



Ekonomická
fakulta
Faculty
of Economics

Jihočeská univerzita
v Českých Budějovicích
University of South Bohemia
in České Budějovice

Jihočeská univerzita v Českých Budějovicích
Ekonomická fakulta
Katedra aplikované matematiky a informatiky

Bakalářská práce

Software STATISTICA a její využití při analýze kategoriálních dat

Vypracovala: Petra Horká

Vedoucí práce: Ing. Michael Rost, Ph.D.

České Budějovice 2016/2017

ZADÁNÍ BAKALÁŘSKÉ PRÁCE
(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Petra HORKÁ**
Osobní číslo: **E14116**
Studijní program: **B6208 Ekonomika a management**
Studijní obor: **Účetnictví a finanční řízení podniku**
Název tématu: **Software STATISTICA a její využití při analýze
kategoriálních dat**
Zadávající katedra: **Katedra aplikované matematiky a informatiky**

Z á s a d y p r o v y p r a c o v á n í :

Cíl práce spočívá v představení možností statistického balíčku STATISTICA pro analýzu kategoriálních dat plynoucích z dotazníkových šetření v oblasti daňové problematiky a účetnictví.

Metodický postup:

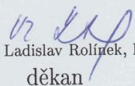
1. Studium literatury a manuálu software STATISTICA.
2. Zpracování teoretického základu, zejména představení software a vybraných statistických metod.
3. Hlavní část - možnosti využití softwaru a rozebrání předností a nedostatků při využívání software STATISTICA.
4. Praktická část - uvedení typových příkladů.
5. Závěr.

Rozsah grafických prací: 15 stran
Rozsah pracovní zprávy: 40 - 50 stran
Forma zpracování bakalářské práce: tištěná
Seznam odborné literatury:

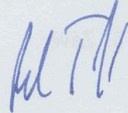
1. AGRESTI, A. (2002). *Categorical data analysis*. 2nd ed. Hoboken: Wiley-interscience. Wiley series in probability and statistics.
2. BORŮVKOVÁ, J., HORÁČKOVÁ P., & HANAČEK, M. (2013). *Statistica: úvod do zpracování dat*. Jihlava: Vysoká škola polytechnická Jihlava.
3. EVERITT, B. (1994). *The analysis of contingency tables. Monographs on statistics and applied probability*, 45. 2nd ed. London: Chapman & Hall.
4. LOUDA, Z. (2004.) *Řešené příklady v systému STATISTICA*. Praha: Česká zemědělská univerzita v Praze, Provozně ekonomická fakulta.
5. ŘEZANKOVÁ, H, & HRONOVA, S. (2000). *Statistická data*. Praha: Vysoká škola ekonomická.

Vedoucí bakalářské práce: Ing. Michael Rost, Ph.D.
Katedra aplikované matematiky a informatiky

Datum zadání bakalářské práce: 15. ledna 2016
Termín odevzdání bakalářské práce: 14. dubna 2017


doc. Ing. Ladislav Rolínek, Ph.D.
děkan

JIHOČESKÁ UNIVERZITA
V ČESKÝCH BUDĚJOVICÍCH
EKONOMICKÁ FAKULTA
Studentská 13 (25)
370 05 České Budějovice


prof. RNDr. Pavel Tlustý, CSc.
vedoucí katedry

V Českých Budějovicích dne 16. března 2016

PROHLÁŠENÍ

Prohlašuji, že svoji bakalářskou práci jsem vypracovala samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47 zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své bakalářské práce, a to - v nezkrácené podobě/v úpravě vzniklé vypuštěním vyznačených částí archivovaných Ekonomickou fakultou - elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

.....
Datum

.....
Podpis studenta

PODĚKOVÁNÍ

Dovolte mi poděkovat v první řadě vedoucímu bakalářské práce panu Ing. Michaelovi Rostovi, PhD. za ochotu a čas, který mi věnoval při zpracování a dokončení práce. Dále bych chtěla poděkovat svým přátelům a blízkým (především Janu Valentovi) za pevné nervy a podporu, díky níž se mi podařilo dojít do konce bakalářského studia.

Obsah

1	Úvod	3
2	Úvod do analýzy kategoriálních dat	4
3	Jednorozměrná analýza	5
3.1	Rozdělení četností.....	5
3.2	Testování hypotéz.....	5
3.2.1	Binomický (znaménkový) test	5
3.2.2	Chí-kvadrát test (test dobré shody)	8
3.2.3	Testy porovávající četnosti dvou kategorií	8
4	Dvourozměrná analýza	9
4.1	Kontingenční tabulky	9
4.2	Zjišťování závislosti mezi dvěma proměnnými	10
4.2.1	Testování závislosti nominálních proměnných	11
4.2.2	Testování závislosti ordinálních proměnných.....	12
5	Software STATISTICA	14
5.1	STATISTICA 13.1	14
5.2	Další statistické softwary.....	15
5.3	Metodika – ovládání softwaru STATISTICA	16
6	Praktická práce	23
6.1	První část – analýza dat ze společnosti GRAND s.r.o.....	24
6.1.1	Znaménkový test s binomickým rozdělením	24
6.1.2	Binomický test s normovaným normálním rozdělením	26
6.1.3	Chí-kvadrát test dobré shody	28
6.1.4	Test porovávající četnosti dvou kategorií	32
6.1.5	Analýza dvourozměrné kontingenční tabulky (nominální proměnné).....	33
6.2	Druhá část – analýza dat z dotazníkového šetření	36
6.2.1	Chí-kvadrát test dobré shody	36
6.2.2	Test porovávající četnosti dvou kategorií	38
6.2.3	Analýza dvourozměrné kontingenční tabulky (nominální proměnné).....	41
6.2.4	Analýza dvourozměrné kontingenční tabulky (ordinální proměnné)	46

7	Závěr a diskuze	52
I	Summary.....	54
II	Citovaná literatura	55
III	Seznam tabulek	1
IV	Seznam obrázků.....	3
V	Seznam příloh.....	4

1 Úvod

V dnešní době, která je dobou plnou nových technologií a světu vládne internet, se už žádné analýzy nezpracovávají bez využití výpočetní techniky. Počítačová analýza se stala nezbytnou součástí inventáře dovedností každého současného uživatele. Dějiny počítačové analýzy byly vytvářeny už nad obecnými úvahami při práci s daty. Nástup moderních technologií umožnil této vědě obrovský krok dopředu a v dnešní době téměř každý uživatel zvládne vytvářet analýzy, graficky tyto výsledky zpracovávat a následně vyhodnocovat.

Pro počítačovou vědu je klíčový rok 1936, kdy byl představen první mechanický model počítače a historie softwaru se datuje roku 1946. Software obecně zajišťuje chod počítače či nějakého stroje. Software je tedy autorské dílo, a tak jej uživatelé mohou využívat po zakoupení licence od autora.

Proto je důležité seznámit čtenáře s ovládním a využitím počítačových aplikací, které nabízí možnosti pro zpracování pozorovaných dat a vytváření počítačových analýz. Dnes je na trhu opravdu mnoho počítačových programů a každý uživatel si může zvolit software podle svých požadavků a možností.

Cíl práce

Cílem práce je čtenáři ukázat možnosti softwaru STATISTICA při analýze kategoriálních dat. Představení předností a nedostatků při využívání a následné srovnání s obdobnými softwary. V první části praktické práce jsou zpracována data poskytnutá vybranou společností a v druhé části praktické práce plynou data pro analýzu z dotazníkového šetření. Následně jsou tyto data zpracována v softwaru STATISTICA podle metodického postupu. Při nastínění na typových příkladech jsou stanoveny nulové a alternativní hypotézy, hodnoty testových statistik, hodnoty potřebné k dokončení analýzy, závěr a vyhodnocení.

Metody a postupy v této práci jsou zaměřeny výhradně na analýzu kategoriálních dat, a od čtenáře se očekává, že má základní znalosti statistiky.

2 Úvod do analýzy kategoriálních dat

Tímto termínem se rozumí analýza, při které se vychází z dat, která jsou slovní, nebo je také lze nazývat kvalitativní. Hodnoty znaku se obvykle zjišťují pomocí dotazování. Odpovědím jsou přiřazeny buď slovní, nebo číselné kódy, a tyto hodnoty se nazývají *kategorie*. Kategoriální znaky se dělí na:

Nominální znaky (proměnné) a jsou vždy slovní. Tyto data nemají své stanovené pořadí, určuje se podle četnosti opakování znaku. Například zda podnik vede účetnictví či daňovou evidenci.

Ordinální znaky (proměnné) jsou také slovní, ale mohou být i číselné. U těchto dat záleží vždy na přirozeném pořadí, které se musí respektovat. Například pokud se hodnotí zadluženost podniku podle stupně (žádná, mírná, vysoká). (Borůvková, Horáčková, & Hanáček, 2013)

Kvantitativní diskrétní proměnné (nabývají pouze celočíselných obměn).

Zvláštním typem znaku je **dichotomická proměnná**, která nabývá pouze dvou hodnot. Dělí se na:

- *symetrické* – kategorie jsou stejně důležité (plátce daně, neplátce daně),
- *asymetrické* – jedna z nich je více důležitá (podnik přežil nebo nepřežil).

(Řezanková & Hronová, Statistická data, 2000)

U těchto dat nemá cenu zjišťovat popisné statistiky jako je například aritmetický průměr, medián, směrodatnou odchylku atd. Pro tyto data má smysl zjišťovat z těchto popisných statistik pouze modus. Modus je hodnota, která se ve sledovaném datovém souboru vyskytuje nejčastěji, tzn. má největší relativní četnost. Někdy se může vyskytovat i více modů a značí se jako \hat{x} .

Ve zkoumaném programu sloužícím ke statistické analýze kategoriálních dat se bude autor zabývat následujícími statistickými testy a metodami.

3 Jednorozměrná analýza

Jednorozměrná analýza zkoumá každý znak jednotlivě. A pokud je třeba současně zkoumat dva nebo více znaků, říká se tomu **vícerozměrná analýza** (viz Kapitola 4).

3.1 Rozdělení četností

Tabulky četností lze využít k analýze kategoriálních dat, a poskytují základní přehled o zjištěných hodnotách, které patří k popisným statistikám.

Tabulka četností obsahuje:

- proměnnou X a její kategorie x_i ,
- absolutní četnost N_i ,
- relativní četnosti p_i a platí, že:

$$p_i = \frac{N_i}{N} \quad (1)$$

- kumulativní relativní četnost P_i (u kategoriálních dat nemá smysl počítat).

Pravděpodobnost zastoupení jednotlivých kategorií budeme označovat jako π_i . (Borůvková, Horáčková, & Hanáček, 2013)

Tabulka 1: Schéma tabulky četností

X	Absolutní četnost	Relativní četnost
$x_{[1]}$	n_1	p_1
...
$x_{[i]}$	n_i	p_i
...
$x_{[k]}$	n_k	p_k
Celkem	n	1

Zdroj: (Borůvková, Horáčková, & Hanáček, 2013)

3.2 Testování hypotéz

3.2.1 Binomický (znaménkový) test

Jedná se o neparametrický test, díky kterému lze zjistit, zda procentuální výskyt zkoumaného jevu odpovídá frekvenci, která je požadována. (Binomický test, nedatováno)

Nulová hypotéza je stanovena jako $H_0: \pi_k = \pi_{k,0}$, kde k je počet kategorií.

Testovaná kategorie se označí indexem 1 a všechny ostatní kategorie se sloučí do jedné a značí indexem 2. Nulovou hypotézu lze tedy přepsat jako $H_0: \pi_1 = \pi_{1,0}$. Využití znaménkového testu pomůže zjistit minimální hladinu významnosti $p - value$, která se porovná se zvolenou hladinou významnosti α (většinou 5 %). Jestli platí vztah, že $p - value \leq \alpha$, tak nulová hypotéza bude **zamítnuta**, a pokud je $p - value > \alpha$, nulovou hypotézu **zamítnout nelze**.

Nulová hypotéza se testuje proti:

- Levostranné alternativě $H_1: \pi_1 < \pi_{1,0}$ – pokud platí, že $n_1 < n_2$ a p_1 (relativní četnost testované kategorie) $< \pi_{1,0}$,
- Pravostranné alternativě $H_1: \pi_1 > \pi_{1,0}$ – pokud platí, že $n_1 < n_2$ a $p_1 > \pi_{1,0}$,
- Oboustranné alternativě $H_1: \pi_1 \neq \pi_{1,0}$. (Agresti, 2002)

Hodnota $p - value$ se zjišťuje jako hodnota distribuční funkce binomického rozdělení v bodě n_1 . Znaménkový test s využitím binomického rozdělení se používá pro $n_1 + n_2 \leq 25$.

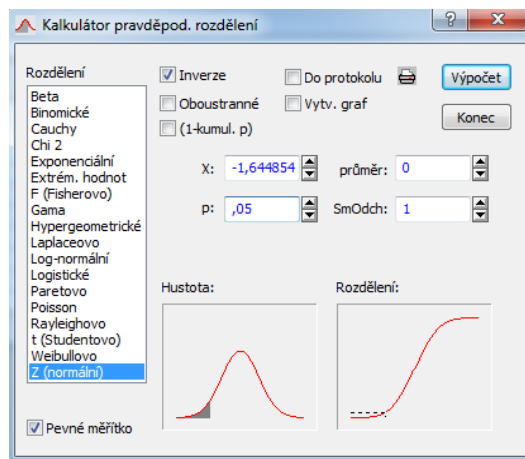
Pokud je $n_1 + n_2 > 25$, použije se test s normovaným normálním rozdělením. Pro toto rozdělení počítáme náhodnou veličinu Z .

$$Z = \frac{n_1 - n\pi + 0,5}{\sqrt{n\pi(1-\pi)}}. \quad (2)$$

Pro levostrannou hypotézu je hodnota $p - value$ distribuční funkce normovaného normálního rozdělení s kvantilem u_α , pro pravostrannou hypotézu je to kvantil $u_{1-\alpha}$ a v případě oboustranné hypotézy s kvantilem $u_{1-\alpha/2}$. (Řezanková, Analýza kategoriálních dat, 2005)

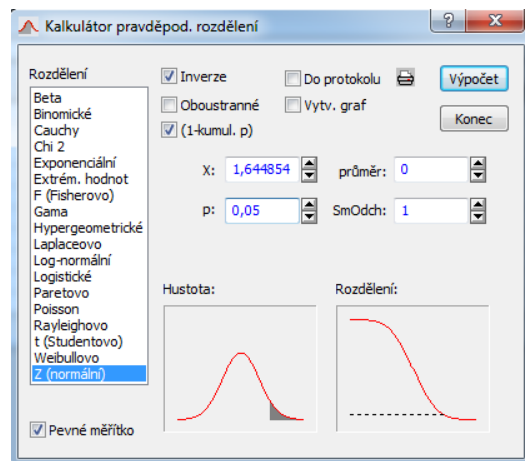
K výpočtu těchto kvantilů lze použít software STATISTICA. **Pravděpodobnostní kalkulátor** vypočítá požadovanou hodnotu kvantilu u . Následující obrázky znázorňují výpočet pro všechny tři hypotézy (oboustrannou, pravostrannou a levostrannou). Výstup z kalkulátoru nabízí i grafické znázornění funkce, sloužící pro snadnější určení alternativní hypotézy.

Obrázek 1: Distribuční funkce pro normální rozdělení - levostranná alternativa s kvantilem $u_{0,05}$



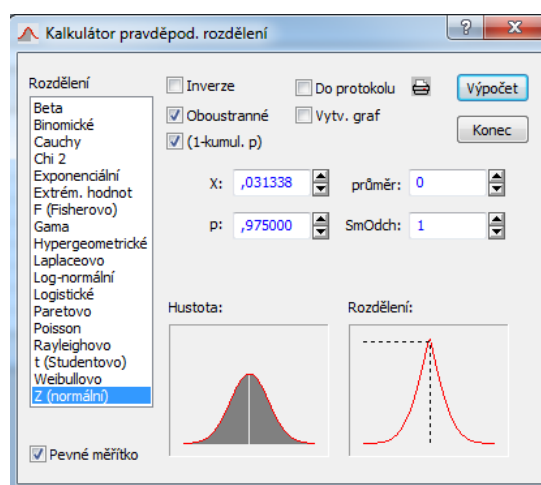
Zdroj: vlastní zpracování v software STATISTICA

Obrázek 2: Distribuční funkce pro normální rozdělení – pravostranná alternativa s kvantilem $u_{1-0,05}$



Zdroj: vlastní zpracování v software STATISTICA

Obrázek 3: Distribuční funkce pro normální rozdělení – pravostranná alternativa s kvantilem $u_{1-0,05/2}$



Zdroj: vlastní zpracování v software STATISTICA

3.2.2 Chí-kvadrát test (test dobré shody)

Jedná se o statistický test, který zachycuje rozdíl mezi pozorovanými a očekávanými četnostmi. Nulová hypotéza je stanovena jako $H_0: \pi_i = \pi_{i,0}$, kde $i = 1, 2, \dots, K$ a $\sum_{i=1}^k \pi_{i,0} = 1$ a testuje se proti alternativní hypotéze $H_1: \text{non } H_0$ (u chí-kvadrát testu je alternativní hypotéza buď pravostranná, nebo levostranná). Pro výpočet slouží testová statistika χ^2 a při vyhledávání kritické hodnoty má $k - 1$ stupně volnosti a za platnosti nulové hypotézy chí-kvadrát rozdělení.

