vlastního datového souboru přejděte na list „data“. Do prvních třech sloupců označených „moje data 1“, „moje data 2“ a „moje data 3“ vložte data.

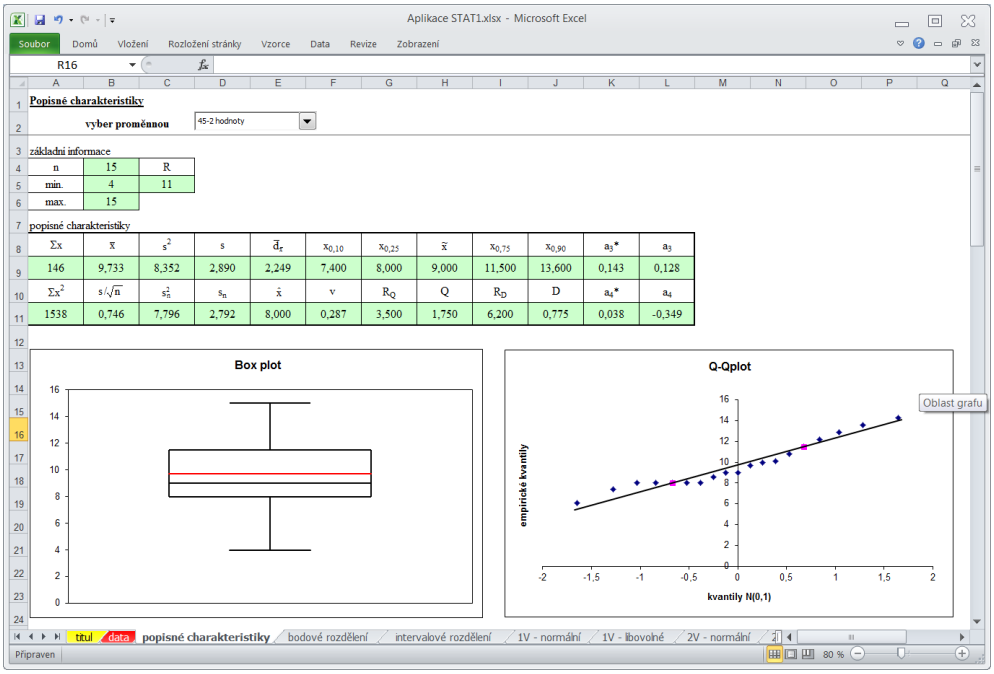


	A	B	C	D	E
1	moje data 1	moje data 2	moje data 3	tuk v mléku	výška chlapců
2	23,19			14,85	83
3	25,5			14,68	85
4	27,36			15,27	81
5	23,64			14,77	82
6	25,57			14,83	84
7	24,12			14,95	82

Data obsažená v knize Základy statistiky jsou uvedena v daném listu v pořadí, v jakém se v knize objevují.

Popisné charakteristiky

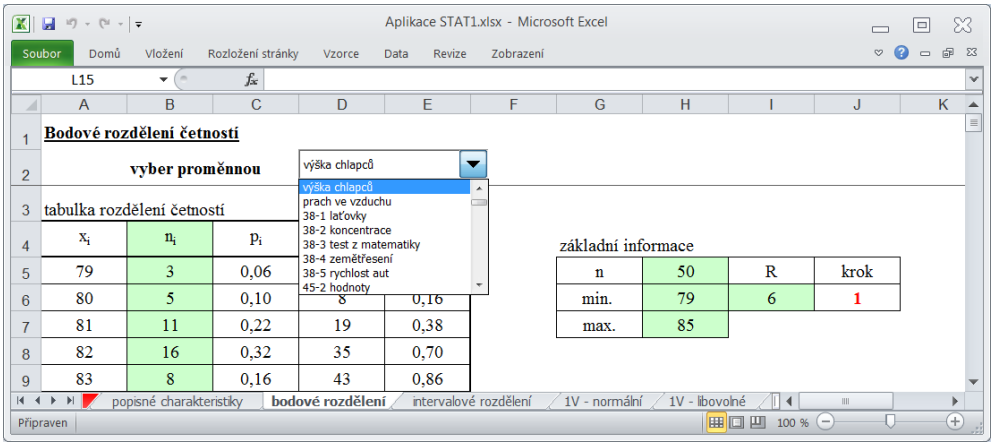
List „popisné charakteristiky“ nabízí výpočet vybraných číselných charakteristik datového souboru. Z nabízeného menu vyberte datový soubor, který máte v úmyslu zpracovávat. (Název datového souboru odpovídá názvu uvedenému v prvním řádku v listu „data“). Číselné charakteristiky v tabulkové podobě se spočítají automaticky.



