

**Univerzita Hradec Králové**  
**Fakulta informatiky a managementu**  
**Katedra informatiky a kvantitativních metod**

**Statistické metody pro zpracování dotazníku**  
Bakalářská práce

Autor: Jakub Horáček  
Studijní obor: Informační management - prezenční

Vedoucí práce: doc. RNDr. Pavel Pražák, Ph.D.

Prohlášení:

Prohlašuji, že jsem bakalářskou práci zpracoval samostatně a s použitím uvedené literatury.

V Hradci Králové dne 14.8.2020

Jakub Horáček

Poděkování:

Děkuji vedoucímu bakalářské práce doc. RNDr. Pavlu Pražákovi, Ph.D za metodické vedení práce, rady a připomínky.

## **Anotace**

Předmětem této bakalářské práce je popsat základní statistické metody pro zpracování dotazníku a na následném problému ukázat jejich použití a interpretaci výsledků. Teoretická část práce se nejprve zabývá principy správné konstrukce dotazníku, následně popisuje převedení výsledků šetření do datového souboru a zaměřuje se na popisnou statistiku a analýzu závislostí nominálních a ordinálních proměnných. Dále je popsána a vysvětlena orientace v programu SPSS. V praktické části je vytvořen dotazník a následně popsáno provedení dotazníkového šetření. Dále je ukázáno převedení získaných dat do programu SPSS a potom vybráno několik proměnných za účelem ukázky analýzy závislostí a provedení popisu proměnných. V závěru jsou uvedeny přínosy práce a problémy, které nastaly.

## **Annotation**

### **Title: Statistical methods in the questionnaire analysis**

The subject of this bachelor thesis is to describe the basic statistical methods for processing the questionnaire and to show their use and interpretation of results on the subsequent problem. The theoretical part of the thesis first deals with the principles of correct construction of the questionnaire, then describes the conversion of survey results into a data set and focuses on descriptive statistics and analysis of the dependencies of nominal and ordinal variables. The orientation in the SPSS program is also described and explained. In the practical part, a questionnaire is created and then the implementation of the questionnaire survey is described. Next, the conversion of the obtained data into the SPSS program is shown and then several variables are selected in order to demonstrate the analysis of dependencies and perform the description of variables. In the end, the benefits of the work and the problems that occurred are presented.

# Obsah

1	Úvod.....	1
2	Pořizování dat.....	2
2.1	Principy konstrukce dotazníku.....	2
2.1.1	Uzavřené otázky.....	3
2.1.2	Otevřené otázky .....	4
2.1.3	Škálové otázky .....	4
2.2	Požadované vlastnosti dobrého dotazníku .....	5
2.2.1	Objektivnost.....	5
2.2.2	Standardnost .....	6
2.2.3	Reliabilita.....	6
2.2.4	Validita.....	6
2.2.5	Úspornost .....	7
2.2.6	Míra návratnosti.....	7
3	Vytvoření datového souboru .....	9
3.1	Škály měření a typy proměnných .....	9
4	Analýza jednotlivých proměnných .....	11
4.1	Rozdělení četností.....	11
4.2	Popisné charakteristiky .....	12
4.2.1	Popisné charakteristiky nominální proměnné .....	12
4.2.2	Popisné charakteristiky ordinální proměnné .....	13
4.2.3	Popisné charakteristiky kvantitativní proměnné .....	15
4.2.4	Grafické rozdělení četností.....	15
5	Analýza závislostí.....	17
5.1	Statistické hypotézy .....	17
5.2	Hladina významnosti .....	18