Pro výpočet testové statistiky χ^2 lze použít vzorec:

$$\chi^2 = \frac{(n_i - \pi_{i,0})^2}{\pi_{i,0}}. \quad (3)$$

(Hendl, 2009)

3.2.3 Testy porovnávací četnosti dvou kategorií

Testuje se nulová hypotéza $H_0: \pi_i = \pi_j$ proti oboustranné hypotéze $H_1: \pi_i \neq \pi_j$. Jestliže je $n_i + n_j \geq 30$, lze použít testovou statistiku chí-kvadrát ve tvaru:

$$\chi^2 = \frac{(n_i - n_j)^2}{n_i + n_j}. \quad (4)$$

(Řezanková, Analýza kategoriálních dat, 2005)

V případě, že platí nulová hypotéza, má tato veličina chí-kvadrát rozdělení s jedním stupněm volnosti. Tato vypočtená hodnota se porovnává s kvantilem $\chi_{1-\alpha/2}^2$.

Jestliže není zamítnuta nulová hypotéza H_0 , lze odhadnout společnou hodnotu π bodovým odhadem:

$$p_{\text{odhad společné hodnoty}} = \frac{1}{2}(p_i + p_j). \quad (5)$$

(Řezanková, Analýza kategoriálních dat, 2005)

Pokud nulová hypotéza zamítnuta je, tak lze odhadnout rozdíl $\pi_i - \pi_j$ odhadem:

$$p_{\text{odhad rozdílu}} = p_i - p_j. \quad (6)$$

(Řezanková, Analýza kategoriálních dat, 2005)

4 Dvourozměrná analýza

Tato analýza vychází z dvourozměrné tabulky četností, která se nazývá kontingenční tabulka a zkoumá vzájemnou závislost proměnných.

4.1 Kontingenční tabulky

V kontingenční tabulce se stanoví rozsah souboru n , počet kategorií pro proměnnou X jako r a počet kategorií pro proměnnou Y jako s . Zjištěné četnosti jsou označeny jako n_{ij} , kde $i = 1, 2, \dots, r$ a $j = 1, 2, \dots, s$. (Everitt, 1994)

Tabulka 2: Schéma kontingenční tabulky pro absolutní četnosti

		Y					Celkem
		$y_{[1]}$...	$y_{[j]}$...	$y_{[s]}$	
X	$x_{[1]}$	n_{11}	...	n_{1j}	...	n_{1s}	n_{1+}

	$x_{[i]}$	n_{i1}	...	n_{ij}	...	n_{is}	n_{i+}

	$x_{[r]}$	n_{r1}	...	n_{rj}	...	n_{rs}	n_{r+}
Celkem		n_{+1}	...	n_{+j}	...	n_{+s}	n

Zdroj: (Řezanková, Analýza kategoriálních dat, 2005)

V tabulce 2 kromě sdružených absolutních četností znamenají symboly n_{+j} a n_{i+} marginální četnosti a platí pro ně, že: $n_{i+} = \sum_{j=1}^s n_{ij}$ a $n_{+j} = \sum_{i=1}^r n_{ij}$. (Řezanková, Analýza kategoriálních dat, 2005)

Tabulka 3: Schéma kontingenční tabulky pro relativní četnosti

		Y					Celkem
		$y_{[1]}$...	$y_{[j]}$...	$y_{[s]}$	
X	$x_{[1]}$	p_{11}	...	p_{1j}	...	p_{1s}	p_{1+}

	$x_{[i]}$	p_{i1}	...	p_{ij}	...	p_{is}	p_{i+}

	$x_{[r]}$	p_{r1}	...	p_{rj}	...	p_{rs}	p_{r+}
Celkem		p_{+1}	...	p_{+j}	...	p_{+s}	p

Zdroj: (Řezanková, Analýza kategoriálních dat, 2005)

Pokud se jedná o podíly na celkovém rozsahu souboru, pak je:

$$p_{ij} = \frac{n_{ij}}{n} \text{ a } \sum_{i=1}^r p_{i+} = \sum_{j=1}^s p_{+j} = 1.$$

Pro marginální relativní četnosti p_{+j} a p_{i+} platí, že: $p_{i+} = \sum_{j=1}^s p_{ij}$ a $p_{+j} = \sum_{i=1}^r p_{ij}$. (Everitt, 1994)

4.2 Zjišťování závislosti mezi dvěma proměnnými

(Řezanková, Analýza kategoriálních dat, 2005) ve své literatuře uvádí, že u dvou proměnných se nejčastěji zjišťuje, zda mezi proměnnými existuje závislost. Závislost může být buď vzájemná, nebo jednostranná. Pro zjišťování této závislosti se nejčastěji používá **chí-kvadrát test o vzájemné nezávislosti v kontingenční tabulce**. Pro sledování síly závislosti jsou využívány různé koeficienty, které nabývají hodnoty z intervalu $< 0; 1 >$, popřípadě $< -1; 1 >$, a pokud nabývají hodnoty 0, jsou proměnné nezávislé. Z toho tedy plyne, že dalšími analýzami jsou **testy o nulovosti těchto koeficientů**. Těchto koeficientů je velké množství a rozdělují se podle typu proměnných.

4.2.1 Testování závislosti nominálních proměnných

Testuje se nulová hypotéza $H_0: \pi_{ij} = \pi_{ij,0}$, kde $\pi_{ij,0}$ je očekávaná relativní četnost v případě nezávislosti, která je dána jako $\pi_{ij,0} = p_{i+}p_{+j}$. Alternativní hypotéza je stanovena jako $H_1: \pi_{ij} \neq \pi_{ij,0}$. Jedná se tedy o Pearsonův chí-kvadrát test o nezávislosti a testová statistika chí-kvadrát má tvar:

$$\chi_p^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n\pi_{ij,0})^2}{n\pi_{ij,0}}. \quad (7)$$

Za platnosti H_0 má tato veličina chí-kvadrát rozdělení s $(r - 1)(s - 1)$ stupně volnosti a testové kritérium je porovnáváno s kvantilem $\chi_{1-\alpha}^2[(r - 1)(s - 1)]$ (pro výpočet je možné opět použít *Pravděpodobnostní kalkulátor* v softwaru STATISTICA). Statistika nabývá hodnot z intervalu $< 0; n \cdot (q - 1) >$, kde q je $\min\{r, s\}$.

Chí-kvadrát statistiku lze použít pouze za předpokladu, že očekávané četnosti aspoň v 80 % polích nabývají hodnotu větší než 5 a ve zbylých polích minimálně hodnoty 1. (Řezanková, Analýza kategoriálních dat, 2005)

Koeficienty závislosti

Jedním ze základních koeficientů je **Pearsonův kontingenční koeficient C_P** , sloužící k měření závislosti dvou kategoriálních proměnných. Pearsonův koeficient nabývá hodnot z intervalu $< 0; \sqrt{\frac{q-1}{q}} >$.

Dalším koeficientem pro měření intenzity závislosti je **koeficient ϕ (fi)**, **Cramérovo V** (nabývá hodnot z intervalu $< 0; 1 >$) a dále pak **Čuprovův kontingenční koeficient** (pokud se $r = s$, tak je potom tento koeficient roven Cramérovu V). (Řezanková, Analýza kategoriálních dat, 2005)

Goodmanovo-Kruskalovo λ (lambda)

Při zkoumání závislosti sloupcové proměnné Y na řádkové proměnné X mohou nastat tyto dvě situace:

- 1) sloupce jsou nezávislé na řádcích,
- 2) sloupce jsou funkcí řádku.

Pokud jsou známé hodnoty X , ale hodnoty Y jsou neznámé, tak lze tyto hodnoty odhadnout pomocí modální kategorie, pro kterou platí $p_{+Mo} = \max_j(p_{+j})$, za předpokladu situace 1). Při předpokladu situace 2) se hodnoty Y odhadnou pomocí maxima $p_{iMo} = \max_j(p_{ij})$. Pomocí těchto odhadů pak lze vypočítat Goodmanovo-Kruskalovo λ . Tento koeficient nabývá hodnot z intervalu $< 0; 1 >$, Hodnota 0 vyjde v případě, že proměnná X vůbec nepřispívá k predikci (odhadu) proměnné Y a hodnota 1 značí, že se v každém řádku vyskytuje nejvýš jedno pole s nenulovou četností. K výpočtu λ se používá variační poměr. (Řezanková, Analýza kategoriálních dat, 2005)

Goodmanovo-Kruskalovo τ (tau)

Dále lze použít koeficient **Goodmanovo-Kruskalovo τ** , který je založen na principu analýzy nominálního rozptylu místo variačního poměru. Podmínkou výpočtu τ je stejně jako u λ nutné, aby byly nenulové četnosti ve více než v jednom sloupci a taktéž nabývá hodnot z intervalu $< 0; 1 >$. Hodnoty 0 nabývá v případě, že $p_{ij} = p_{i+}p_{+j}$ (pro všechna i a j) a hodnoty 1 nabývá v případě, že se v každém řádku tabulky vyskytuje takové p_{ij} , že se $p_{ij} = p_{i+}$. (Řezanková, Analýza kategoriálních dat, 2005)

Koeficient nejistoty

K výpočtu koeficientu lze použít i entropii („míru neuspořádanosti“) a nazývá se koeficientem nejistoty. K těmto koeficientům patří **Cohenův koeficient κ** (kappa) a používá se v případě, že obě proměnné mají shodný počet kategorií, a sleduje míru souhlasu. Nabývá hodnot z intervalu $< 0; 1 >$, a pokud je hodnota koeficientu 1, tak to znamená, že se nenulové četnosti vyskytly pouze na diagonále. Pokud je hodnota větší než 0,75 znamená dobrý souhlas, a jestliže je hodnota menší než 0,4 o souhlasu tedy nevyovídá. (Řezanková, Analýza kategoriálních dat, 2005)

4.2.2 Testování závislosti ordinálních proměnných

U nominálních proměnných se testuje nezávislost, která se nazývá **Kontingence**, u ordinálních proměnných se této nezávislosti říká **Korelace**. Korelace se rozlišuje na *pozitivní* (nízkým hodnotám odpovídají nízké hodnoty druhé proměnné) a *negativní* (nízkým hodnotám jedné proměnné odpovídají vysoké hodnoty druhé proměnné).

Spearmanův koeficient

Tento koeficient patří k základním koeficientům pořadové korelace. Nabývá hodnot z intervalu $\langle -1; 1 \rangle$. Jestliže platí, že je pořadí u obou proměnných stejné, koeficient má hodnotu 1 (pozitivní korelace). Hodnota -1 znamená negativní korelaci, hodnota 0 pak lineární nezávislost.

Lze testovat nulovost tohoto koeficientu $H_0: \rho_s = 0$ pomocí testové statistiky:

$$t = r_s \sqrt{\frac{n-1}{1-r_s^2}}, \quad (8)$$

která má za platnosti nulové hypotézy Studentovo rozdělení s $(n-2)$ stupni volnosti. (Řezanková, Analýza kategoriálních dat, 2005)

Symetrické míry

Tyto míry zkoumají dvojice objektů, mohou být buď *konkordantním párem* (pokud jsou hodnoty u jednoho objektu u obou proměnných menší, resp. větší než u druhého objektu) nebo *diskordantním párem* (jsou-li hodnoty u jedné proměnné menší a u druhé proměnné větší). Pokud se hodnoty jedné nebo obou proměnných shodují, nazývají se *vázaným párem*. K těmto mírám závislosti patří **Goodmanova-Kruskalova** γ (gama) a platí, že $|\gamma|$ je z intervalu $\langle 0; 1 \rangle$. Hodnota 0 značí nezávislost a hodnota 1 úplnou závislost (pokud jsou všechna nenulová políčka pouze na diagonále). (Řezanková, Analýza kategoriálních dat, 2005)

Další mírou je pak **Kendallovo** τ_b a **Kendallovo** τ_c , což jsou koeficienty pořadové korelace. Oba tyto koeficienty nabývají hodnoty z intervalu $\langle -1; 1 \rangle$. (Řezanková, Analýza kategoriálních dat, 2005)

Kruskal-Wallisův test

Podle (Řezanková, Analýza kategoriálních dat, 2005) se tento test používá v případě, že je vysvětlovaná proměnná Y ordinálního typu a její hodnoty lze rozdělit do skupin podle kategorií vysvětlující proměnné X . Pomocí tohoto testu se tyto skupiny porovnávají. Nulová hypotéza říká, že všechny skupiny mají stejný medián a alternativní hypotéza je stanovena tak, že aspoň jeden medián je jiný. Za platnosti nulové hypotézy o nezávislosti má tato veličina chí-kvadrát rozdělení s $(r-1)$ stupni volnosti.

5 Software STATISTICA

Jedná se o analytický software, který slouží pro správu dat, jejich analýzu, vizualizaci a vývoj uživatelských aplikací.

Společnost StatSoft byla založena v roce 1984 v USA a jejími zakladateli byli vědci a profesori z univerzity, kteří se zabývali kvantitativní analýzou dat. Dnes je společnost rozšířena ve dvaceti zemích světa a stala se jedním z předních světových dodavatelů statistického softwaru pro analýzu dat a reportování. (Profil společnosti, nedatováno)

STATISTICA umožňuje zpracování dat v celé řadě průmyslových odvětví (zejména v oblasti bankovníctví, pojišťovnictví, zdravotnictví, výroba, energetika, potravinářství, telekomunikace atd.). Obsahuje celou řadu užitečných funkcí, které vyhoví všem potřebám každého uživatele. Nabízí mnoho grafických funkcí a velmi boduje ve srovnávacích testech. Ovládání je jednoduché a podobné nástrojům MS Office. V dialogových oknech jednotlivých analýz se snadno orientuje, a ke všemu se lze snadno „doklikat“. Využití počítačové analýzy umožňuje zpracování až desítky tisíc dat, což značně usnadňuje práci a šetří čas.

V horní části STATISTICA obsahuje *klasickou nabídku* nebo si lze přes tlačítko **Zobrazit** otevřít *pás karet* a zobrazí se nabídka obdobná MS Office.

Pro vytvoření jakékoliv analýzy je nejprve zapotřebí mít ověřená data. Tyto data je možno načíst z různých zdrojů či databází. Importovat data lze prakticky ze všech zdrojů (např. text, Excel, MS Word, atd.) nebo data do softwaru vepsat „ručně“ (ovšem pokud je dat velký počet, je to velice zdlouhavé a náročné). Nativní formát po uložení výsledků přímo ze softwaru STATISTICA má potom koncovku *sta*.

5.1 STATISTICA 13.1

Jedná se o nejnovější verzi tohoto softwaru, která nabízí mnoho vylepšení a zjednodušení při používání. Tato verze nabízí řadu funkcí v datové analýze a uživatelé mohou vytvářet jednorázové analytické modely a pracovní postupy. Eliminuje se nadbytečná práce a uživatelům to umožňuje efektivně řešit obchodní problémy bez nadměrných technických znalostí. Další významnou funkcí je vylepšená webová grafika, což uživatelům dovoluje lépe pracovat s výstupy, a zároveň je zdokonaleno

ověřování zadávání dat, která se vkládají do softwaru manuálně. (Nová verze softwaru Dell Statistica pomůže s lepší přípravou dat, 2016) Tato verze je nyní na trhu dostupná v anglickém jazyce.

Praktická práce je zpracována ve verzi STATISTICA 12, která je v českém jazyce a byla autorovi poskytnuta ekonomickou fakultou Jihočeské univerzity.