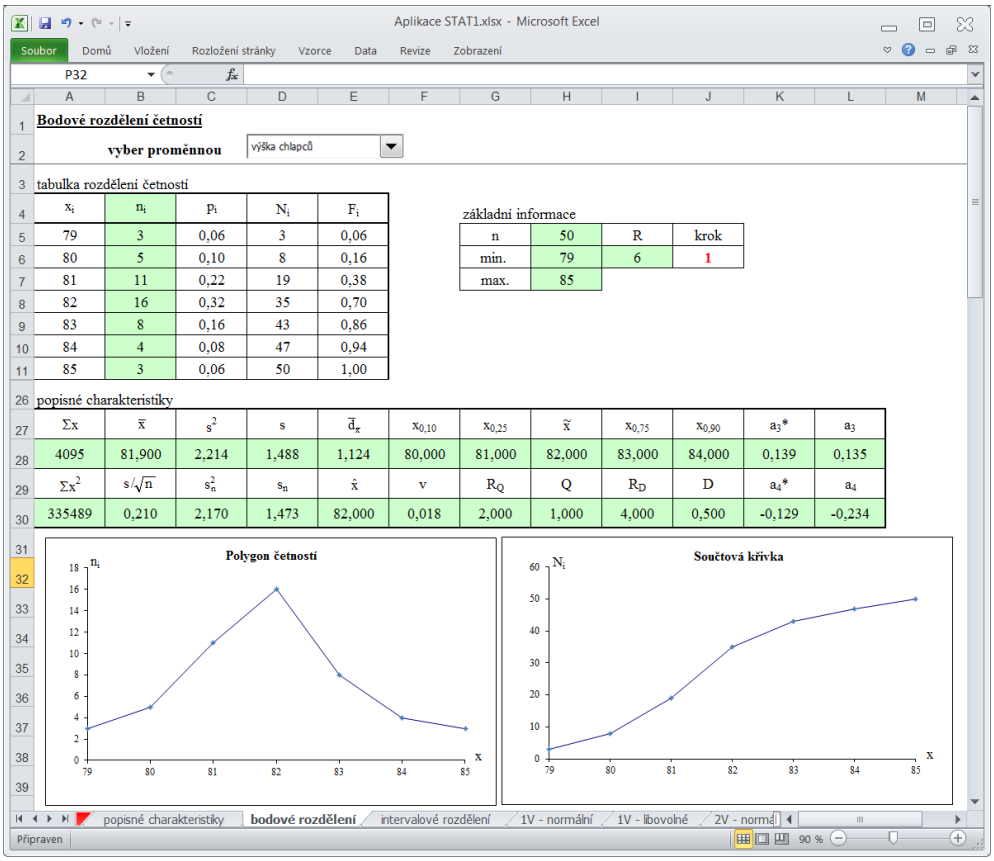
Kromě těchto charakteristik lze na listu nalézt dva grafy: krabicový diagram (boxplot) a Q-Q plot. Krabicový diagram zachycuje minimální a maximální hodnotu datového souboru, dolní kvartil, medián, horní kvartil a aritmetický průměr (červená linka). Q-Q plot porovnává teoretické kvantily normovaného rozdělení N(0,1) s empirickými kvantily určených z dat. Dále jsou spočteny testy normality založené na koeficientech šikmosti a špičatosti – viz testy hypotéz.

Bodové rozdělení četností

Pro vytvoření tabulky bodového rozdělení četností a grafů popisující toto rozdělení přejděte na list „bodové rozdělení“. Z nabízeného menu vyberte datový soubor, který máte v úmyslu zpracovávat. (Název datového souboru odpovídá názvu uvedenému v prvním řádku v listu „data“).



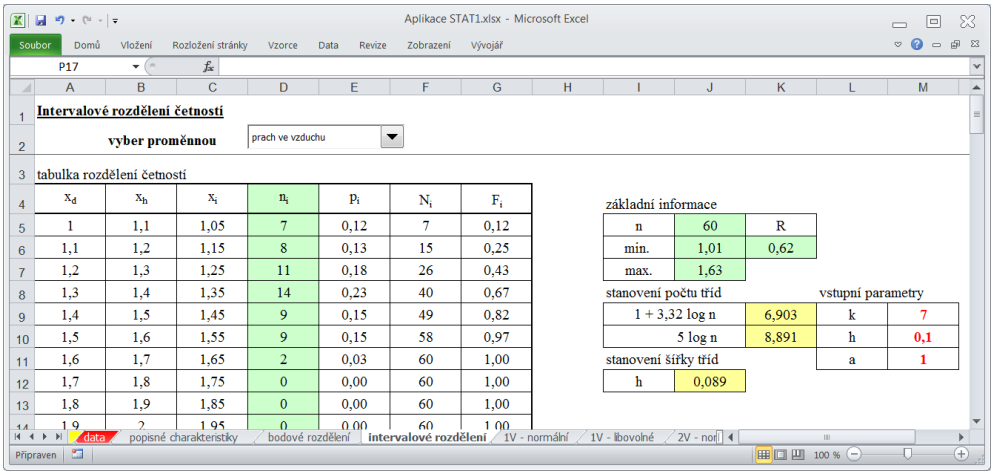
Tabulka rozdělení četností se vytvoří automaticky s krokem uvedeným v políčku „krok“, nastavenou hodnotu „1“ lze měnit v závislosti na vlastnostech datového souboru. Spolu s tabulkou se vypočítají základní popisné charakteristiky a vytvoří se dva grafy: polygon četností a součtová křivka. Řádky s nulovými četnostmi je možné skrýt (tyto hodnoty se nebudou objevovat v grafech).