5.3	Dvourozměrné rozdělení četností .....	18
5.4	Princip zjišťování závislostí dvou proměnných .....	19
5.5	Tabulka pro dvě nominální proměnné .....	20
5.5.1	Chí-kvadrát test o nezávislosti .....	20
5.5.2	Pearsonův koeficient kontingence.....	21
5.5.3	Cramérovo V .....	21
5.6	Tabulka pro dvě ordinální proměnné .....	22
5.6.1	Spearmanův koeficient pořadové korelace .....	23
5.6.2	Kendallův koeficient pořadové korelace .....	24
6	Software SPSS .....	25
6.1	Rozhraní programu .....	25
6.1.1	Vložení dat do SPSS.....	26
6.1.2	Vytvoření četnostní tabulky .....	27
6.1.3	Vytvoření kontingenční tabulky .....	28
6.1.4	Překódování proměnných.....	29
7	Dotazník pro studenty předmětu ZMI1 .....	31
7.1	Dotazníkové šetření .....	31
7.2	Vložení dat do souboru SPSS .....	32
8	Popisná statistika a analýza závislostí v SPSS.....	35
8.1	Popisná statistika nominálních proměnných.....	35
8.1.1	Proměnná D1 – Vystudovaná střední škola .....	36
8.1.2	Proměnná D2 – Maturita z matematiky.....	37
8.2	Popisná statistika ordinálních proměnných.....	38
8.2.1	Proměnná O1 – Náročnost předmětu ZMI1 .....	39
8.2.2	Proměnná O4 – Účast studenta na přednáškách.....	40

8.3	Analýza závislostí s nominální proměnnou – závislost maturity z matematiky a úspěšnosti v zápočtovém testu.....	42
8.4	Analýza závislostí ordinálních proměnných – korelace mezi studenty využívající poznámky z přednášek a ze cvičení .....	44
9	Závěr.....	47
10	Seznam použité literatury .....	49
11	Přílohy.....	50

## Seznam obrázků

Obrázek č. 2 - Navigace pro vytvoření datového souboru .....	26
Obrázek č. 3 - Kódování možností odpovědí.....	27
Obrázek č. 4 - Navigace pro vytvoření četnostní tabulky.....	28
Obrázek č. 5 - Možnosti popisné statistiky pro četnostní tabulku.....	28
Obrázek č. 6 - Navigace pro vytvoření kontingenční tabulky .....	29
Obrázek č. 7 - Okno výběru možností pro kontingenční tabulku.....	29
Obrázek č. 8 - Variable View .....	32
Obrázek č. 9 - Data View.....	34

## Seznam tabulek

Tabulka č. 1 - Akceptovatelná a doporučená míra návratnosti. ....	8
Tabulka č. 2 - Symbolika pro tabulku rozdělení četností .....	11
Tabulka č. 3 - Značení pro kontingenční tabulku absolutních četností .....	19
Tabulka č. 4 - Přibližná interpretace hodnot korelačního koeficientu.....	23
Tabulka č. 5 - Kódování otázek.....	33
Tabulka č. 6 - Četnostní tabulka proměnné D1 .....	36
Tabulka č. 7 - pomocná tabulka k proměnné D1 .....	37
Tabulka č. 8 - Četnostní tabulka proměnné D2 .....	37
Tabulka č. 9 - Pomocná tabulka k proměnné D2 .....	38
Tabulka č. 10 - Četnostní tabulka proměnné O1.....	39
Tabulka č. 11 - Pomocná tabulka k proměnné O1.....	40
Tabulka č. 12 - Četnostní tabulka proměnné O4.....	40
Tabulka č. 13 - Pomocná tabulka k proměnné O4.....	41
Tabulka č. 14 - Kontingenční tabulka proměnné D2 a D3 .....	42
Tabulka č. 15 - Kontingenční tabulka proměnné D2 a D3 - očekávaná četnost.....	42
Tabulka č. 16 - Kontingenční tabulka proměnné D2 a D3 (překódováno).....	43
Tabulka č. 17 - Kontingenční tabulka proměnné D2 a D3 (překódováno) - očekávaná četnost .....	43
Tabulka č. 18 - Výsledek testu nezávislosti.....	43
Tabulka č. 19 - Výpočty kontingenčních koeficientů .....	44



Tabulka č. 20 - Kontingenční tabulka proměnných O10C a O10D.....	45
Tabulka č. 21 - Výsledky korelačních koeficientů .....	45

## **Seznam grafů**

Graf č. 1 - Četnosti proměnné D1 .....	36
Graf č. 2 - Četnosti proměnné D2 .....	38
Graf č. 3 - Četnosti proměnné O1 .....	39
Graf č. 4 - Četnosti proměnné O4 .....	41

# 1 Úvod

Tato bakalářská práce se zabývá vybranými statistickými metodami pro zpracování dotazníku. Jednou z nejužívanějších metod kvantitativních výzkumů jsou právě dotazníková šetření. Aby bylo možné na praktickém problému ukázat použití těchto vybraných metod, rozhodl se autor vytvořit a zpracovat dotazník týkající se zpětné vazby studentů prvního ročníku Fakulty informatiky a managementu na předmět Základy matematiky pro informatiky I (ZMI1).