5.2 Další statistické softwary

Na trhu se objevuje spousta statistických softwarů, které pomáhají uživatelům vytvářet analýzy v mnoha oborech.

- SPSS (IBM SPSS Statistics)

Podle (Matoušek, 2010) je tento statistický software vládce trhu, ale STATISTICA je mu velkým konkurentem. Má převahu nad ostatními softwary hned z několika důvodů. Klade skutečný důraz na efektivitu a rychlost ve zpracování dat. U tohoto softwaru lze využít jak klávesové zkratky, tak i systém „klikání“, který je doplněn podrobnou a vysoce funkční nápovědou.

Výhody: vysoká efektivita,
snadné ovládání,
nápověda.

Nevýhody: vysoká cena na trhu,
vysoký nárok na hardware.

Tento software je taktéž dostupný na ekonomické fakultě Jihočeské univerzity.

- NCSS

Nabízí spoustu funkcí, které jsou zaměřeny na uživatele, kteří používají základní tabulky i pokročilou statistiku. Velkou výhodou je, že je software dostupný za méně než poloviční cenu na trhu. Stejně jako u SPSS je zde dobře a kvalitně zpracovaná nápověda, ale z hlediska rychlosti už zaostává, proto není vhodný pro každodenní používání. (Matoušek, 2010)

- R

Program R je spíše matematický software, který se specializuje na statistiku. Nevýhodou softwaru je, že se ovládá příkazy a ne z menu, což je o dost složitější

na ovládání. Tyto příkazy je nutné znát a zároveň je umět používat. Ovšem, pokud patříte mezi tyto uživatele, tak tento software nabízí téměř vše. Další výhodou je, že je tento software dostupný zdarma. (Matoušek, 2010)

- **MATLAB**

Poslední software, který bude zmíněn, je MATLAB. Je to také matematický software, ve kterém lze provádět analýzu dat a stejně jako program R se ovládá příkazy. Takové ovládání může být pro některé uživatele značná nevýhoda, pokud neznají tento programovací jazyk. Ovšem naopak pro určitou skupinu uživatelů, znající programování a příkazy, je to velké plus, jelikož software pak nabízí spoustu možností při analýze a zpracování dat.

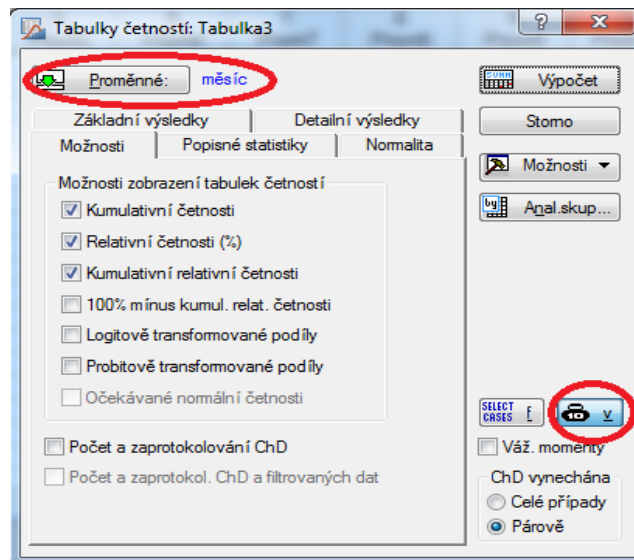
5.3 Metodika – ovládání softwaru STATISTICA

1) Tabulky četností

Nejprve je zapotřebí si do softwaru připravit data pro vytvoření tabulky četností. Do prvního sloupce se napíše proměnná (např. měsíce), pro kterou jsou jednotlivé četnosti zkoumané. Do druhého sloupce se dále tyto četnosti ke každé proměnné nakopírují (přepíší).

Postupuje se přes nabídku *Statistiky* → *Základní statistiky/tabulky* → *Tabulky četností* → *OK*. Přes tlačítko *Proměnné* se zvolí zkoumaná proměnná a pomocí tlačítka **podtržené v**, nacházející se v pravé části formuláře, se nastaví pro proměnnou váha, což je sloupec s pozorovanými četnostmi.

Obrázek 4: Tabulky četností v software STATISTICA



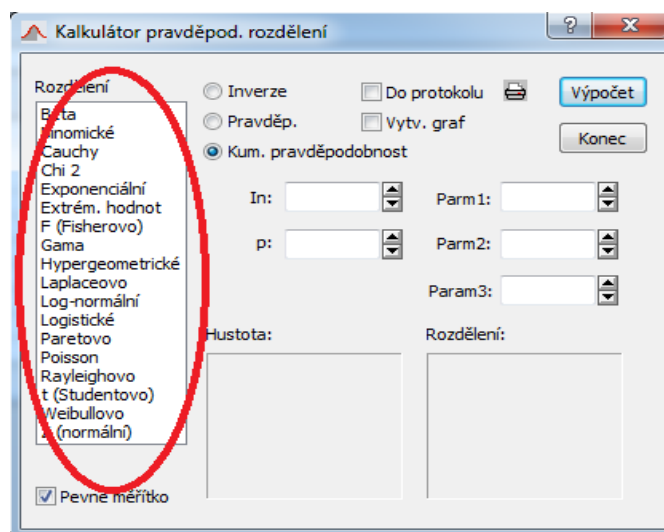
Zdroj: vlastní zpracování v softwaru STATISTICA

Na kartě **Možnosti** si lze vybrat, jaké četnosti jsou požadovány ve výstupní tabulce výpočtů a postup se dokončí tlačítkem **Výpočet**.

2) Binomický test

Aplikuje se na základě vytvoření tabulky četností. Zjištěné údaje z tabulek jsou použity pro výpočet p – *value*, kterou vygeneruje **Pravděpodobnostní kalkulátor**. Nalezne se přes nabídku **Statistiky** → **Pravděpodobnostní kalkulátor** → **Rozdělení** a v levém sloupci je na výběr rozdělení, které je požadováno k výpočtům.

Obrázek 5: Pravděpodobnostní kalkulátor

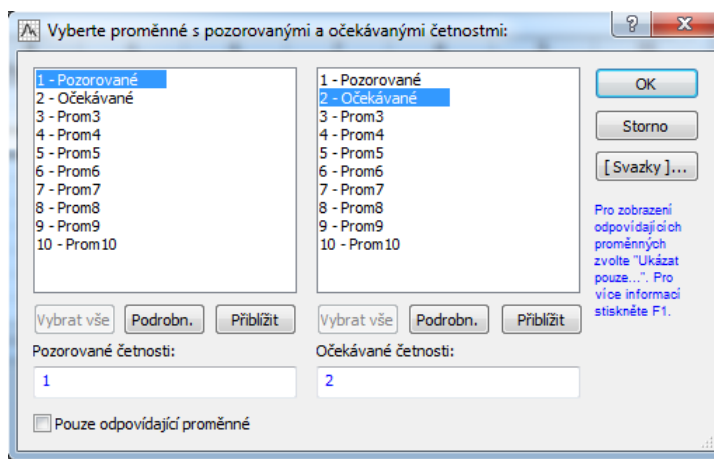


Zdroj: vlastní zpracování v softwaru STATISTICA

3) Chí-kvadrát test

V první řadě je důležité si pro tento test správně připravit data. Do prvního sloupce jsou vypsány pozorované četnosti a do druhého sloupce pak četnosti očekávané. Dále se postupuje přes nabídku *Statistiky* → *Neparametrická statistika* → *Pozorované vs. očekávané X2* → **OK**. Nyní je třeba si nastavit *Proměnné* a potvrdit tlačítkem **OK**.

Obrázek 6: Proměnné pro chí-kvadrát test



Zdroj: vlastní zpracování v softwaru STATISTICA

Pro test, který porovnává četnosti dvou kategorií, lze odhadnout buď společnou hodnotu, nebo jejich rozdíl. Pro tyto odhady lze použít vzorce (5) a (6). Do těchto vzorců se pouze dosadí hodnoty relativních četností vypočtených pro chí-kvadrát test.

4) Kontingenční tabulky pro nominální proměnné

Pro vytvoření kontingenční tabulky v softwaru STATISTICA je opět potřeba si správně připravit data pro výpočet. Do prvního sloupce je důležité vypsát první proměnnou pod sebe tolikrát, kolik je počet obměn. Do druhého sloupce se napíší obměny, stejně pro všechny proměnné. Ve třetím sloupci jsou přiřazeny četnosti, které jsou v dalším kroku třeba nastavit jako váhy k řádkům. V následující tabulce je příklad připravených dat kontingenční tabulky pro plátce a neplátce DPH.

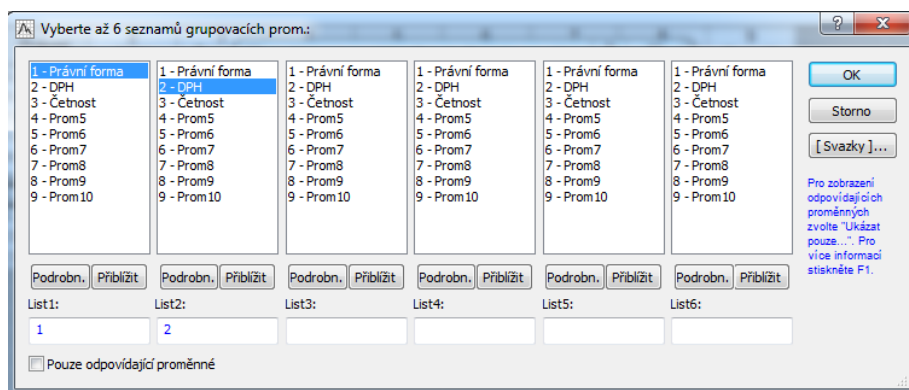
Tabulka 4: Příprava dat pro kontingenční tabulku

	1 Právní forma	2 DPH	3 Četnost
1	s.r.o.	plátce	26
2	s.r.o.	neplátce	19
3	v.o.s.	plátce	10
4	v.o.s.	neplátce	21
5	a.s.	plátce	15
6	a.s.	neplátce	17
7	k.s.	plátce	18
8	k.s.	neplátce	21
9	OSVČ	plátce	0
10	OSVČ	neplátce	6

Zdroj: vlastní zpracování v softwaru STATISTICA

Po přípravě dat se postupuje přes nabídku *Statistiky* → *Základní statistiky a tabulky* → *Kontingenční tabulky* → *OK*. Zobrazí se vyskakovací okno, kde se opět nastaví proměnné, což umožní tlačítko *Speciv. tabulky (vyberte proměn.)*. Lze vybrat až 6 grupovacích proměnných, v těchto analýzách stačí pouze 2, jako například takto:

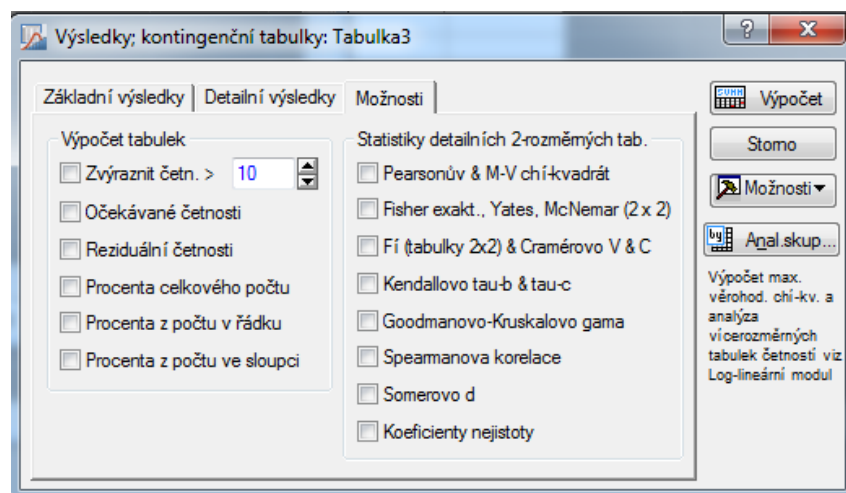
Obrázek 7: Proměnné u kontingenční tabulky



Zdroj: vlastní zpracování v softwaru STATISTICA

Pokračuje se tlačítkem *OK*, které analýzu vrátí o krok zpět k předchozímu vyskakovacímu oknu. Tam je opět třeba nastavit váhy pro proměnné, které jsou vypsány ve třetím sloupci, znovu pomocí tlačítka *podtržené v*. Po kliknutí na tlačítko *OK* se zobrazí další vyskakovací okno pro výpočet. Karta *Možnosti* nabízí spoustu koeficientů a výpočtů, kde si lze zvolit ty, jejichž hodnoty jsou požadovány ve výstupní tabulce.

Obrázek 8: Možnosti kontingenčních tabulek



Zdroj: vlastní zpracování v softwaru STATISTICA

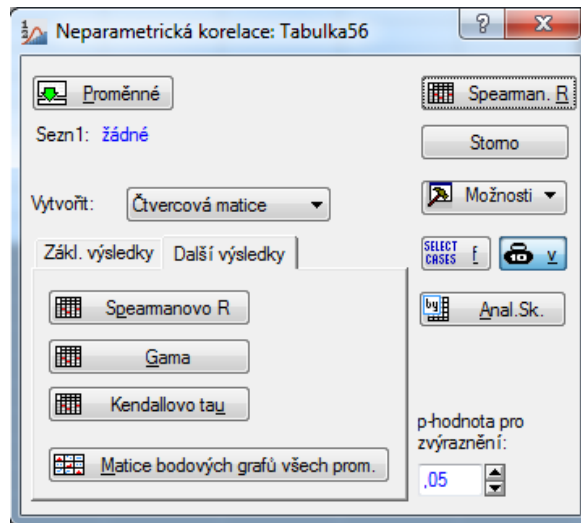
Po výběru výsledků do tabulky se postupuje tlačítkem *Výpočet* a zobrazí se požadovaná kontingenční tabulka. Ovšem pro výpočet detailních výsledků koeficientů je nutné se přepnout na kartu *Detailní výsledky* a pokračovat přes tlačítko *detailní 2-rozm. tabulky*.

5) Kontingenční tabulky pro ordinální proměnné

Data pro analýzu ordinálních proměnných je zapotřebí si připravit přesně tak, jako u analýzy pro proměnné nominální.

K výpočtu koeficientů se postupuje přes nabídku *Statistiky* → *Neparametrická statistika* → *Korelace (Spearman. Kendalovo tau. gama)* → *OK*. Dále je nutné si nastavit proměnné stejným postupem jako u proměnných nominálních. Na kartě *Další výsledky* jsou na výběr požadované koeficienty, jejichž hodnoty budou ve výstupní tabulce.

Obrázek 9: Neparametrická korelace



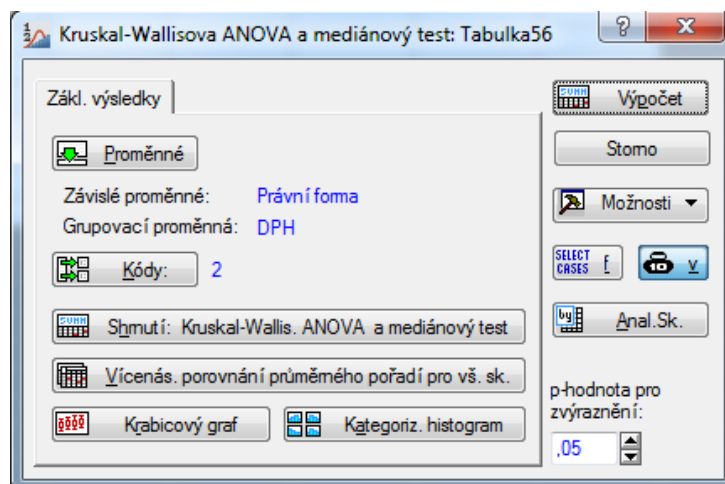
Zdroj: vlastní zpracování v softwaru STATISTICA

6) Kruskal-Wallisova ANOVA

Tento test navazuje na předchozí analýzu Korelace, proto jsou data již v softwaru připravená.

Postupuje se přes nabídku *Statistiky* → *Neparametrická statistika* → *Porovnání více nezávislých vzorků (skupiny)* → *OK*. Znovu se nastaví proměnné přesně jako v předchozí analýze, a pokračuje se tlačítkem *Shrnutí: Kruskal-Wallis.. ANOVA a mediánový test*.