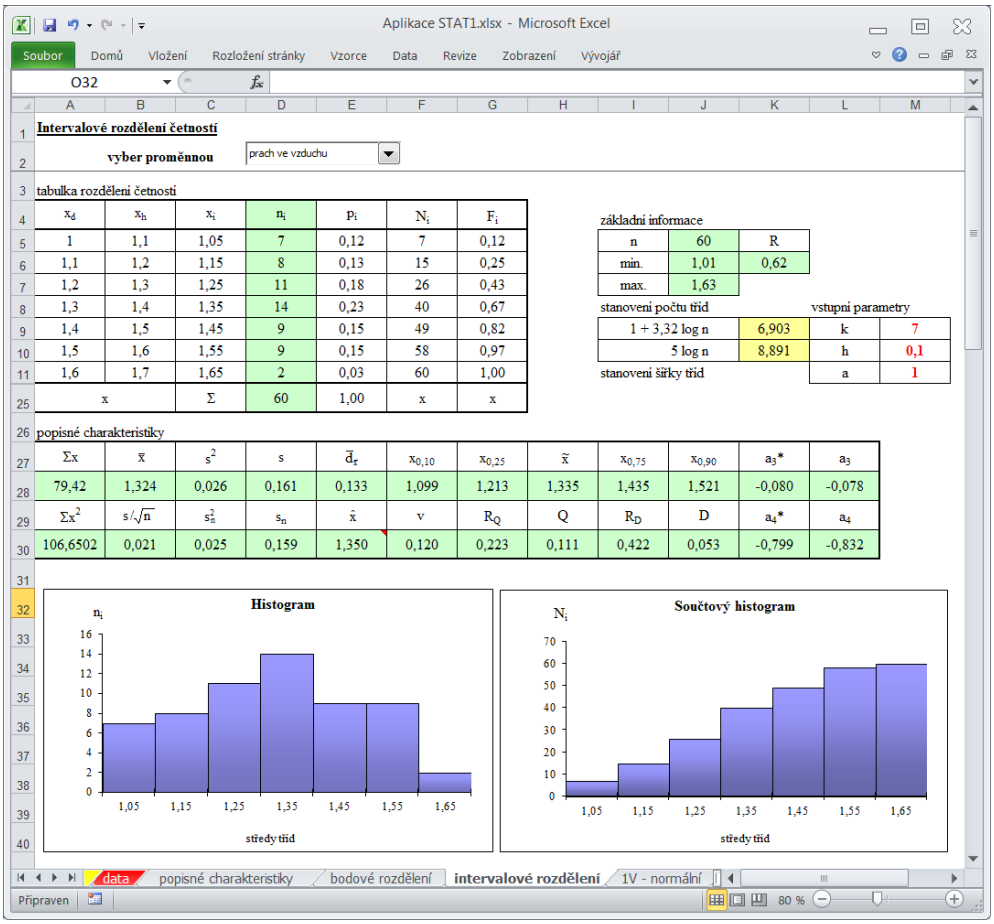
Dále jsou spočteny testy normality založené na koeficientech šikmosti a špičatosti – viz testy hypotéz.

Intervalové rozdělení četností

Pro vytvoření tabulky intervalového rozdělení četností a grafů popisující toto rozdělení přejděte na list „intervalové rozdělení“. Z nabízeného menu vyberte datový soubor, který máte v úmyslu zpracovávat. (Název datového souboru odpovídá názvu uvedenému v prvním řádku v listu „data“). Ke správnému vytvoření tabulky rozdělení četností je potřeba zadat následující tři parametry (označené červeně): k ... plánovaný počet tříd (řádků) v tabulce, h ... šířka třídy (intervalu), a ... počáteční hodnota, od které se začne tabulka vytvářet. Jako pomůcka pro určení optimálního počtu tříd jsou zde uvedena dvě pravidla. Konkrétní volba potom závisí na zpracovateli.



Spolu s tabulkou se vypočítají základní popisné charakteristiky a vytvoří se dva grafy: histogram a součtový histogram. Řádky s nulovými četnostmi je možné skrýt (tyto hodnoty se nebudou objevovat v grafech).