Cílem práce je popsat a na praktickém problému ukázat použití vybraných statistických metod při zpracování dotazníku. Zaměřuje se především na popisnou statistiku nominálních a ordinálních dat a následně se zabývá analýzou závislosti těchto dat.

V teoretické části práce jsou popsány principy konstrukce dotazníku, dále uvedeny vybrané statistické metody pro zpracování dotazníku a popsány základy softwaru SPSS. V praktické části od kapitoly 7 jsou interpretovány výsledky provedeného šetření pomocí uvedených statistických metod na několika vybraných proměnných. Na závěr jsou popsány problémy, které se při práci vyskytly a přínosy práce.

## 2 Pořizování dat

Základními metodami sběru dat jsou podle Pecákové (2008) pozorování a dotazování. Chráska (2016) uvádí, že dotazování se rozděluje na interview a dotazník, který můžeme vymezit podle P. Gavora (2000) jako „*způsob písemného kladení otázek a získávání písemných odpovědí*“.

Při pořizování dat je třeba se rozhodnout, jak rozsáhlou skupinu lidí budeme šetřit a jakým způsobem. V oblasti sociologických výzkumů se považuje konkrétní osoba za nositele primárních informací. Tyto konkrétní osoby potom spadají pod značně rozsáhlou skupinu osob, která se nazývá základní soubor (upraveno, Pecáková (2008)). Pecáková (2008) však uvádí, že prošetření celého základního souboru je nákladné, vyčerpávající a ve spoustě případech i nemožné. Proto se velmi často používá výběrový soubor, při kterém vybereme jen část základního souboru, a to buď náhodným výběrem nebo jiným způsobem (například skupinovým či oblastním výběrem). Pecáková (2008) dále upozorňuje, že pokud chceme posudky získané o výběrovém souboru zobecnit na základní soubor, je třeba aby byly tyto soubory svým věrným obrazem, jinak řečeno chceme, aby byly soubory reprezentativní. Podle Pecákové (2008) můžeme reprezentativnost chápat i tak, že *„vzorek musí být tak velký, aby umožňoval provádění dostatečně přesných a spolehlivých odhadů populačních parametrů a aby mohl reagovat na hypotézy vyslovené o populaci s přijatelně malými pravděpodobnostmi chyb“*. Pokud se rozhodneme některé jednotky pominout (například z organizačních důvodů), je třeba si tuto skutečnost uvědomit a zohlednit ji při zobecňování.

### 2.1 Principy konstrukce dotazníku

Jak uvádí Řezanková (2007), pro každé šetření musí být formulované nějaké cíle. Jako cíl uvádí například zjištění hodnot, které si lidé nejvíce cení a zdali to závisí na vztahu k náboženství nebo třeba k sociálně ekonomickému postavení.

Podle Řezankové (2007) se mohou otázky rozdělovat do dvou základních skupin:

- otázky týkající se názorů a chování respondentů
- otázky za účelem získání jiných údajů, například demografických

První skupina otázek slouží k šetření samotného problému. Jedná se tedy o otázky, které nám poskytnou informace o postoji respondenta ke zkoumané problematice. Tyto otázky se nazývají meritorní. Druhá skupina otázek nám pomáhá respondenty třídit do různých kategorií. Může se jednat hlavně o otázky demografické nebo jiné třídící, které souvisí s problematikou (například „Kolik jste dostal/a bodů ze zápočtového testu“?). Toto jsou otázky analytické. Většinou je uvádíme uprostřed nebo na konci dotazníku (převzato Řezanková, 2007).

Kohoutek (2010) a Řezanková (2007) rozdělují dále otázky podle variant odpovědí:

- uzavřené
- otevřené
- škálové

### **2.1.1 Uzavřené otázky**

Jedná se o otázky, které nabízejí respondentovi nějaké varianty odpovědí. Podle Řezankové (2007) se tyto otázky rozdělují dále na alternativní a selektivní. Alternativní otázka nabízí jen dvě varianty odpovědí (například muž – žena, ano – ne), selektivní pak obsahuje větší množství variant, ze kterých si může respondent vybrat. Řezanková (2007) dále uvádí, že je třeba, aby všechny varianty odpovědí byly jednoznačné, nepřekrývaly se a také byly zahrnuty všechny možné odpovědi. Nemělo by se stát, že respondent zaškrtně odpověď, se kterou se neztotožňuje jen proto, že některá odpověď chybí. Proto se často do otázek zařazuje varianta typu „nevím“ či „jiné“ apod.