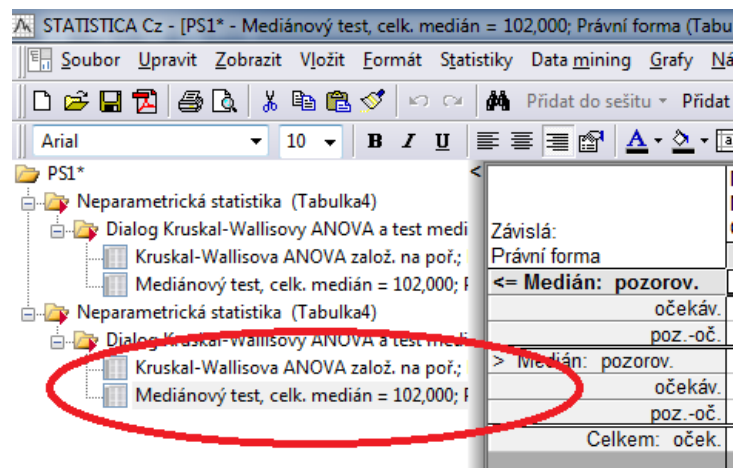
Obrázek 10: Kruskal-Wallisova ANOVA 1



Zdroj: vlastní zpracování v softwaru STATISTICA

Zobrazí se výstupní tabulka pro mediánový test, proto je třeba v levé části okna přepnout na Kruskal-Wallisův test.

Obrázek 11: Kruskal-Wallisova ANOVA 2



Zdroj: vlastní zpracování v softwaru STATISTICA

6 Praktická práce

Zaměřuje se na aplikování testů, hypotéz a koeficientů na konkrétních datech. Budou použity některé vzorce, které jsou uvedené v teoretické části, a pokud není uvedeno jinak, je ke všem výpočtům použit software STATISTICA.

Data pro první část práce autorovi poskytla společnost GRAND, s. r. o. Společnost se zabývá výrobou a prodejem informačních systémů EPIS (Ekonomicko-právní informační systém) na DVD a PROFESIS (obdoba systému EPIS, určen pro jinou skupinu zákazníků) na DVD. Budou zkoumány četnosti prodaných licencí na DVD pro rok 2014 v jednotlivých měsících.

Pro druhou část práce byl vytvořen na serveru www.surveio.com krátký dotazník, který byl elektronicky rozeslán náhodným firmám v ČR.

Struktura dotazníku:

- **Název firmy:** (nepovinné)
- **Právní forma:**
 - Společnost s ručením omezeným
 - Veřejná obchodní společnost
 - Akciová společnost
 - Komanditní společnost
 - OSVČ – podnikatel
- **Základní kapitál (mimo v.o.s. a OSVČ)Kč**
- **Počet společníků (mimo OSVČ)**
- **DPH (daň z přidané hodnoty):**
 - Plátce
 - Neplátce
- **Forma evidence činnosti podniku:**
 - Daňová evidence
 - Účetnictví

Výsledky tohoto dotazníku byly zpracovány do excelové tabulky, která je součástí práce jako Příloha, a následně postupně rozdělena do jednotlivých tabulek použitých v druhé části praktické práce.

6.1 První část – analýza dat ze společnosti GRAND s.r.o.

6.1.1 Znaménkový test s binomickým rozdělením

Uvažují se data Počtu prodaných licencí na DVD programu PROFESIS (jedná se o počet nových uživatelů programu PROFESIS) pro rok 2014, které jsou uvedeny v následující tabulce:

Tabulka 5: Počet prodaných licencí PROFESIS za rok 2014

Měsíc	Licence PROFESIS
leden	5
únor	2
březen	3
duben	2
květen	1
červen	1
červenec	2
srpen	1
září	3
říjen	1
listopad	2
prosinec	1

Zdroj: GRAND, s. r. o.

Do softwaru STATISTICA byly naimportovány data a vytvořena tabulka četností.

Tabulka 6: Tabulka četností – Licence PROFESIS za rok 2014

Kategorie	Tabulka četností: Licence PROFESIS	
	Četnost	Rel. četnost Oček. četnosti
leden	5	20,83333
únor	2	8,33333
březen	3	12,5
duben	2	8,33333
květen	1	4,16667
červen	1	4,16667
červenec	2	8,33333
srpen	1	4,16667
září	3	12,5
říjen	1	4,16667
listopad	2	8,33333
prosinec	1	4,16667

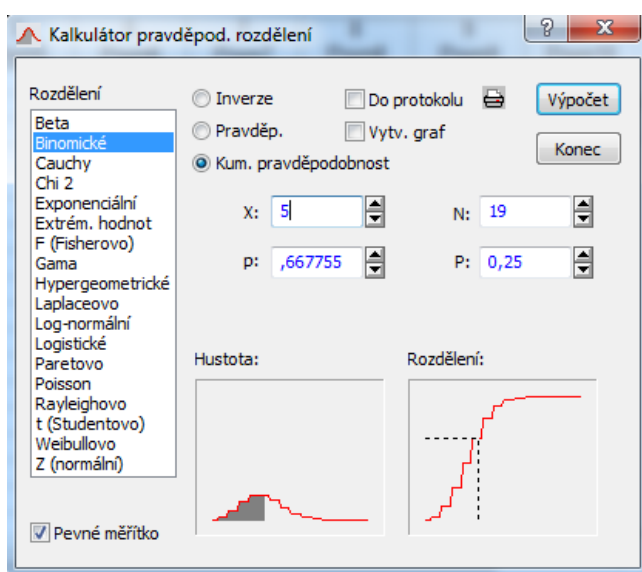
Zdroj: vlastní zpracování v softwaru STATISTICA

Je testována nulová hypotéza $H_0: \pi_1 = \frac{1}{4}$ proti alternativní hypotéze $H_1: \pi_1 < \frac{1}{4}$.

Testuje se, že počet prodaných licencí na DVD v lednu tvoří jednu čtvrtinu z celkového počtu prodaných licencí. Testovaná kategorie je *leden*, která je označena jako n_1 a ostatní kategorie jsou sloučeny do jedné a druhá kategorie je označena jako n_2 .

Alternativní hypotéza je levostranná, jelikož platí, že $n_1 = 5 < n_2 = 19$ a zároveň platí, že p_1 (relativní četnost pro leden) = 0,2083333 < 0,25. Dále platí, že $n_1 + n_2 = 24 \leq 25$, proto je aplikován znaménkový test s binomickým rozdělením. Hypotéza se testuje na hladině $\alpha = 0,05$.

Obrázek 12: Kalkulátor pro binomické rozdělení



Zdroj: vlastní zpracování v softwaru STATISTICA

Pomocí *Pravděpodobnostního kalkulátoru* se v softwaru STATISTICA vypočítá pro *Binomické rozdělení* hodnota p – *value*, která se bude porovnávat se stanovenou hladinou významnosti $\alpha = 0,05$.

Do kalkulátoru se zadají tyto hodnoty, které následně vygenerují hodnotu p – *value*:

$$n_1 = 5 (X); n_2 = 19 (N); \pi_{1,0} = 0,25 (P)$$

Porovnání: p – *value* = 0,667755 > $\alpha = 0,05$.

Závěr: Tento vztah udává, že nulová hypotéza není zamítnuta a alternativní hypotézu nelze přijmout. Tímto testem bylo prokázáno, že počet prodaných licencí systému PROFESIS v lednu tvoří jednu čtvrtinu z celkového počtu prodaných licencí.

6.1.2 Binomický test s normovaným normálním rozdělením

Tentokrát se uvažují data Počet prodaných licencí na DVD produktu EPIS pro rok 2014.

Tabulka 7: Počet prodaných Licencí EPIS za rok 2014

Měsíc	Licence EPIS
leden	75
únor	53
březen	36
duben	39
květen	29
červen	34
červenec	19
srpen	28
září	59
říjen	34
listopad	31
prosinec	25

Zdroj: GRAND, s. r. o.

V softwaru byla jako u prvního příkladu vytvořena tabulka četností, nyní pro licenci EPIS.

Tabulka 8: Tabulka četností – Licence EPIS za rok 2014

Kategorie	Tabulka četností: Licence EPIS	
	Četnost	Rel. četnost Oček. četnosti
leden	75	16,23377
únor	53	11,47186
březen	36	7,79221
duben	39	8,44156
květen	29	6,27706
červen	34	7,35931
červenec	19	4,11255
srpen	28	6,06061
září	59	12,77056
říjen	34	7,35931
listopad	31	6,70996
prosinec	25	5,41126

Zdroj: vlastní zpracování v softwaru STATISTICA

Testuje se nulová hypotéza $H_0: \pi_1 = \frac{1}{8}$ proti levostranné alternativní hypotéze $H_1: \pi_1 < \frac{1}{8}$, že počet prodaných DVD EPIS v únoru tvoří $\frac{1}{8}$ z celkového počtu prodaných licencí.

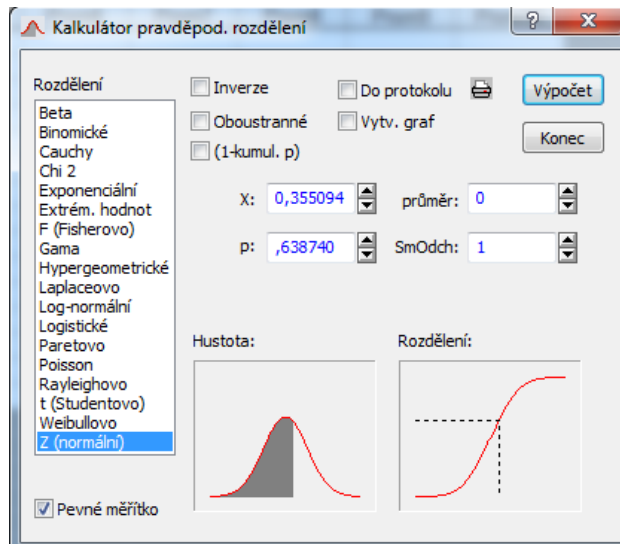
Opět jsou splněny podmínky pro levostrannou alternativu hypotézy, protože $n_1 = 53 < n_2 = 409$ a $p_1 = 0,1147186 < 0,125$. Zde je použit binomický test s normálním normovaným rozdělením, jelikož platí že $n_1 + n_2 = 462 > 25$ a testujeme na hladině významnosti $\alpha = 0,05$.

Nejprve je nutné si vypočítat testovou statistiku Z použitím vzorce (2).

$$Z = \frac{53 - (409 * 0,125) + 0,5}{\sqrt{(409 * 0,125) * (1 - 0,125)}} = 0,355094$$

Pomocí *Pravděpodobnostního kalkulátoru* se určí hodnota p – value pro *Normální normované rozdělení* se směrodatnou odchylkou = 1, průměrem = 0 a vypočítanou testovou statistikou $Z = 0,355094$.

Obrázek 13: Kalkulátor pro normální rozdělení



Zdroj: vlastní zpracování v softwaru STATISTICA

Porovnání: p – value = 0,638740 > $\alpha = 0,05$.

Závěr: Hodnota p – value je prokazatelně větší než stanovená hladina významnosti α , proto nulovou hypotézu nelze zamítnout a alternativní hypotéza tedy není přijata. Tento test potvrdil, že počet prodaných licencí na DVD systému EPIS v únoru tvoří jednu osminu z celkového počtu prodaných licencí.

6.1.3 Chí-kvadrát test dobré shody

Tabulka 9: Pozorované a očekávané četnosti – Licence EPIS za rok 2014

Měsíc	Licence EPIS			
	Pozorované četnosti	Očekávané četnosti	Pozorované relativní četnosti	Očekávané relativní četnosti
leden	75	38,5	0,162338	0,083333
únor	53	38,5	0,114719	0,083333
březen	36	38,5	0,077922	0,083333
duben	39	38,5	0,084416	0,083333
květen	29	38,5	0,062771	0,083333
červen	34	38,5	0,073593	0,083333
červenec	19	38,5	0,041126	0,083333
srpen	28	38,5	0,060606	0,083333
září	59	38,5	0,127706	0,083333
říjen	34	38,5	0,073593	0,083333
listopad	31	38,5	0,0671	0,083333
prosinec	25	38,5	0,054113	0,083333

Zdroj: vlastní zpracování, vychází z tabulky 6

Očekávané četnosti jsou dány jako $n\pi_{i,0} = 462 * \frac{1}{12} = 38,5$ a očekávané relativní četnosti $= \frac{1}{12}$. Na hladině významnosti $\alpha = 0,05$ je testována nulová hypotéza $H_0: \pi_i = \frac{1}{12}$, $i = 1, 2, \dots, 12$ proti alternativní oboustranné hypotéze $H_1: \text{non } H_0$. Nulová hypotéza udává, že je relativní četnost v jednotlivých měsících stejná.

V softwaru STATISTICA je pomocí *Neparametrické statistiky* vytvořena následující tabulka pozorovaných a očekávaných četností pro chí-kvadrát test.

Tabulka 10: Chí-kvadrát test dobré shody pro shodné očekávané četnosti – Licence EPIS za rok 2014

Případ	Pozorované vs. očekávané četnosti Chi-Kvadr. = 73,48052 sv = 11 p = 0,000000			
	Pozor. četnosti	Očekáv. četnosti Oček. četnosti	P - O	(P-O) ² /O
leden	75	38,5	36,5	34,60390
únor	53	38,5	14,5	5,46104
březen	36	38,5	-2,5	0,16234
duben	39	38,5	0,5	0,00649
květen	29	38,5	-9,5	2,34416
červen	34	38,5	-4,5	0,52597
červenec	19	38,5	-19,5	9,87662
srpen	28	38,5	-10,5	2,86364
září	59	38,5	20,5	10,91558
říjen	34	38,5	-4,5	0,52597
listopad	31	38,5	-7,5	1,46104
prosinec	25	38,5	-13,5	4,73377
Sčt	462	462	0	73,48052

Zdroj: vlastní zpracování v softwaru STATISTICA

Testová statistika $\chi^2 = 73,48052$.

Porovnání: $p - value = 0,0000 < \alpha = 0,05$.

Závěr: Hodnota p – value je výrazně nižší než hladina významnosti α , z toho plyne, že je zamítnuta nulová hypotéza ve prospěch alternativní hypotézy. Nebylo prokázáno, že je počet prodaných DVD systému EPIS v jednotlivých měsících stejný. Což tedy znamená, že se četnosti v jednotlivých měsících od sebe prokazatelně liší.

Nyní jsou uvažována data pozorovaných a očekávaných četností pro Licenci EPIS za rok 2014. V následující tabulce se změnilo očekávané četnosti, které jsou pro každý měsíc různé.

Tabulka 11: Pozorované a očekávané četnosti 2 – Licence EPIS za rok 2014

Měsíc	DVD EPIS		
	Pozorované četnosti	Očekávané relativní četnosti 2	Očekávané četnosti 2
leden	75	0,2	92,4
únor	53	0,1	46,2
březen	36	0,08	36,96
duben	39	0,08	36,96
květen	29	0,06	27,72
červen	34	0,07	32,34
červenec	19	0,04	18,48
srpen	28	0,06	27,72
září	59	0,11	50,82
říjen	34	0,07	32,34
listopad	31	0,07	32,34
prosinec	25	0,06	27,72
Celkem	462	1	462

Zdroj: vlastní zpracování, vychází z tabulky 8

Tyto očekávané relativní četnosti jsou stanoveny na základě relativních pozorovaných četností z předchozí tabulky a očekávané četnosti jsou vypočítány jako $n\pi_{i,0}$.

Testuje se na hladině významnosti $\alpha = 0,05$ nulová hypotéza $H_0: \pi_{i,0} = \pi_{i,0}^2$ proti oboustranné alternativní hypotéze $H_1: \text{non } H_0$. Nulová hypotéza stanovuje, že pozorované relativní četnosti jsou postupně rovny očekávaným relativním četnostem 2.

Tabulka 12: Chí-kvadrát test dobré shody pro různé očekávané četnosti – Licence EPIS za rok 2014

Případ	Pozorované vs. očekávané četnosti Chi-Kvadr. = 6,301077 sv = 11 p = 0,852537			
	Pozor. četnosti	Očekáv. četnosti Oček. četnosti	P - O	(P-O) ² /O
leden	75	92,4	-17,4	3,276623
únor	53	46,2	6,8	1,000866
březen	36	36,96	-0,96	0,024935
duben	39	36,96	2,04	0,112597
květen	29	27,72	1,28	0,059105
červen	34	32,34	1,66	0,085207
červenec	19	18,48	0,52	0,014632
srpen	28	27,72	0,28	0,002828
září	59	50,82	8,18	1,316655
říjen	34	32,34	1,66	0,085207
listopad	31	32,34	-1,34	0,055523
prosinec	25	27,72	-2,72	0,266898
Sčt	462	462	- 0	6,301077

Zdroj: vlastní zpracování v softwaru STATISTICA

Testová statistika $\chi^2 = 6,301077$.