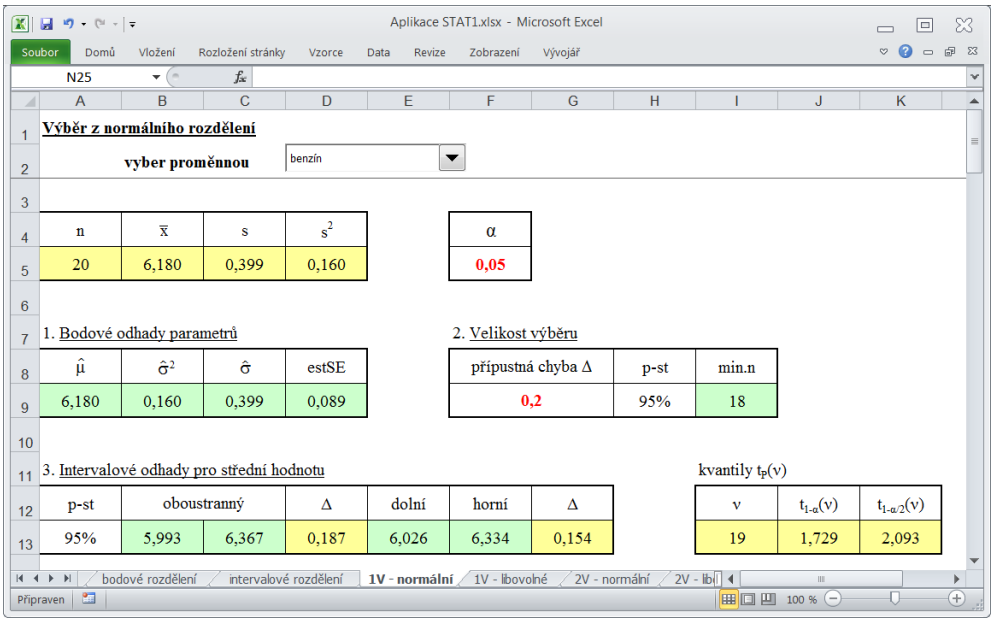


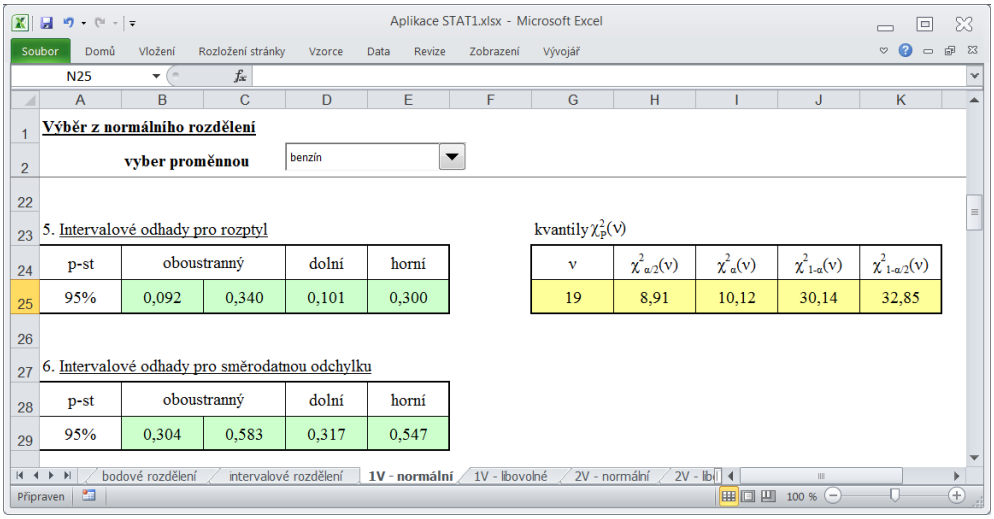
Dále jsou spočteny testy normality založené na koeficientech šikmosti a špičatosti – viz testy hypotéz.

Bodové a intervalové odhady

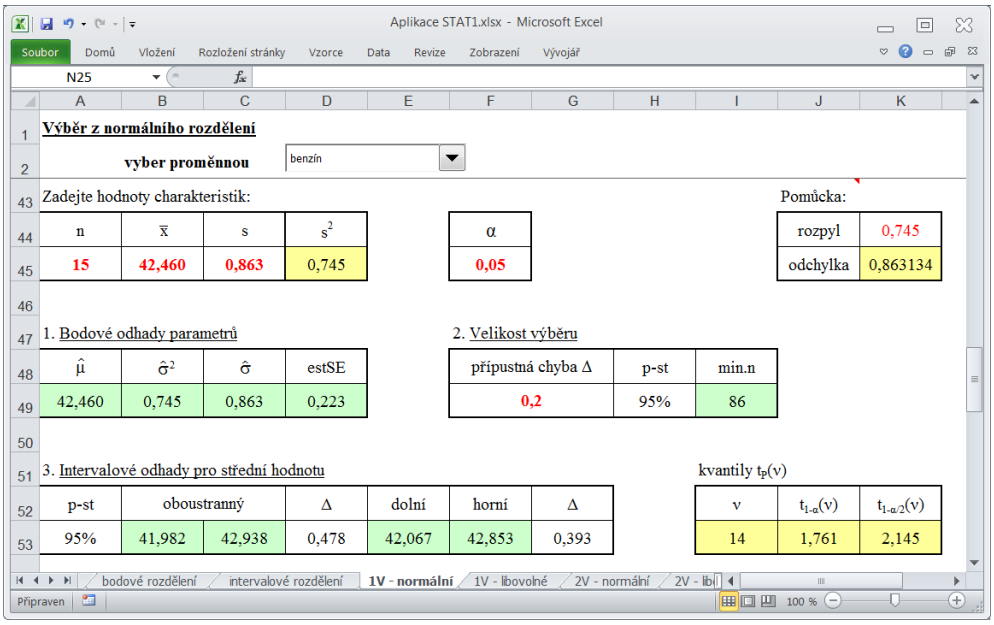
Odhady parametrů normálního rozdělení

Bodové a intervalové odhady střední hodnoty a rozptylu (příp. směrodatné odchylky) získáme přepnutím na list „1V – normální“. Poté, co vybereme analyzovaný datový soubor a zadáme riziko odhadu α (implicitně nastaveno na hodnotu 0,05), bodové odhady a intervaly spolehlivosti pro střední hodnotu, rozptyl i směrodatnou odchylku (oboustranný, dolní i horní) se vypočítají.





Odhady parametrů lze také získat přímým zadáním číselných charakteristik (rozsahu, aritmetického průměru a výběrové směrodatné odchylky).