Porovnání: $p - value = 0,85237 > \alpha = 0,05$.

Závěr: Hodnota p – value je prokazatelně větší než stanovená hladina významnosti α , což znamená, že není zamítnuta nulová hypotéza a nelze přijmout hypotézu alternativní. Je podloženo, že se relativní četnosti prodaných licencí EPIS na DVD rovnají druhým přepočítaným očekávaným četnostem.

6.1.4 Test porovnávací četnosti dvou kategorií

Při této analýze budou porovnány četnosti prodeje licencí EPIS ve dvou vybraných měsících (červenec a září).

Tabulka 13: Počet prodaných licencí systému EPIS – červenec a září

Měsíc	Licence EPIS			
	Pozorované četnosti	Očekávané četnosti	Pozorované relativní četnosti	Očekávané relativní četnosti
červenec	19	39	24,359	0,5
září	59	39	75,641	0,5

Zdroj: vlastní zpracování, vychází z tabulky 8

Je testována nulová hypotéza $H_0: \pi_1 = \pi_2$ proti alternativní oboustranné hypotéze $H_0: \pi_1 \neq \pi_2$ na hladině významnosti $\alpha = 0,05$. Nulová hypotéza udává, že se četnosti v obou měsících sobě rovnají. Je zde splněna podmínka, že rozsah souboru $n = 78 > 30$, a proto lze použít statistiku chí-kvadrát, jehož hodnota je uvedena v následující tabulce.

Tabulka 14: Test dobré shody – Měsíc červenec a září

Případ	Pozorované vs. očekávané četnosti Chi-Kvadr. = 20,51282 sv = 1 p = 0,000006			
	Pozor. četnosti	Očekáv. četnosti Oček. četnosti	P - O	(P-O) ² / O
červenec	19	39	-20	10,25641
září	59	39	20	10,25641
Sčt	78	78	0	20,51282

Zdroj: vlastní zpracování v softwaru STATISTICA

Testová statistika $\chi^2 = 20,51282$.

Porovnání: $p - value = 0,000006 < \alpha = 0,05$.

Jestliže byla zamítnuta nulová hypotéza, lze odhadnout hodnotu rozdílu podle vzorce (6). K výpočtu je třeba znát hodnoty p_1 a p_2 , což jsou pozorované relativní četnosti pro měsíce červenec a září, které udává tabulka 9.

$$p_{\text{odhad rozdílu}} = 0,041126 - 0,127706 = -0,08658$$

Závěr: Hodnota $p - value$ je výrazně blízká k 0, a to dává signál, že je zamítnuta nulová hypotéza ve prospěch alternativní hypotézy, kterou lze přijmout. Není potvrzeno,

že se relativní četnosti prodaných licencí EPIS v červenci a v září sobě rovnají. Tyto hodnoty se od sebe liší o hodnotu 0,08658.

6.1.5 Analýza dvourozměrné kontingenční tabulky (nominální proměnné)

Nyní bude zkoumána závislost druhu prodané licence v jednotlivých měsících roku 2014 a jejich přehled znázorňuje následující kontingenční tabulka.

Tabulka 15: Kontingenční tabulka – Prodané licence za rok 2014

Kontingenční tabulka Tab.:			
Měsíc	licence EPIS	licence PROFESIS	Řádk. součty
leden	75	5	80
únor	53	2	55
březen	36	3	39
duben	39	2	41
květen	29	1	30
červen	34	1	35
červenec	19	2	21
srpen	28	1	29
září	59	3	62
říjen	34	1	35
listopad	31	2	33
prosinec	25	1	26
Vš. skup.	462	24	486

Zdroj: vlastní zpracování v softwaru STATISTICA

Je testována nulová hypotéza $H_0: \pi_{ij} = \pi_{ij,0}$, která stanovuje, že se pozorované četnosti rovnají četnostem očekávaným. Na hladině významnosti $\alpha = 0,05$ se testuje proti alternativní oboustranné hypotéze $H_1: \pi_{ij} \neq \pi_{ij,0}$.

V softwaru STATISTICA se zpracuje tabulka očekávaných četností, která vychází z předchozí kontingenční tabulky, a dále je vytvořena tabulka statistických koeficientů závislosti.

Tabulka 16: Očekávané četnosti – Prodané licence za rok 2014

Souhrnná tab.: Očekávané četnosti Pearsonův chí-kv.: 3,16637 sv=11 p=0,988339			
Měsíc	licence EPIS	licence PROFESIS	Řádk. součty
leden	76,0494	3,95062	80
únor	52,2840	2,71605	55
březen	37,0741	1,92593	39
duben	38,9753	2,02469	41
květen	28,5185	1,49148	30
červen	33,2716	1,72840	35
červenec	19,9630	1,03704	21
srpen	27,5679	1,43210	29
září	58,9383	3,06173	62
říjen	33,2716	1,72840	35
listopad	31,3704	1,62963	33
prosinec	24,7160	1,28395	26
Vš. skup.	462	24	486

Zdroj: vlastní zpracování v softwaru STATISTICA

Testová statistika $\chi^2 = 3,16637$.

Tabulka 17: Koeficienty závislosti – Podané licence za rok 2014

Statist.	Statist.: měsíc (12) x licence (2)		
	Chí-kvadr.	sv	p
Pearsonův chí-kv.	3,166372	df=11	p=0,98834
Fí	0,0807166		
Čuprovův kontingenční koef.	0,0804549		
Cramér. V	0,0807166		
Koeficient nejistoty	X=0,0012927	Y=0,0158479	X/Y=0,00239

Zdroj: vlastní zpracování v softwaru STATISTICA

Koeficient $f_i = 0,0807166$,

Čuprovův kontingenční koeficient = $0,0804549$,

Cramérovo V = $0,0807166$,

Koeficient nejistoty = $0,0012927$.

blízké k 0

Pro tuto analýzu slouží ještě koeficienty Goodmanova-Kruskalova lambda a Goodmanovo-Kruskalovo tau. Bohužel software STATISTICA tyto koeficienty

nenabízí, proto je třeba pro výpočet použít jiný software (např. SPSS). Ovšem v tomto případě to není třeba, jelikož hodnota p – *value* dává jasný signál pro nezamítnutí nulové hypotézy.

Porovnání: $p - value = 0,98834 > \alpha = 0,05$.

Závěr: Hodnota p – value je výrazně vyšší než hladina významnosti α , proto není zamítnuta nulová hypotéza, která tvrdí, že jsou tyto dvě proměnné na sobě nezávislé. Hodnoty všech koeficientů se přibližují k 0, a proto tuto nezávislost a nezamítnutí nulové hypotézy potvrzují. Lze tedy konstatovat, že druh prodané licence není závislý na měsíci prodeje.

6.2 Druhá část – analýza dat z dotazníkového šetření

6.2.1 Chí-kvadrát test dobré shody

Analýza se zaměřuje na četnosti společností podle právní formy podnikání. Nejprve jsou ve STATISTICE vypočítány relativní četnosti, a následně je vytvořena tabulka pozorovaných a očekávaných četností.

Tabulka 18: Relativní četnosti – Právní forma

Kategorie	Tabulka četností: Právní forma	
	Četnost	Rel. četnost Oček. četnosti
s.r.o.	45	29,41176
v.o.s.	31	20,26144
k.s.	39	25,49020
a.s.	32	20,91503
OSVČ	6	3,92157

Zdroj: vlastní zpracování v softwaru STATISTICA

Očekávané četnosti $n\pi_{i,0} = 153 * \frac{1}{5} = 30,6$ a očekávané relativní četnosti $= \frac{1}{5}$.

Tabulka 19: Pozorované a očekávané četnosti – Právní forma

Právní forma	Pozorované četnosti	Očekávané četnosti	Pozorované relativní četnosti	Očekávané relativní četnosti
s.r.o.	45	30,6	29,41176	0,2
v.o.s.	31	30,6	20,26144	0,2
k.s.	39	30,6	25,49020	0,2
a.s.	32	30,6	20,91503	0,2
OSVČ	6	30,6	3,92157	0,2

Zdroj: vlastní zpracování, vychází z Přílohy

Testuje se na hladině významnosti $\alpha = 0,05$ nulová hypotéza $H_0: \pi_i = \frac{1}{5}$, $i = 1, 2, \dots, 5$ proti alternativní oboustranné hypotéze $H_1: \text{non } H_0$. Nulová hypotéza konstatuje, že je relativní četnost zastoupení společností stejná.

Tabulka 20: Chí-kvadrát test dobré shody pro shodné očekávané četnosti – Právní forma

Případ	Pozorované vs. očekávané četnosti Chi-Kvadr. = 28,92810 sv = 4 p = 0,000008			
	Pozor. četnosti	Očekáv. četnosti Oček. četnosti	P - O	(P-O)^2 /O
s.r.o.	45	31	14,4	6,77647
v.o.s.	31	31	0,4	0,00523
k.s.	39	31	8,4	2,30588
a.s.	32	31	1,4	0,06405
OSVČ	6	31	-24,6	0,831169
Sčt	153	153	- 0	28,92810

Zdroj: vlastní zpracování v softwaru STATISTICA

Testová statistika $\chi^2 = 28,92810$.

Porovnání: $p - value = 0,000008 < \alpha = 0,05$.

Závěr: Hodnota p – value je prokazatelně blízká 0, a proto lze na této hladině významnosti α zamítnout nulovou hypotézu ve prospěch alternativní hypotézy. Alternativní hypotéza říká, že se pozorované četnosti nerovnjají četnostem očekávaným.

Nyní jsou použita stejná data jako v předchozím příkladu, jen bez posledního řádku OSVČ. Podle toho jsou upraveny očekávané a relativní četnosti.

Tabulka 21: Pozorované a očekávané četnosti – Právní forma (bez OSVČ)

Právní forma	Pozorované četnosti	Očekávané četnosti	Pozorované relativní četnosti	Očekávané relativní četnosti
s.r.o.	45	36,75	0,30612	0,25
v.o.s.	31	36,75	0,21088	0,25
k.s.	39	36,75	0,26531	0,25
a.s.	32	36,75	0,21769	0,25

Zdroj: vlastní zpracování

Testuje se na hladině významnosti $\alpha = 0,05$ nulová hypotéza $H_0: \pi_i = \frac{1}{4}$, $i = 1, 2, \dots, 4$ proti alternativní oboustranné hypotéze $H_1: \text{non } H_0$. Nulová hypotéza stanovuje, že jsou relativní četnosti zastoupení společností stejné. Toto tvrzení podloží nový chí-kvadrát test.

Tabulka 22: Chí-kvadrát test dobré shody pro shodné očekávané četnosti – právní forma (bez OSVČ)

Případ	Pozorované vs. očekávané četnosti Chi-Kvadr. = 3,503401 sv = 3 p = 0,320323			
	Pozor. četnosti	Očekáv. četnosti Oček. četnosti	P - O	(P-O) ² /O
s.r.o.	45	36,75	8,25	1,852041
v.o.s.	31	36,75	-5,75	0,899660
k.s.	39	36,75	2,25	0,137755
a.s.	32	36,75	-4,75	0,613946
Sčt	147	147	0	3,503401

Zdroj: vlastní zpracování v softwaru STATISTICA

Testová statistika $\chi^2 = 3,503401$.

Porovnání: $p - value = 0,320323 > 0,05$.

Závěr: Hodnota p – value je prokazatelně větší než stanovená hladina významnosti α , a proto není nulová hypotéza zamítnuta a alternativní hypotézu nelze nepřijmout. Tím, že bylo odstraněno zastoupení právní formy OSVČ, které mělo znatelně nižší četnost, bylo prokázáno, že se očekávané četnosti rovnají četnostem očekávaným.

6.2.2 Test porovnávací četnosti dvou kategorií

Vytvořením tabulky pozorovaných a očekávaných četností, je zkoumáno, zda společnost odvádí DPH.

Tabulka 23: Pozorované a očekávané četnosti – DPH

DPH	Pozorované četnosti	Očekávané četnosti	Pozorované relativní četnosti	Očekávané relativní četnosti
plátce	68	77	0,444444	0,5
neplátce	85	77	0,555556	0,5

Zdroj: vlastní zpracování, vychází z Přílohy

Je testována nulová hypotéza $H_0: \pi_1 = \pi_2$ proti alternativní oboustranné hypotéze $H_0: \pi_1 \neq \pi_2$ a porovnává se s hladinou významnosti $\alpha = 0,05$. Nulová hypotéza stanovuje, že se četnosti pro plátce a neplátce DPH sobě rovnají. Je splněna podmínka, že rozsah souboru $n = 153 > 30$, proto lze použít statistiku chí-kvadrát, kterou software STATISTICI určí.

Tabulka 24: Chí-kvadrát test – DPH

Případ	Pozorované vs. očekávané četnosti Chi-Kvadr. = 1,888889 sv = 1 p = 0,169328			
	Pozor. četnosti	Očekáv. četnosti Oček. četnosti	P - O	(P-O) ² /O
plátce	68	76,5	-8,5	0,944444
neplátce	85	76,5	8,5	0,944444
Sčt	153	153	0	1,662338

Zdroj: vlastní zpracování v softwaru STATISTICA

Testová statistika $\chi^2 = 1,888889$.

Porovnání: $p - value = 0,169328 > 0,05$.

Toto porovnání určuje, že není zamítnuta nulová hypotéza, proto lze odhadnout společnou hodnotu π pomocí vzorce (5). Opět je třeba znát hodnoty p_1 a p_2 , což jsou nyní pozorované relativní četnosti pro plátce a neplátce DPH, které jsou uvedeny v tabulce 23.

$$\text{Podhad společné hodnoty} = \frac{1}{2} * (0,444444 + 0,555556) = 0,5$$

Závěr: Hodnota p – value je větší než stanovená hladina významnosti α , proto není zamítnuta nulová hypotéza a nepřijímá se alternativní hypotéza. Je doloženo, že se relativní četnosti plátců a neplátců DPH sobě rovnají, a jejich společná hodnota $\pi = 0,5$.

Nyní jsou porovnány četnosti, jakou formu evidence příjmů a výdajů využívají pozorované společnosti.

Tabulka 25: Pozorované a očekávané četnosti – evidence

Evidence	Pozorované četnosti	Očekávané četnosti	Pozorované relativní četnosti	Očekávané relativní četnosti
účetnictví	147	76,5	0,954545	0,5
daňová	6	76,5	0,039216	0,5

Zdroj: vlastní zpracování, vychází z Přílohy

Je testována nulová hypotéza $H_0: \pi_1 = \pi_2$ proti alternativní oboustranné hypotéze $H_0: \pi_1 \neq \pi_2$ a porovnává se s hladinou významnosti $\alpha = 0,05$. Nulová hypotéza stanovuje, že se četnosti pro vedení účetnictví a daňové evidence sobě rovnají.

Je splněna podmínka, že rozsah souboru $n = 153 > 30$, proto lze použít statistiku chí-kvadrát, která je zpracována ve výstupní tabulce.

Tabulka 26: Chí-kvadrát test – evidence

Případ	Pozorované vs. očekávané četnosti Chi-Kvadr. = 129,9412 sv = 1 p = 0,000000			
	Pozor. četnosti	Očekáv. četnosti Oček. četnosti	P - O	(P-O) ² /O
účetnictví	147	76,5	70,5	64,9706
daňová evidence	6	76,5	-70,5	64,9706
Sčt	153	153	0	129,9412

Zdroj: vlastní zpracování v softwaru STATISTICA

Testová statistika $\chi^2 = 129,9412$.

Porovnání: **$p - value = 0,000000 > 0,05$** .

Z této hodnoty je jasné, že je zamítnuta nulová hypotéza, proto lze odhadnout rozdíl pomocí vzorce (6). Tento rozdíl se odhadne pomocí hodnot p_1 a p_2 , což jsou nyní pozorované relativní četnosti vedení účetnictví a daňové evidence.

$$p_{\text{odhad rozdíl}} = 0,954545 - 0,039216 = 0,915329$$

Závěr: Hodnota p – value je rovna 0, což znamená, že je menší než stanovená hladina významnosti α , a proto je zamítnuta nulová hypotéza a přijímá se alternativní hypotéza. Je prokázáno, že se relativní četnosti firem, které vedou účetnictví či daňovou evidenci, se sobě nerovnají a jejich rozdíl = 0,915329. Tato analýza je jednoznačná, jelikož účetnictví mají povinnost vést všechny právnické osoby, což je důsledek tohoto nepoměru četností.