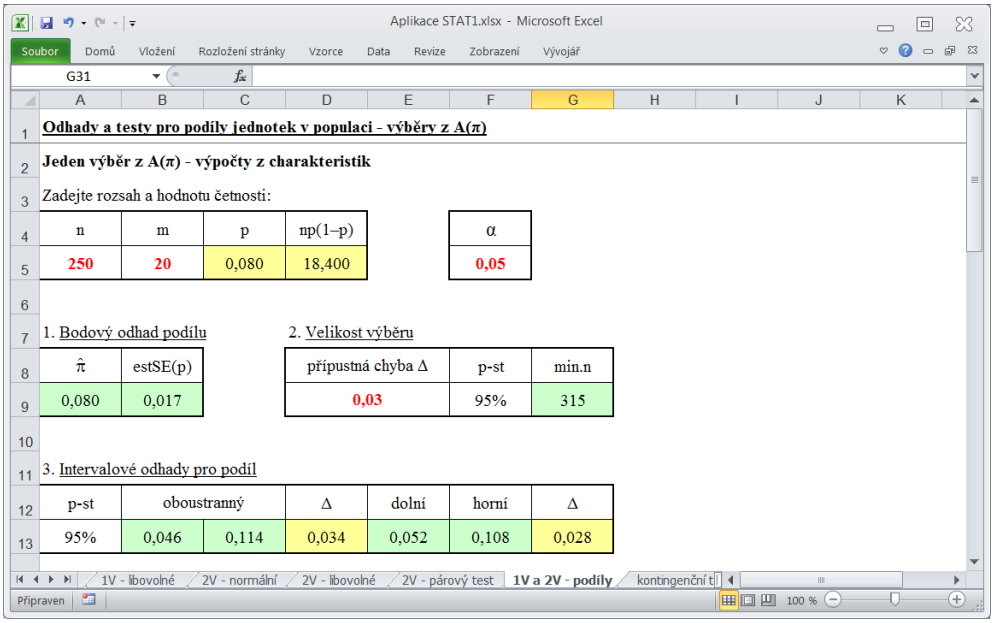


Odhady střední hodnoty pro výběry velkého rozsahu

Bodové a intervalové odhady střední hodnoty získáme přepnutím na list „1V – libovolné“. Ovládání je obdobné jako u odhadů parametrů normálního rozdělení.

Odhady parametru alternativního rozdělení

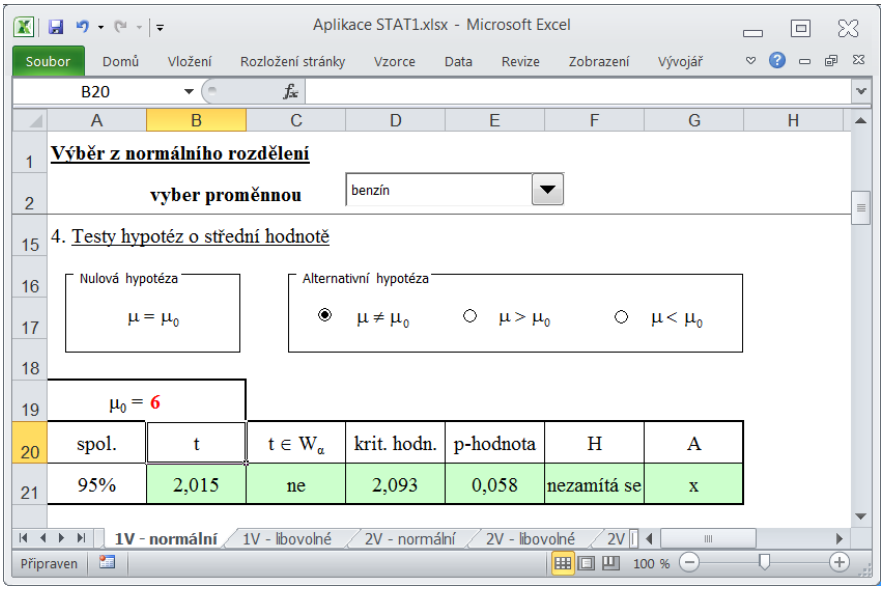
Bodové a intervalové odhady parametru alternativního rozdělení získáme přepnutím na list „1V a 2V – podíly“. Zde je nutné zadat vstupní n a m, kde podíl m/n je bodovým odhadem parametru π alternativního rozdělení.



Testy statistických hypotéz

Jednovýběrové testy

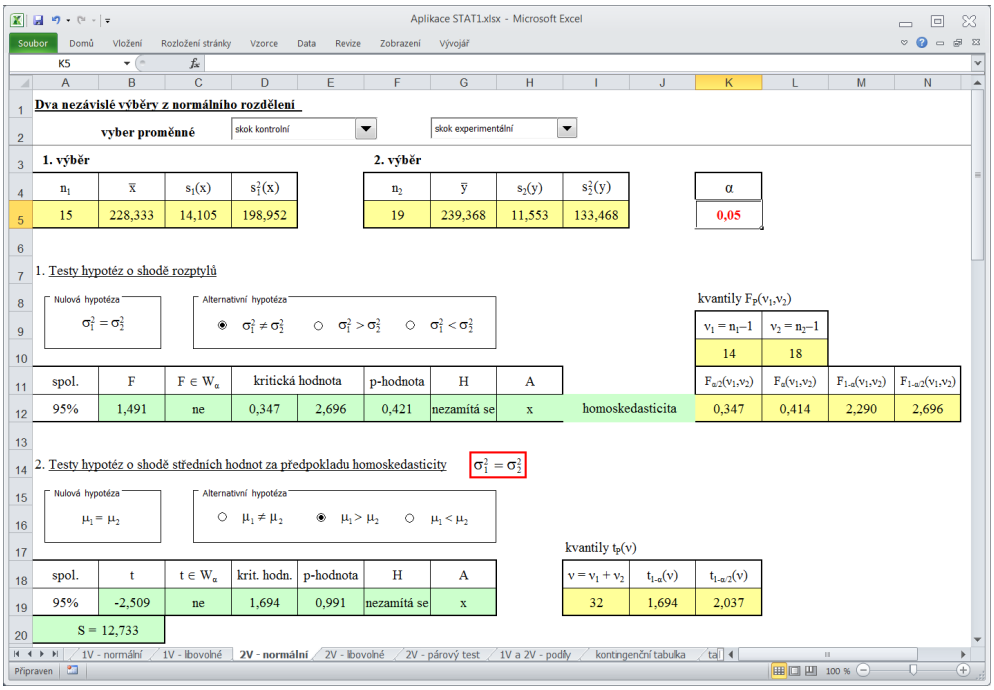
Aplikace STAT1 obsahuje tyto jednovýběrové testy hypotéz: test střední hodnoty a rozptylu normálního rozdělení (list „1V – normální“), test střední hodnoty pro velké výběry (list „1V – libovolné“) a test parametru alternativního rozdělení pro velké výběry (list 1V a 2V – podíly). Testování se ve všech případech provádí podobně, zaměříme se na jeden konkrétní – test střední hodnoty normálního rozdělení. Přejdeme na list „1V – normální“ a vybereme datový soubor. Zvolíme hladinu významnosti α (implicitně nastaveno na hodnotu 0,05), zadáme nulovou hypotézu H a vybereme jednu ze tří nabízených alternativních hypotéz A.



Jako výstup obdržíme hodnotu testového kritéria, kritickou hodnotu, p-hodnotu a slovní odpověď (H se nezamítá, nebo H se zamítá A se přijímá). Testy je možné také počítat zadáním číselných charakteristik (v dolní části listu).

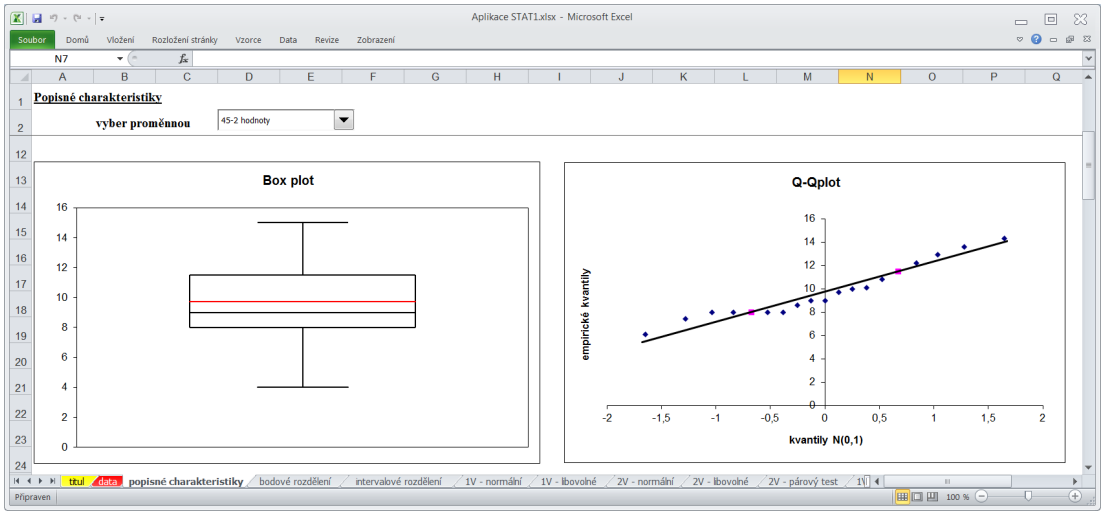
Dvouvýběrové testy

Aplikace STAT1 obsahuje tyto dvouvýběrové testy hypotéz: test shody dvou rozptylů nezávislých normálních rozdělení (list „2V – normální“), test shody dvou středních hodnot nezávislých normálních rozdělení (za předpokladu homoskedasticity a heteroskedasticity – list „2V – normální“), test shody dvou středních hodnot pro velké nezávislé výběry (list „2V – libovolné“), test shody dvou středních hodnot pro závislé výběry (párový test – list „2V – párový test“) a test shody dvou parametrů alternativního rozdělení pro velké nezávislé výběry (list 1V a 2V – podíly). Testování se provádí podobně jako u jednovýběrových testů, zde je třeba vybrat dva datové soubory.