6.2.3 Analýza dvourozměrné kontingenční tabulky (nominální proměnné)

Je pozorována nezávislost druhu právní formy společnosti podle kritéria, zda je plátce či neplátce DPH pomocí následující kontingenční tabulky.

Tabulka 27: Kontingenční tabulka – DPH (nomin. proměn.)

Kontingenční tabulka Tab.:			
Právní forma	DPH plátce	DPH neplátce	Řádk. součty
s.r.o.	26	19	45
v.o.s.	10	21	31
k.s.	15	17	32
a.s.	18	21	39
OSVČ	0	6	6
Vš. skup.	69	84	153

Zdroj: vlastní zpracování v softwaru STATISTICA

Testuje se nulová hypotéza $H_0: \pi_{ij} = \pi_{ij,0}$, která konstatuje, že se četnosti π_{ij} rovnají četnostem $\pi_{ij,0}$ při nezávislosti. Na hladině významnosti $\alpha = 0,05$ se testuje proti alternativní oboustranné hypotéze $H_1: \pi_{ij} \neq \pi_{ij,0}$.

Vytvoří se tabulka očekávaných četností, která vychází z předchozí kontingenční tabulky a následně tabulka statistických koeficientů závislosti.

Tabulka 28: Očekávané četnosti – DPH

Souhrnná tab.: Očekávané četnosti Pearsonův chí-kv.: 9,97315 sv=4 p=0,040882			
Právní forma	DPH plátce	DPH neplátce	Řádk. součty
s.r.o.	20,29412	24,70588	45
v.o.s.	13,98039	17,01961	31
k.s.	14,43137	17,56863	32
a.s.	17,58824	21,41176	39
OSVČ	2,70588	3,29412	6
Vš. skup.	69	84	153

Zdroj: vlastní zpracování v softwaru STATISTICA

Testová statistika $\chi^2 = 9,97315$.

Tabulka 29: Koeficienty závislosti – DPH (nomin. proměn.)

Statist.	Statist.: Právní forma (5) x DPH (2)		
	Chí-kvadr.	sv	p
Pearsonův chí-kv.	9,973154	df=4	p=0,04088
Fí	0,2553116		
Čuprovův kontingenční koef.	0,2473764		
Cramér. V	0,2553116		
Koeficient nejistoty	X=0,0270127	Y=0,0583192	X/Y=0,03692

Zdroj: vlastní zpracování v softwaru STATISTICA

Koeficient $f_i = 0,2553116$,

Čuprovův kontingenční koeficient = $0,2473764$,

Cramérovo $V = 0,2553116$,

Koeficient nejistoty = $0,0270127$.

} > 0

Porovnání: $p - value = 0,04088 < \alpha = 0,05$.

Závěr: Hodnota p – value je nižší než hladina významnosti α , proto zamítáme nulovou hypotézu a lze říci, že tyto dvě proměnné jsou na sobě závislé. Hodnoty všech koeficientů jsou větší než 0, a proto potvrzují zamítnutí nulové hypotézy a závislost. Výsledek této analýzy udává, že jestli je společnost plátce či neplátce DPH závisí na právní formě společnosti.

Nyní se autor věnuje základnímu kapitálu u akciové společnosti, komanditní společnosti a u společnosti s ručením omezeným. OSVČ a veřejná obchodní společnost jsou vynechány z důvodu, že povinně netvoří základní kapitál. Přehled základního kapitálu u vybraných právních forem zobrazuje následující kontingenční tabulka.

Tabulka 30: Kontingenční tabulka – Základní kapitál (nomin. proměn.)

Kontingenční tabulka Tab.:					
Právní forma	Základní kapitál - 1 000 000	Základní kapitál 1 000 001 – 5 000 000	Základní kapitál 5 000 001 – 10 000 000	Základní kapitál 10 000 000 +	Řádk. součty
s.r.o.	29	5	5	5	44
a.s.	5	12	3	12	32
k.s.	5	35	3	1	45
Vš. skup.	40	52	11	18	121

Zdroj: vlastní zpracování v softwaru STATISTICA

Na hladině významnosti $\alpha = 0,05$ se testuje nulová hypotéza $H_0: \pi_{ij} = \pi_{ij,0}$, která udává, že se četnosti π_{ij} rovnají četnostem $\pi_{ij,0}$ při nezávislosti. Nulová hypotéza se porovnává proti alternativní oboustranné hypotéze $H_1: \pi_{ij} \neq \pi_{ij,0}$.

V softwaru se zpracuje tabulka očekávaných četností, která vychází z předchozí tabulky a dále tabulka statistických koeficientů závislosti.

Tabulka 31: Očekávané četnosti – Základní kapitál

Souhrnná tabulka: Očekávané četnosti Pearsonův chí-kv.: 62,5019 sv=6 p=0,000000					
Právní forma	Základní kapitál - 1 000 000	Základní kapitál 1 000 001 – 5 000 000	Základní kapitál 5 000 001 – 10 000 000	Základní kapitál 10 000 000 +	Řádk. součty
s.r.o.	14,514545	18,90909	4	6,54545	44
a.s.	10,57851	13,75207	2,90909	4,76033	32
k.s.	14,87603	19,33834	4,09091	6,69421	45
Vš. skup.	40	52	11	18	121

Zdroj: vlastní zpracování v softwaru STATISTICA

Testová statistika $\chi^2 = 62,5019$.

Tabulka 32: Koeficienty závislosti – Základní kapitál (nomin. proměn.)

Statist.	Statist.: Právní forma (3) x Základní kapitál (4)		
	Chí-kvadr.	sv	p
Pearsonův chí-kv.	62,50190	df=6	p=0,00000
F _i	0,7187104		
Čuprovův kontingenční koef.	0,5836148		
Cramér. V	0,5082050		
Koeficient nejistoty	X=0,2387625	Y=0,2110408	X/Y=0,22405

Zdroj: vlastní zpracování v softwaru STATISTICA

Koeficient $f_i = 0,7187104$,
 Čuprovův kontingenční koeficient = $0,5836148$,
 Cramérovu $V = 0,5082050$,
 Koeficient nejistoty = $0,2387625$.

} > 0

Porovnání: $p - value = 0,00000 < \alpha = 0,05$.

Závěr: Hodnota p – value je rovna 0 a je tedy nižší než hladina významnosti α , proto je zamítnuta nulová hypotéza a lze říci, že tyto dvě proměnné jsou na sobě závislé. Hodnoty všech koeficientů jsou výrazně větší než 0, a proto potvrzují tuto závislost. Je tedy jednoznačné, že právní forma závisí na výši základního kapitálu.

V posledním případě je tato analýza zaměřena na počet společníků. Věnuje se všem právním formám mimo OSVČ a přehled zobrazuje následující kontingenční tabulka.

Tabulka 33: Kontingenční tabulka – Společníci (nomin. proměn.)

Kontingenční tabulka					
Tab.:					
Právní forma	Společníci 1 - 2	Společníci 3 - 5	Společníci 6 – 10	Společníci 11 +	Řádk. součty
s.r.o.	5	10	17	1	33
v.o.s.	16	5	4	6	31
a.s.	28	9	5	2	44
k.s.	14	18	6	1	39
Vš. skup.	63	42	32	10	147

Zdroj: vlastní zpracování v softwaru STATISTICA

Opět se testuje nulová hypotéza $H_0: \pi_{ij} = \pi_{ij,0}$, která říká, že se četnosti π_{ij} rovnají četnostem $\pi_{ij,0}$ při nezávislosti. Na hladině významnosti $\alpha = 0,05$ se testuje proti alternativní oboustranné hypotéze $H_1: \pi_{ij} \neq \pi_{ij,0}$.

Následující tabulka očekávaných četností vychází z předchozí kontingenční tabulky a zároveň je vytvořena tabulka koeficientů závislosti.

Tabulka 34: Očekávané četnosti – Společníci

Souhrnná tabulka: Očekávané četnosti Pearsonův chí-kv.: 44,9688 sv=9 p=0,000001					
Právní forma	Společníci 1 - 2	Společníci 3 - 5	Společníci 6 - 10	Společníci 11 +	Řádk. součty
s.r.o.	14,14286	9,42857	7,18367	2,24490	33
v.o.s.	13,28571	8,85714	6,74830	2,10884	31
a.s.	18,85714	12,57143	9,57823	2,99320	44
k.s.	16,71429	11,14286	8,48980	2,65306	39
Vš. skup.	63	42	32	10	147

Zdroj: vlastní zpracování v softwaru STATISTICA

Testová statistika $\chi^2 = 44,9688$.

Tabulka 35: Koeficienty závislosti – Společníci (nomin. proměn.)

Statist.	Statist.: Právní forma (4) x Společníci (4)		
	Chí-kvadr.	sv	p
Pearsonův chí-kv.	44,96875	df=9	p=0,00000
Fí	0,5530912		
Čuprovův kontingenční koef.	0,4839942		
Cramér. V	0,3193274		
Koeficient nejistoty	X=0,1018880	Y=0,1135019	X/Y=0,10738

Zdroj: vlastní zpracování v softwaru STATISTICA

Koeficient $f_i = 0,5530912$,

Čuprovův kontingenční koeficient = $0,4839942$,

Cramérovu $V = 0,3193274$,

Koeficient nejistoty = $0,1018880$.

} > 0

Porovnání: $p - value = 0,00000 < \alpha = 0,05$.

Závěr: Jednoznačně se zamítá nulová hypotéza, jelikož je hodnota $p - value$ rovna 0 a je tedy nižší než hladina významnosti α . Zamítnutí nulové hypotézy znamená,

že tyto dvě proměnné jsou na sobě závislé. Hodnoty všech koeficientů jsou výrazně větší než 0, tak tuto závislost potvrzují. Lze tedy konstatovat, že právní forma závisí na tom, kolik je ve společnosti společníků.

6.2.4 Analýza dvourozměrné kontingenční tabulky (ordinální proměnné)

Při těchto analýzách nulová hypotéza udává, že mezi testovanými proměnnými není žádná závislost a alternativní hypotéza tvrdí opak (závislost mezi nimi je).

Je pozorována nezávislost druhu právní formy společnosti podle kritéria, zda je plátce či neplátce DPH. Tyto data jsou uvažována jako ordinální a testuje se, zda jsou tyto proměnné na sobě nezávislé. Tedy H_0 : nezávislost a H_1 : závislost.

Tabulka 36: Kontingenční tabulka – DPH (ordin. proměn.)

Kontingenční tabulka Tab.:			
Právní forma	DPH plátce	DPH neplátce	Řádk. součty
s.r.o.	26	19	45
v.o.s.	10	21	31
k.s.	15	17	32
a.s.	18	21	39
OSVČ	0	6	6
Vš. skup.	69	84	153

Zdroj: vlastní zpracování v softwaru STATISTICA

Tabulka 37: Koeficienty závislosti – DPH (ordin. proměn.)

Statist.	Statist.: Právní forma (5) x DPH (2)		
	Chí-kvadr.	sv	p
Kendall. tau b & c	b=0,1150097	c=0,1413132	
Somers. D(X/Y), D(Y/X)	X/Y=0,14268	Y/X=0,09270	
Goodman.-Kruskal. gama	0,1854676		
Spearmanovo poř. R	0,1268056	t=1,5709	p=0,11830

Zdroj: vlastní zpracování v softwaru STATISTICA

Z tabulky je znatelné, že se hodnota Spearmanova koeficientu = 0,1268 přibližuje k 0, testová statistika pro Spearmanův koeficient $t = 1,5709$ a p – value pro testovou statistiku = **0,1183** > $\alpha = 0,05$, proto nulovost koeficientu nelze zamítnout. Ostatní koeficienty jsou rovněž blízké k 0, lze konstatovat, že mezi právní formou společnosti a plátcem DPH je nezávislost.

Závěr: Na základě výpočtů tedy nelze zamítnout hypotézu, že právní forma nezávisí na tom, zda je plátce či neplátce DPH. Tuto analýzu potvrdil Spearmanův koeficient blízký k 0, ostatní koeficienty taktéž blízké k 0 a hodnota p – value pro Spearmanův koeficient, která je prokazatelně vyšší než stanovená hladina významnosti α . Tato analýza vyvrátila výsledky analýzy pro nominální proměnné, jelikož hodnota p – value byla jen o minimální hodnotu nižší než stanovená hladina významnosti α , takže už se k potvrzení nezávislosti blížila.

Kruskal-Wallisův test

Je stanovena H_0 : všechny mediány jsou stejné a H_1 : non H_0 . Výsledky jsou uvedeny v následujících tabulkách.

Tabulka 38: Mediánový test - DPH

Závislá: Právní forma	Mediánový test, celk. medián = 103; Právní forma Nezávislá (grupovací) proměnná: DPH Chi-Kvadr. = 0,6691770 sv = 1 p = 0,4133		
	plátce	neplátce	Celkem
<= Medián: pozorov.	51	57	108
očekáv.	48,70588	59,29412	
poz.-oč.	2,29412	-2,29412	
> Medián: pozorov.	18	27	45
očekáv.	20,29412	24,70588	
poz.-oč.	-2,29412	2,29412	
Celkem: oček.	69	84	153

Zdroj: vlastní zpracování v softwaru STATISTICA

Tabulka 39 Kruskal-Wallisův test – DPH

Závislá: Právní forma	Kruskal-Wallisova ANOVA založ. na poř.: Právní forma Nezávislá (grupovací) proměnná: DPH Kruskal-Wallisův test: $H(1, N=153) = 2,444110 p = 0,4133$			
	Kód	Počet platných	Součet pořadí	Prům. Pořadí
plátce	101	69	4899,5	71,00725
neplátce	102	84	6881,5	81,92262

Zdroj: vlastní zpracování v softwaru STATISTICA

Porovnání: p – value pro Wallisův test = 0,4133 > $\alpha = 0,05$.

Závěr: Hodnota p – value pro Kruskal-Wallisův test výrazně převyšuje stanovenou hladinu významnosti α , proto není zamítnuta nulová hypotéza, že jsou všechny mediány stejné.

Nyní se prověří nezávislost základního kapitálu u akciové společnosti, komanditní společnosti a u společnosti s ručením omezeným. Data jsou uvažována jako ordinální a H_0 : nezávislost a H_1 : závislost. Opět se vytvoří kontingenční tabulka a tabulka pro koeficienty nezávislosti.

Tabulka 40: Kontingenční tabulka – Základní kapitál (ordin. proměn.)

Kontingenční tabulka					
Tab.:					
Právní forma	Základní kapitál - 1 000 000	Základní kapitál 1 000 001 – 5 000 000	Základní kapitál 5 000 001 – 10 000 000	Základní kapitál 10 000 000 +	Řádk. součty
s.r.o.	29	5	5	5	44
a.s.	5	12	3	12	32
k.s.	5	35	3	1	45
Vš. skup.	40	52	11	18	121

Zdroj: vlastní zpracování v softwaru STATISTIC

Tabulka 41: Koeficienty závislosti – Základní kapitál (ordin. proměn.)

Statist.	Statist.: Právní forma (3) x Základní kapitál (4)		
	Chí-kvadr.	sv	p
Kendall. tau b & c	b=0,1907239	c=0,1909706	
Somers. D(X/Y), D(Y/X)	X/Y=0,18843	Y/X=0,19304	
Goodman.-Kruskal. gama	0,2513484		
Spearmanovo poř. R	0,2389259	t=2,6841	p=0,00831

Zdroj: vlastní zpracování v softwaru STATISTICA

Z tabulky je viditelné, že hodnota Spearmanova koeficientu = 0,2389259, testová statistika pro Spearmanův koeficient $t = 2,6841$ a **hodnota p – value pro testovou statistiku = 0,00831** $< \alpha = 0,05$, proto se nulovost koeficientu zamítá. Ostatní koeficienty jsou rovněž větší než 0.

Závěr: Z výsledků je prokazatelné, že se zamítá hypotéza o nezávislosti. Tuto analýzu potvrdil Spearmanův koeficient větší než 0 a hodnota p – value pro Spearmanův

koeficient je prokazatelně nižší než stanovená hladina významnosti α . Závislost potvrdily i ostatní koeficienty, jejichž hodnoty jsou vyšší než 0.