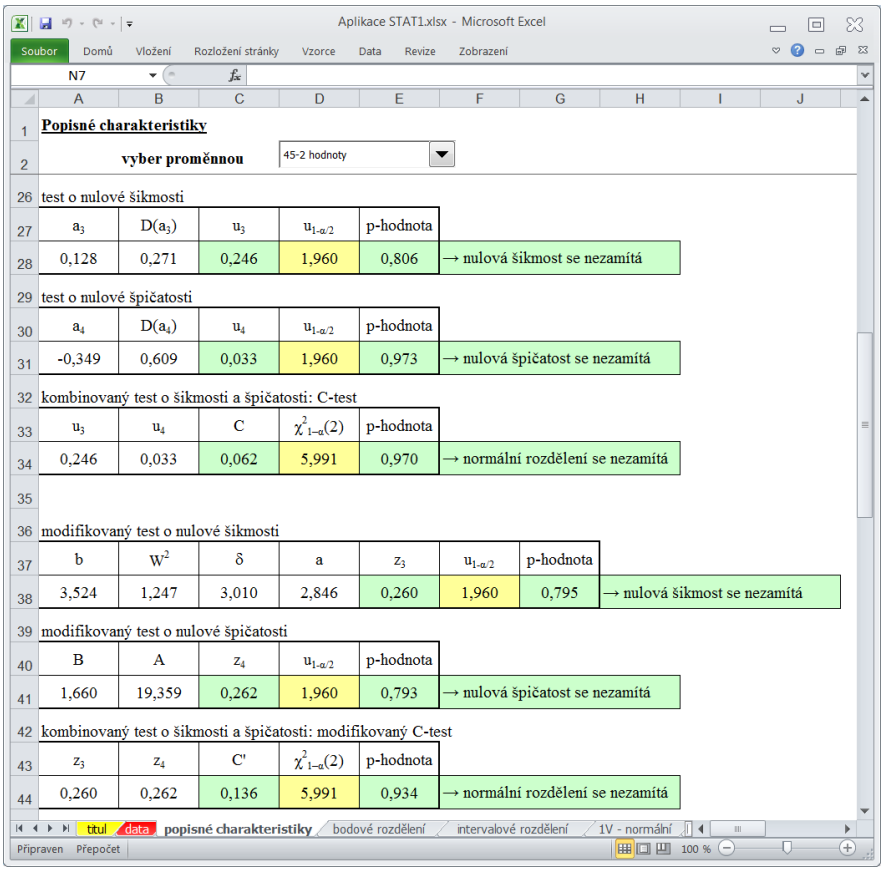
Testy normality

Základní představu o tvaru rozdělení datového souboru můžeme získat konstrukcí histogramu, případně polygonu četností (viz intervalové a bodové rozdělení četností). V listu „popisné charakteristiky“ lze nalézt kromě krabicového diagramu i Q-Q plot porovnávající teoretické kvantily normovaného rozdělení N(0,1) s empirickými kvantily určených z dat. Leží-li tyto body přibližně na přímce, můžeme usoudit, že zkoumaný náhodný výběr pochází z normálního rozdělení.



Listy „popisné charakteristiky“, „bodové rozdělení“ a „intervalové rozdělení“ obsahují v dolní části testy normality založené na výběrových koeficientech šik-

mosti a špičatosti.



Aplikace STAT1.xlsx - Microsoft Excel

Soubor | Domů | Vložení | Rozložení stránky | Vzorce | Data | Revize | Zobrazení

N7

Popisné charakteristiky

vyber proměnnou 45-2 hodnoty

test o nulové šikmosti

a_3	$D(a_3)$	u_3	$u_{1-\alpha/2}$	p-hodnota	
0,128	0,271	0,246	1,960	0,806	→ nulová šikmost se nezamítá

test o nulové špičatosti

a_4	$D(a_4)$	u_4	$u_{1-\alpha/2}$	p-hodnota	
-0,349	0,609	0,033	1,960	0,973	→ nulová špičatost se nezamítá

kombinovaný test o šikmosti a špičatosti: C-test

u_3	u_4	C	$\chi^2_{1-\alpha}(2)$	p-hodnota	
0,246	0,033	0,062	5,991	0,970	→ normální rozdělení se nezamítá

modifikovaný test o nulové šikmosti

b	W^2	δ	a	z_3	$u_{1-\alpha/2}$	p-hodnota	
3,524	1,247	3,010	2,846	0,260	1,960	0,795	→ nulová šikmost se nezamítá

modifikovaný test o nulové špičatosti

B	A	z_4	$u_{1-\alpha/2}$	p-hodnota	
1,660	19,359	0,262	1,960	0,793	→ nulová špičatost se nezamítá

kombinovaný test o šikmosti a špičatosti: modifikovaný C-test

z_3	z_4	C'	$\chi^2_{1-\alpha}(2)$	p-hodnota	
0,260	0,262	0,136	5,991	0,934	→ normální rozdělení se nezamítá

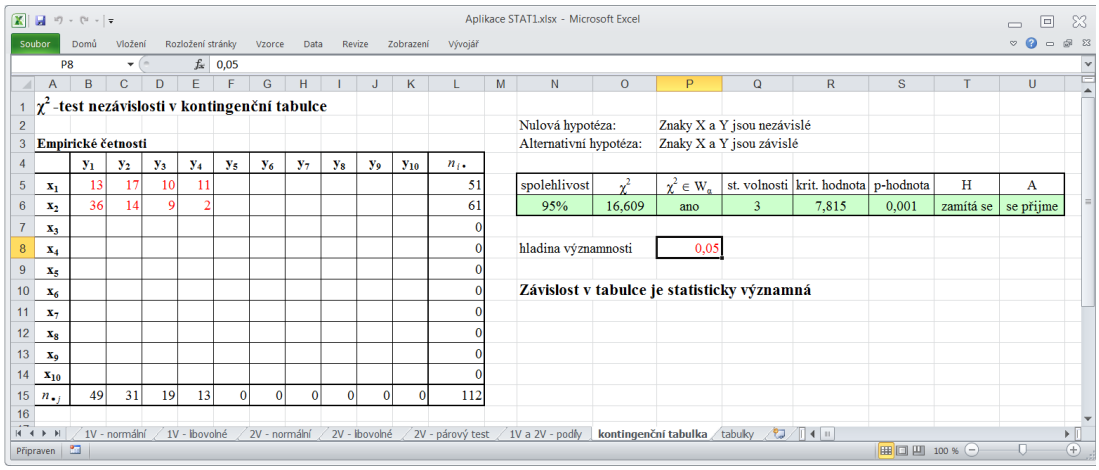
titul | data | popisné charakteristiky | bodové rozdělení | intervalové rozdělení | 1V - normální