Kruskal-Wallisův test

Testuje se H_0 : všechny mediány jsou stejné a H_1 : non H_0 . Přehled výsledků je uveden v následujících výstupních tabulkách.

Tabulka 42: Mediánový test – Základní kapitál

Závislá: Právní forma	Mediánový test, celk. medián = 103; Právní forma Nezávislá (grupovací) proměnná: Základní kapitál Chi-Kvadr. = 18,89542 sv = 3 p = 0,0003				
	- 1 000 000	1 000 001 – 5 000 000	5 000 001 – 10 000 000	10 000 000 +	Celkem
<= Medián: pozorov. očekáv.	34 28,6	40 38,13333	8 8,06667	6 13,2	88
poz.-oč.	5,4	1,86667	-0,06667	-7,2	
> Medián: pozorov. očekáv.	5 10,4	12 13,86667	3 2,93333	12 4,8	32
poz.-oč.	-5,4	-1,86667	0,06667	7,2	
Celkem: oček.	39	52	11	18	120

Zdroj: vlastní zpracování v softwaru STATISTICA

Tabulka 43: Kruskal-Wallisův test – Základní kapitál

Závislá: Právní forma	Kruskal-Wallisova ANOVA založ. na poř.: Právní forma Nezávislá (grupovací) proměnná: Základní kapitál Kruskal-Wallisův test: H (3, N=121) = 42,8373 p= 0,0000			
	Kód	Počet platných	Součet pořadí	Prům. Pořadí
- 1 000 000	101	40	1549	38,725
1 000 001 – 5 000 000	102	52	4303,5	82,75962
5 000 001 – 10 000 000	103	11	591	53,72727
10 000 000 +	104	18	937,5	52,08333

Zdroj: vlastní zpracování v softwaru STATISTICA

Porovnání: p – value pro Wallisův test = 0,0000 < α = 0,05.

Závěr: Hodnota p – value pro tento test je rovna 0, proto je zamítnuta nulová hypotéza, že jsou všechny mediány stejné. Je tedy přijata alternativní hypotéza, že se aspoň jeden medián v tabulce liší.

V poslední řadě se analyzuje závislost právní formy společnosti na počtu společníků. Opět jsou tyto data chápána jako ordinální a přehled dat znázorňuje následující kontingenční tabulka a výsledky tabulka koeficientů.

Tabulka 44: Kontingenční tabulka – Společníci (ordin. proměn.)

Kontingenční tabulka					
Tab.:					
Právní forma	Společníci 1 - 2	Společníci 3 - 5	Společníci 6 - 10	Společníci 11 +	Řádk. součty
s.r.o.	5	10	17	1	33
v.o.s.	16	5	4	6	31
a.s.	28	9	5	2	44
k.s.	14	18	6	1	39
Vš. skup.	63	42	32	10	147

Zdroj: vlastní zpracování v softwaru STATISTICA

Tabulka 45: Koeficienty závislosti – Společníci (ordin. proměn.)

Statist.	Statist.: Právní forma (4) x Společníci (4)		
	Chí-kvadr.	sv	p
Kendall. tau b & c	b=-0,177002	c=-0,168325	
Somers. D(X/Y), D(Y/X)	X/Y=-0,1849	Y/X=-0,1694	
Goodman.-Kruskal. gama	-0,238378		
Spearmanovo poř. R	-0,227337	t=-2,811	p=0,00562

Zdroj: vlastní zpracování v softwaru STATISTICA

Z tabulky lze pozorovat, že hodnota Spearmanova koeficientu = $-0,227337$, testová statistika pro Spearmanův koeficient $t = -2,811$ a **hodnota p – value pro testovou statistiku = $0,00562 < \alpha = 0,05$** , proto se nulovost koeficientu zamítá. Ostatní koeficienty jsou rovněž větší než 0, a proto lze konstatovat, že mezi právní formou společnosti a počtem společníků je závislost.

Závěr: Na základě koeficientů, p – value pro Spearmanův koeficient je jednoznačné, že mezi proměnnými je závislost, jelikož je zamítnuta nulová hypotéza na stanovené hladině významnosti α . Lze říci, že právní forma závisí na tom, kolik má společnost společníků.

Kruskal-Wallisův test

Je stanovena H_0 : všechny mediány jsou stejné a H_1 : non H_0 . Přehled výsledků nabízí výstupy ze softwaru v následujících tabulkách.

Tabulka 46: Mediánový test – Společníci

Závislá: Právní forma	Mediánový test, celk. medián = 103; Právní forma Nezávislá (grupovací) proměnná: Společníci Chi-Kvadr. = 11,36808 sv = 3 p = 0,0099				
	1-2	3-5	6-10	11+	Celkem
<= Medián: pozorov. očekáv.	35	33	27	8	103
poz.-oč.	44,14286	29,42857	22,42177	7,00680	
> Medián: pozorov. očekáv.	-9,14286	3,57143	4,57823	0,99320	
poz.-oč.	28	9	5	2	44
<= Medián: pozorov. očekáv.	18,85714	12,57143	9,57823	2,99320	
poz.-oč.	9,14286	-3,57143	-4,57823	-0,99320	
Celkem: oček.	63	42	32	10	147

Zdroj: vlastní zpracování v softwaru STATISTICA

Tabulka 47: Kruskal-Wallisův test – Společníci

Závislá: Právní forma	Kruskal-Wallisova ANOVA založ. na poř.: Právní forma Nezávislá (grupovací) proměnná: Společníci Kruskal-Wallisův test: H (3, N=147) = 13,3604 p= 0,0039			
	Kód	Počet platných	Součet pořadí	Prům. Pořadí
1 - 2	101	63	5083	80,68254
3 - 5	102	42	3497,5	83,27381
6 - 10	103	32	1685,5	52,67188
11 +	104	10	612	61,2

Zdroj: vlastní zpracování v softwaru STATISTICA

Porovnání: p – value pro Wallisův test = 0,0039 < α = 0,05.

Závěr: Hodnota p – value je jednoznačně nižší než stanovená hladina významnosti α , a proto je zamítnuta nulová hypotéza, že všechny mediány v tabulce jsou stejné. Platí tedy alternativní hypotéza, že se alespoň jeden medián liší.

7 Závěr a diskuze

V úvodu práce byla popsána statistická kategoriální data a základní statistické metody pro jejich analýzu (tabulky četností, chí-kvadrát test, kontingenční tabulky, Kruskal-Wallisova ANOVA,...), které poskytly základní fundament pro pochopení zkoumané problematiky. Detailně jsou rozebrány jednotlivé analýzy, jejich nulové a alternativní hypotézy, testové statistiky, koeficienty závislosti apod. Dále je popsán software STATISTICA, který umožňuje zpracovat analýzu nejen kategoriálních dat. Tento software je srovnáván i s obdobnými softwary, které se objevují na trhu. V další kapitole je popsán podrobný postup pro vytvoření analýzy kategoriálních dat přímo ve sledovaném statistickém softwaru, který je doplněn obrázky pro snadnější pochopení a orientaci při jeho ovládní a zpracování výsledků.

Praktická práce byla rozdělena na dvě části. V první části bylo zjištěno, že druh prodané licence na DVD od společnosti Grand s.r.o. (buď EPIS či PROFESIS) závisí na měsíci prodeje. Z tabulek četností bylo vypořováno a potvrzeno, že počet prodaných licencí systému PROFESIS tvoří v lednu jednu čtvrtinu z celkového prodeje a počet prodaných licencí systému EPIS tvořil v únoru jednu osminu z celkového prodeje. V těchto měsících má společnost největší nárůst zákazníků, pravděpodobně z důvodu začátku kalendářního roku. Analýza očekávaných četností potvrdila, že společnost dosáhla takového prodeje, jaký sama očekávala.

V 2. části práce po zpracování dat z dotazníkového šetření bylo zjištěno, že na dotazník neodpovídal stejný počet společností podle druhu právní formy podnikání. Z jednotlivých analýz se podařilo zjistit, že právní forma společnosti nezávisí na tom, zda je společnost plátce či neplátce DPH. Dále bylo vypořováno, že právní forma závisí na výši základního kapitálu i na počtu společníků ve společnosti.

Software STATISTICA nabídl spoustu možností pro tuto analýzu. Podařilo se nastítnit statistické metody na typových příkladech z ekonomické oblasti. V softwaru STATISTICA bohužel chyběly některé koeficienty, které je možno pro analýzu kategoriálních dat využít. Pokud je nutné dělat ještě rozšířenější a detailnější analýzu, je třeba k výpočtům využít jiný statistický software (např. SSPS, který tyto možnosti nabízí). Práce s tímto softwarem byla jednoduchá, je v něm snadná orientace a velmi dobře se pracuje s výstupy. Velkou výhodou je tedy uživatelská přívětivost, a dále že je

program v českém jazyce, což většina statistických aplikací není a lze s nimi pracovat pouze v anglickém jazyce.

I Summary

STATISTICA is analytical software containing data management resources, their analysis and development of user applications. It provides wide range of basic and advanced techniques developed especially for Business Intelligence, Big Data, quality management, research and other spheres.

This thesis involves primarily advantages and disadvantages of using the software to other similar softwares, focusing on analysis categorical data and representation of selected statistical methods (Pivot table, chi-square test).

The main purpose of this thesis is demonstration of options the software offers and the main focus is put on possibility of processing categorical data. Typical examples are shown whether the software can be used in practice as well as in accounting, economics and tax issues. All those information come from questionnaire and are processed in my thesis.

Key words: software STATISTICA, categorial data, statistical methods (Pivot table, chi-square test)

II Citovaná literatura

- Agresti, A. (2002). *Categorical data analysis*. 2nd ed. Hoboken: Wiley-interscience, New Jersey: Wiley series in probability and statistics.
- Binomický test*. (nedatováno). Načteno z *Základy informatiky pro biologii*: <http://portal.matematickabiologie.cz/index.php?pg=zaklady-informatiky-pro-biology--databazove-systemy-v-biomedicine--analyticke-a-statisticke-funkce-sql--statisticke-funkce--binomicky-test>
- Borůvková, J., Horáčková, P., & Hanáček, M. (2013). *STATISTICA: úvod do zpracování dat*. Jihlava, Czech: Vysoká škola polytechnická Jihlava.
- Everitt, B. (1994). *The analysis of contingency tables*. London: Chapman & Hall, England: Monographs on statistics and applied probability.
- Hendl, J. (2009). *Přehled statistických metod*. Praha, Czech: nakladatelství Portál, s.r.o.
- Matoušek, J. (2. duben 2010). *Statistický software - jak na výběr statistického software*. Načteno z Datamind: <http://www.datamind.cz/cz/blog/Statisticky-software-jak-na-vyber-statistickeho-software>
- Nová verze softwaru Dell Statistica pomůže s lepší přípravou dat*. (22. duben 2016). Načteno z Dell aktualizuje analytický software Statistica: http://www.statsoft.cz/file1/PDF/Dell_aktualizuje_analytický_software_Statistica.pdf
- Profil společnosti*. (nedatováno). Načteno z O firmě: <http://www.statsoft.cz/o-firme/profil-spolecnosti/>
- Řezanková, H. (2005). *Analýza kategoriálních dat*. Praha, Czech: Vysoká škola ekonomická v Praze.
- Řezanková, H., & Hronová, S. (2000). *Statistická data*. Praha, Czech: Vysoká škola ekonomická v Praze.

III Seznam tabulek

Tabulka 1: Schéma tabulky četností.....	5
Tabulka 2: Schéma kontingenční tabulky pro absolutní četnosti.....	9
Tabulka 3: Schéma kontingenční tabulky pro relativní četnosti.....	10
Tabulka 4: Příprava dat pro kontingenční tabulku.....	19
Tabulka 5: Počet prodaných licencí PROFESIS za rok 2014.....	24
Tabulka 6: Tabulka četností – Licence PROFESIS za rok 2014.....	24
Tabulka 7: Počet prodaných Licencí EPIS za rok 2014.....	26
Tabulka 8: Tabulka četností – Licence EPIS za rok 2014.....	26
Tabulka 9: Pozorované a očekávané četnosti – Licence EPIS za rok 2014.....	28
Tabulka 10: Chí-kvadrát test dobré shody pro shodné očekávané četnosti – Licence EPIS za rok 2014.....	29
Tabulka 11: Pozorované a očekávané četnosti 2 – Licence EPIS za rok 2014 ...	30
Tabulka 12: Chí-kvadrát test dobré shody pro různé očekávané četnosti – Licence EPIS za rok 2014.....	31
Tabulka 13: Počet prodaných licencí systému EPIS – červenec a září.....	32
Tabulka 14: Test dobré shody – Měsíc červenec a září.....	32
Tabulka 15: Kontingenční tabulka – Prodané licence za rok 2014.....	33
Tabulka 16: Očekávané četnosti – Prodané licence za rok 2014.....	34
Tabulka 17: Koeficienty závislosti – Podané licence za rok 2014.....	34
Tabulka 18: Relativní četnosti – Právní forma.....	36
Tabulka 19: Pozorované a očekávané četnosti – Právní forma.....	36
Tabulka 20: Chí-kvadrát test dobré shody pro shodné očekávané četnosti – Právní forma.....	37
Tabulka 21: Pozorované a očekávané četnosti – Právní forma (bez OSVČ).....	37
Tabulka 22: Chí-kvadrát test dobré shody pro shodné očekávané četnosti – právní forma (bez OSVČ).....	38
Tabulka 23: Pozorované a očekávané četnosti – DPH.....	38
Tabulka 24: Chí-kvadrát test – DPH.....	39
Tabulka 25: Pozorované a očekávané četnosti – evidence.....	39
Tabulka 26: Chí-kvadrát test – evidence.....	40
Tabulka 27: Kontingenční tabulka – DPH (nomin. proměn.).....	41
Tabulka 28: Očekávané četnosti – DPH.....	41

Tabulka 29: Koeficienty závislosti – DPH (nomin. proměn.).....	42
Tabulka 30: Kontingenční tabulka – Základní kapitál (nomin. proměn.)	43
Tabulka 31: Očekávané četnosti – Základní kapitál.....	43
Tabulka 32: Koeficienty závislosti – Základní kapitál (nomin. proměn.).....	44
Tabulka 33: Kontingenční tabulka – Společníci (nomin. proměn.)	44
Tabulka 34: Očekávané četnosti – Společníci.....	45
Tabulka 35: Koeficienty závislosti – Společníci (nomin. proměn.).....	45
Tabulka 36: Kontingenční tabulka – DPH (ordin. proměn.).....	46
Tabulka 37: Koeficienty závislosti – DPH (ordin. proměn.).....	46
Tabulka 38: Mediánový test - DPH.....	47
Tabulka 39 Kruskal-Wallisův test – DPH.....	47
Tabulka 40: Kontingenční tabulka – Základní kapitál (ordin. proměn.).....	48
Tabulka 41: Koeficienty závislosti – Základní kapitál (ordin. proměn.).....	48
Tabulka 42: Mediánový test – Základní kapitál	49
Tabulka 43: Kruskal-Wallisův test – Základní kapitál.....	49
Tabulka 44: Kontingenční tabulka – Společníci (ordin. proměn.)	50
Tabulka 45: Koeficienty závislosti – Společníci (ordin. proměn.).....	50
Tabulka 46: Mediánový test – Společníci	51
Tabulka 47: Kruskal-Wallisův test – Společníci	51

IV Seznam obrázků

Obrázek 1: Distribuční funkce pro normální rozdělení - levostranná alternativa s kvantilem $u_{0,05}$	7
Obrázek 2: Distribuční funkce pro normální rozdělení – pravostranná alternativa s kvantilem $u_{1-0,05}$	7
Obrázek 3: Distribuční funkce pro normální rozdělení – pravostranná alternativa s kvantilem $u_{1-0,05/2}$	7
Obrázek 4: Tabulky četností v software STATISTICA	17
Obrázek 5: Pravděpodobnostní kalkulátor	17
Obrázek 6: Proměnné pro chí-kvadrát test	18
Obrázek 7: Proměnné u kontingenční tabulky	19
Obrázek 8: Možnosti kontingenčních tabulek	20
Obrázek 9: Neparametrická korelace	21
Obrázek 10: Kruskal-Wallisova ANOVA 1	21
Obrázek 11: Kruskal-Wallisova ANOVA 2	22
Obrázek 12: Kalkulátor pro binomické rozdělení	25
Obrázek 13: Kalkulátor pro normální rozdělení.....	27