Připraven | Přepočet | 100 %

Výpočet těchto testů zadáním potřebných charakteristik (rozsah, koeficient šikmosti a špičatosti) lze provést v dolní části listu „popisné charakteristiky“.

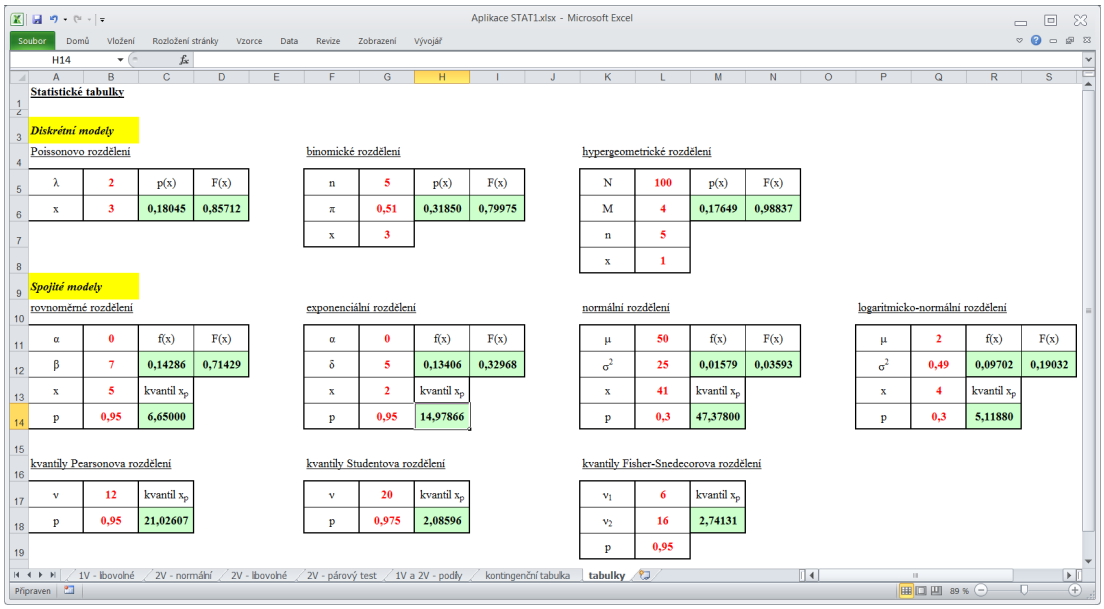
Chí-kvadrát test nezávislosti v kontingenční tabulce

List „kontingenční tabulka“ je určen pro testování nezávislosti v kontingenční tabulce užitím tzv. chí-kvadrát testu nezávislosti dvou statistických znaků. Tento test patří mezi neparametrické metody, to znamená, že nevyžaduje znalost rozdělení zkoumaných statistických proměnných. Při chí-kvadrát testu nezávislosti testujeme nulovou hypotézu H, že sledované znaky jsou nezávislé, proti alternativní hypotéze A, která je naopak hypotézou o jejich závislosti. Uživatel doplní poze hodnoty absolutních četností do připravené kontingenční tabulky a zvolí hladinu významnosti testu α (implicitně nastaveno na hodnotu 0,05).



Statistické tabulky

Poslední list „tabulky“ obsahuje hodnoty pravděpodobnostních a distribučních funkcí Poissonova, binomického a hypergeometrického rozdělení dále funkce hustoty pravděpodobnosti, distribuční funkce a kvantily rozdělení rovnoměrného, exponenciálního, normálního a log-normálního (u verze pro MS Office 2003 a 2007 - STAT.xls - chybí distribuční funkce hypergeometrického rozdělení a funkce hustoty pravděpodobnosti log-normálního rozdělení, která nejsou dispozici). Jsou zde uvedeny i kvantily Pearsonova, Studentova a Fisher-Snedecorova rozdělení.



Použité zdroje

- Anděl, J. *Základy matematické statistiky.* 1. vyd. Praha: Matfyzpress, 2005. ISBN 80-86732-40-1.
- Budíková, M., M. Králová a B. MAROŠ. *Průvodce základními statistickými metodami.* 1. vyd. Praha: Grada, 2010. ISBN 978-80-247-3243-5.
- Chajdiak, J. *Štatistika v Exceli 2007.* 1. vyd. Bratislava: Statis, 2009. ISBN 978-80-85659-49-8.
- Neubauer, J., Sedlačík, M. a O. Kříž *Základy statistiky: Aplikace v technických a ekonomických oborech.* Praha: Grada, 2012. ISBN 978-80-247-4273-1.
- Schels, I. *Excel 2007 – vzorce a funkce.* 1. vyd. Praha: Grada, 2008. ISBN 978-80-247-2074-6.