V Seznam příloh

Příloha 1: Přehled firem – ZK, DPH, společníci, forma evidence

**Příloha č. 1: Přehled firem – ZK, DPH, společníci, forma evidence
(zpracovaná data z dotazníkového šetření)**

Název firmy	Právní forma	DPH	Základní kapitál	Forma evidence	Počet společníků
Šulc a syn a.s.	akciová společnost	plátce	2 000 000 Kč	účetnictví	1
G.A.P., a.s.	akciová společnost	neplátce	1 000 000 Kč	účetnictví	2
A.Z.A. Kyjov a.s.	akciová společnost	plátce	135 000 000 Kč	účetnictví	2
ADVISER-EURO, a.s.	akciová společnost	neplátce	1 100 000 Kč	účetnictví	2
V.R.A.T.J.A., a.s.	akciová společnost	plátce	2 000 000 Kč	účetnictví	2
RELAXA a.s.	akciová společnost	plátce	9 049 000 Kč	účetnictví	3
Real a.s.	akciová společnost	neplátce	10 000 000 Kč	účetnictví	3
STRIKE a.s.	akciová společnost	neplátce	1 000 000 Kč	účetnictví	4
A.A.S.I. a.s.	akciová společnost	plátce	2 000 000 Kč	účetnictví	4
A-Z PREZIP a.s.	akciová společnost	plátce	3 870 000 Kč	účetnictví	4
P.A.M. Holding, a.s.	akciová společnost	plátce	1 000 000 Kč	účetnictví	5
V-KMV, a.s.	akciová společnost	plátce	2 000 000 Kč	účetnictví	5
C - FAS GROUP a.s.	akciová společnost	neplátce	8 000 000 Kč	účetnictví	5
NELI S.P.A., a.s.	akciová společnost	plátce	14 465 000 Kč	účetnictví	5
LARA a.s.	akciová společnost	plátce	140 000 000 Kč	účetnictví	5
Úspěch 2000 Alfa a.s.	akciová společnost	neplátce	1 000 000 Kč	účetnictví	6
L.A. Consulting, a.s.	akciová společnost	neplátce	1 000 000 Kč	účetnictví	6
ZLATÝ LEV A&D a.s.	akciová společnost	plátce	2 000 000 Kč	účetnictví	6
INDUSTRY DESIGN a.s.	akciová společnost	neplátce	2 000 000 Kč	účetnictví	6
MAXIMA UNION a.s.	akciová společnost	neplátce	2 000 000 Kč	účetnictví	6
AUTOART a. s.	akciová společnost	plátce	21 360 000 Kč	účetnictví	6
Moody's Central Europe a.s.	akciová společnost	plátce	31 000 000 Kč	účetnictví	6
A - imme, a.s.	akciová společnost	plátce	2 000 000 Kč	účetnictví	7
Starducks a.s.	akciová společnost	neplátce	2 000 000 Kč	účetnictví	7
GTH CATERING a.s.	akciová společnost	neplátce	3 000 000 Kč	účetnictví	7
MEI Czech Offices, a.s.	akciová společnost	neplátce	12 000 000 Kč	účetnictví	7
AIRFIN PRAHA a.s.	akciová společnost	neplátce	72 450 000 Kč	účetnictví	7
Agropol Group, a.s.	akciová společnost	plátce	240 000 000 Kč	účetnictví	7
Plzeňská energetika a.s.	akciová společnost	neplátce	500 000 000 Kč	účetnictví	7

Skanska CZ a.s.	akciová společnost	neplátce	1 094 229 000 Kč	účetnictví	9
GTS NOVERA a.s.	akciová společnost	neplátce	5 474 340 000 Kč	účetnictví	10
ODS-Dopravní stavby Ostrava, a.s.	akciová společnost	neplátce	367 461 000 Kč	účetnictví	11
ADITUS S.I., k.s.	komanditní společnost	neplátce	30 000 Kč	účetnictví	2
Ústecký pivovar, k.s.	komanditní společnost	neplátce	70 000 Kč	účetnictví	2
AGROCOM k.s.	komanditní společnost	plátce	80 000 Kč	účetnictví	2
NIKOL, k.s.	komanditní společnost	neplátce	100 000 Kč	účetnictví	2
DaS Havířov k.s.	komanditní společnost	plátce	120 000 Kč	účetnictví	2
GRANIT PARTS, k.s.	komanditní společnost	plátce	150 000 Kč	účetnictví	2
Net 4E, k.s.	komanditní společnost	neplátce	150 000 Kč	účetnictví	2
ANACO, k.s.	komanditní společnost	plátce	200 000 Kč	účetnictví	2
KASYG, k.s.	komanditní společnost	plátce	220 000 Kč	účetnictví	2
Würz Bauklotz, k.s.	komanditní společnost	neplátce	230 000 Kč	účetnictví	2
EUPOZ, k.s.	komanditní společnost	neplátce	250 000 Kč	účetnictví	2
Orio k.s.	komanditní společnost	plátce	500 000 Kč	účetnictví	2
B - KONTAKT, k.s.	komanditní společnost	plátce	10 000 000 Kč	účetnictví	2
SAVANA, k.s.	komanditní společnost	neplátce	12 000 000 Kč	účetnictví	2
BALTEX, k.s.	komanditní společnost	neplátce	50 000 Kč	účetnictví	3
PANKO, k.s.	komanditní společnost	neplátce	50 000 Kč	účetnictví	3
BLUE ELEPHANTS k.s.	komanditní společnost	plátce	80 000 Kč	účetnictví	3
KREUZER PANGRAC, k.s.	komanditní společnost	neplátce	80 000 Kč	účetnictví	3
Pila Lhůta, k.s.	komanditní společnost	plátce	100 000 Kč	účetnictví	3
ARISE k.s.	komanditní společnost	plátce	100 000 Kč	účetnictví	3
DG.L.servis, k.s.	komanditní společnost	plátce	100 000 Kč	účetnictví	3
REK k.s.	komanditní společnost	neplátce	100 000 Kč	účetnictví	3

AUXILIA, k.s.	komanditní společnost	neplátce	150 000 Kč	účetnictví	3
KALMA, k.s.	komanditní společnost	plátce	200 000 Kč	účetnictví	3
AUTOCENTER, k.s.	komanditní společnost	plátce	250 000 Kč	účetnictví	3
FINVIA, k.s.	komanditní společnost	neplátce	2 000 000 Kč	účetnictví	3
I.J.M. Kolín, k.s.	komanditní společnost	neplátce	40 000 Kč	účetnictví	4
PLEHASO, k.s.	komanditní společnost	plátce	50 000 Kč	účetnictví	4
ALFA FLY k.s.	komanditní společnost	neplátce	60 000 Kč	účetnictví	4
DEMARKO k.s.	komanditní společnost	neplátce	100 000 Kč	účetnictví	4
KOMPAS, k.s.	komanditní společnost	plátce	700 000 Kč	účetnictví	4
DROMADER CZECH k.s.	komanditní společnost	plátce	6 000 000 Kč	účetnictví	5
JW Consulting Group, k.s.	komanditní společnost	plátce	17 000 Kč	účetnictví	6
Hajný a spol., k.s.	komanditní společnost	neplátce	300 000 Kč	účetnictví	6
GAL, k.s.	komanditní společnost	plátce	450 000 Kč	účetnictví	6
Velkobřezenský pivovar, k.s.	komanditní společnost	neplátce	200 000 Kč	účetnictví	8
ISMAJLER, k.s.	komanditní společnost	neplátce	300 000 Kč	účetnictví	8
NAKOV, k.s.	komanditní společnost	neplátce	500 000 Kč	účetnictví	8
WOC Praha Management, k.s.	komanditní společnost	neplátce	120 000 Kč	účetnictví	11
Agata Romanska	podnikatel - FO	neplátce	-	daňová evidence	-
Ing. Demeter Sulima	podnikatel - FO	neplátce	-	daňová evidence	-
Ing. Pavel Abrahám - AB TRADING	podnikatel - FO	neplátce	-	daňová evidence	-
Klára Šedivá - ATTON MUSIC arts management	podnikatel - FO	neplátce	-	daňová evidence	-
Milan Černý - realitní kancelář "ARCHA"	podnikatel - FO	neplátce	-	daňová evidence	-

Nataliia Leshchenko	podnikatel - FO	neplátce	-	daňová evidence	-
Adal a s.r.o.	společnost s ručením omezeným	neplátce	1 000 Kč	účetnictví	1
TAMAR s.r.o.	společnost s ručením omezeným	plátce	100 000 Kč	účetnictví	1
Arris, s.r.o.	společnost s ručením omezeným	plátce	100 000 Kč	účetnictví	1
FURALO spol. s r.o.	společnost s ručením omezeným	neplátce	100 000 Kč	účetnictví	1
NORMA SaS, s.r.o.	společnost s ručením omezeným	neplátce	100 000 Kč	účetnictví	1
GLT LIBEREC s.r.o.	společnost s ručením omezeným	neplátce	100 000 Kč	účetnictví	1
Atlas Care s.r.o.	společnost s ručením omezeným	plátce	200 000 Kč	účetnictví	1
KGP projekt s.r.o.	společnost s ručením omezeným	plátce	200 000 Kč	účetnictví	1
OB trade s.r.o.	společnost s ručením omezeným	plátce	200 000 Kč	účetnictví	1
OEPD s.r.o.	společnost s ručením omezeným	plátce	200 000 Kč	účetnictví	1
GYNBUS s.r.o.	společnost s ručením omezeným	neplátce	200 000 Kč	účetnictví	1
KRISTENSSON, s r.o.	společnost s ručením omezeným	plátce	300 000 Kč	účetnictví	1
KOMEST, s.r.o.	společnost s ručením omezeným	plátce	500 000 Kč	účetnictví	1
Amplua Company s.r.o.	společnost s ručením omezeným	neplátce	1 000 000 Kč	účetnictví	1
ADATA s.r.o.	společnost s ručením omezeným	neplátce	1 200 000 Kč	účetnictví	1
Hyposervis Partner s.r.o.	společnost s ručením omezeným	plátce	1 200 000 Kč	účetnictví	1

Ivana Halvová s.r.o.	společnost s ručením omezeným	neplátce	6 000 000 Kč	účetnictví	1
MinCIS, s.r.o.	společnost s ručením omezeným	plátce	6 500 000 Kč	účetnictví	1
MONOTEC, s.r.o.	společnost s ručením omezeným	plátce	7 000 000 Kč	účetnictví	1
JARMONT spol. s r.o.	společnost s ručením omezeným	plátce	100 000 Kč	účetnictví	2
ALKOM, s.r.o.	společnost s ručením omezeným	plátce	100 000 Kč	účetnictví	2
DISTRIB spol.s r.o.	společnost s ručením omezeným	neplátce	120 000 Kč	účetnictví	2
ERBA-PET s. r.o.	společnost s ručením omezeným	neplátce	120 000 Kč	účetnictví	2
DESIGNLAND Company s.r.o.	společnost s ručením omezeným	neplátce	200 000 Kč	účetnictví	2
MAV Trading s.r.o.	společnost s ručením omezeným	plátce	1 100 000 Kč	účetnictví	2
SPESSART s.r.o.	společnost s ručením omezeným	neplátce	1 500 000 Kč	účetnictví	2
CREDITFIELD, s.r.o.	společnost s ručením omezeným	neplátce	2 000 000 Kč	účetnictví	2
IRS Company Praha, s.r.o.	společnost s ručením omezeným	neplátce	10 000 000 Kč	účetnictví	2
FALA, s.r.o.	společnost s ručením omezeným	plátce	16 730 000 Kč	účetnictví	2
ROMAGRIS, s.r.o.	společnost s ručením omezeným	plátce	100 000 Kč	účetnictví	3
KOVPAL DOBRUŠKA, s.r.o.	společnost s ručením omezeným	plátce	100 000 Kč	účetnictví	3
A. Perola s.r.o.	společnost s ručením omezeným	neplátce	100 000 Kč	účetnictví	3
AURELIA, spol. s r.o.	společnost s ručením omezeným	plátce	105 000 Kč	účetnictví	3

TRESKO s.r.o.	společnost s ručením omezeným	plátce	115 000 Kč	účetnictví	3
ALASIA s.r.o.	společnost s ručením omezeným	plátce	200 000 Kč	účetnictví	3
R.O.S. Vacek s.r.o.	společnost s ručením omezeným	plátce	5 500 000 Kč	účetnictví	3
Siemens Nízkonapěťová spínací technika s. r. o.	společnost s ručením omezeným	neplátce	134 500 000 Kč	účetnictví	4
Siemens Elektromotory s.r.o.	společnost s ručením omezeným	neplátce	800 000 000 Kč	účetnictví	5
R. T. S. cs, spol. s r. o.	společnost s ručením omezeným	plátce	400 000 Kč	účetnictví	6
Sklo servis, s.r.o.	společnost s ručením omezeným	plátce	1 000 000 Kč	účetnictví	6
SK Chomutov S-R-O s.r.o.	společnost s ručením omezeným	plátce	100 000 Kč	účetnictví	6
Lucas Varity s.r.o.	společnost s ručením omezeným	neplátce	570 100 000 Kč	účetnictví	6
JURSERVIS, s.r.o.	společnost s ručením omezeným	plátce	100 000 Kč	účetnictví	7
S.Y.B. Praha CZ, s.r.o.	společnost s ručením omezeným	neplátce	20 000 000 Kč	účetnictví	11
BECKER s.r.o.	společnost s ručením omezeným	plátce	25 100 000 Kč	účetnictví	13
CP - UNI, v.o.s.	veřejná obchodní společnost	plátce	-	účetnictví	2
Mrázek, v.o.s.	veřejná obchodní společnost	plátce	-	účetnictví	2
STAS, v.o.s.	veřejná obchodní společnost	plátce	-	účetnictví	2
CP - UNI, v.o.s.	veřejná obchodní společnost	plátce	-	účetnictví	2
TERAGON Group v.o.s.	veřejná obchodní společnost	plátce	-	účetnictví	2
IZAURA v.o.s.	veřejná obchodní společnost	plátce	-	účetnictví	2
ELEKTRO - uni, v.o.s.	veřejná obchodní společnost	plátce	-	účetnictví	2

V.V.Royal, v.o.s.	veřejná obchodní společnost	neplátce	-	účetnictví	2
J & J, dřevařské výrobky, v.o.s.	veřejná obchodní společnost	neplátce	-	účetnictví	2
OÁZA V v.o.s.	veřejná obchodní společnost	neplátce	-	účetnictví	2
H + V PRIMA v.o.s.	veřejná obchodní společnost	neplátce	-	účetnictví	2
SPORT CONNEX, v.o.s.	veřejná obchodní společnost	neplátce	-	účetnictví	2
V + S ELEKTRO v. o. s.	veřejná obchodní společnost	neplátce	-	účetnictví	2
V + J software, v.o.s.	veřejná obchodní společnost	neplátce	-	účetnictví	2
V & Š, v.o.s.	veřejná obchodní společnost	neplátce	-	účetnictví	2
JAKASPOJ, v.o.s	veřejná obchodní společnost	neplátce	-	účetnictví	2
Foiti řeznictví a uzenářství v.o.s.	veřejná obchodní společnost	plátce	-	účetnictví	3
K O V I G, v.o.s.	veřejná obchodní společnost	neplátce	-	účetnictví	3
AQ invest, v.o.s.	veřejná obchodní společnost	neplátce	-	účetnictví	4
V.P.V. KASTEX, v.o.s.	veřejná obchodní společnost	neplátce	-	účetnictví	4
Revitax v.o.s.	veřejná obchodní společnost	neplátce	-	účetnictví	4
O.K. reklama v.o.s.	veřejná obchodní společnost	plátce	-	účetnictví	6
Finareal, v.o.s.	veřejná obchodní společnost	plátce	-	účetnictví	6
JUKIMA v.o.s.	veřejná obchodní společnost	neplátce	-	účetnictví	7
SMULEAC v.o.s.	veřejná obchodní společnost	neplátce	-	účetnictví	10
ARTEMIDA v.o.s.	veřejná obchodní společnost	neplátce	-	účetnictví	16
STARIN import, v.o.s.	veřejná obchodní společnost	neplátce	-	účetnictví	18
FEYER a spol.v.o.s.	veřejná obchodní společnost	neplátce	-	účetnictví	20
ZLUKA import v.o.s.	veřejná obchodní společnost	neplátce	-	účetnictví	22
NIKA I.V.T. v.o.s.	veřejná obchodní společnost	neplátce	-	účetnictví	28
NAPOLEON, v.o.s.	veřejná obchodní společnost	neplátce	-	účetnictví	50