### 2.1.2 Otevřené otázky

Podle Pecákové (2008) se jedná o otázky, ve kterých respondent formuluje odpověď samostatně. Můžeme se pak respondenta přímo zeptat například na jeho/její měsíční příjem. Pecáková (2008) upozorňuje, že na takovou otázku nemusí být respondent ochoten odpovědět. Proto je v těchto případech vhodné použít kategorizaci, tedy rozdělit příjem do určitých intervalů a nechat respondenta vybrat ten, do kterého jeho/její měsíční příjem spadá. Tím se ale otázka stává uzavřenou. Pecáková (2008) dále uvádí, že i slovní odpovědi je třeba nějak kategorizovat, pokud mají být použity pro kvantitativní výzkum, a proto se otevřené otázky hodí spíše pro výzkum kvalitativní. Řezanková (2007) dodává, že lze kombinovat otevřené a uzavřené otázky a potom lze tázat polouzavřené otázky. V nich si může respondent vybrat z variant odpovědí nebo uvést nějakou vlastní.

### 2.1.3 Škálové otázky

U škálových otázek jsou odpovědi respondentů hodnotami určité posuzovací škály (Řezanková 2007). Kohoutek (2010) definuje posuzovací škálu (hodnotící stupnici) jako *„druh dotazníku sloužící k záznamu jednotlivých vlastností posuzované osoby nebo posuzovaného předmětu (např. dopisů, projekčních testových materiálů atp.) posuzovatelem, a to způsobem, který zajišťuje určitou objektivnost a zároveň umožňuje kvantitativní zachycení jevu“*. Řezanková (2007) rozděluje škály na:

- nominální - nelze určit, která varianta je významnější (např. barva)
- ordinální - varianty označují určitou úroveň (např. vzdělání, spokojenost)
- kvantitativní – můžeme číselně vyjádřit rozdíl mezi odpověďmi (např. výška, věk)

Kohoutek (2010) a Chráska (2016) uvádějí jako nejpoužívanější škálu Likertova typu konstruovanou podle Rensise Likerta. Jedná se o bodovací škálu. U

této škály je prezentováno několik tvrzení a respondent má za úkol každé tvrzení ohodnotit stupněm souhlasu na několika bodové škále. Většinou se jedná o pětibodové nebo sedmibodové škály, které se odvíjejí podle Řezankové (2007) a Pecákové (2008) od neutrálního bodu (např. „smíšené pocity“ nebo „nemám názor“).

## **2.2 Požadované vlastnosti dobrého dotazníku**

Podle Kohoutka (2010) by měly mít všechny dotazníky určité vlastnosti, které zajistí dobré měření. Jedná se o:

- Objektivnost
- Standardnost
- Reliabilitu (spolehlivost)
- Validitu (platnost)
- Úspornost
- Míra návratnosti

### **2.2.1 Objektivnost**

Kohoutek (2010) uvádí, že objektivnost metody je dána mírou její nezávislosti na osobě uživatele. Zaručuje nám jednoznačnost výsledků – omezuje nebezpečí, že osoba provádějící průzkum by nějak zkreslila data. Zároveň by neměla mít možnost jakýmkoliv způsobem působit na výsledky. Kohoutek (2010) tedy shrnuje, že u objektivních testů by měli různí posuzovatelé vždy dojít ke stejným výsledkům.

### **2.2.2 Standardnost**

Standardnost Kohoutek (2010) interpretuje jako požadavek, že by se mělo jednat o identickou metodu šetření. Jedná se tedy o sjednocení všech podmínek při vyšetření. Všichni respondenti by měli mít při vyplňování dotazníku stejné úkony, postupovat se stejnými instrukcemi apod. Proces vyhodnocování výsledků by měl být stejný pro každou zkoumanou osobu, a to podle stanovených norem. Kohoutek (2010) normy chápe jako „*typické reakce příslušného vzorku populace podle věku a pohlaví*“. Kohoutek (2010) dodává, že pro standardizaci metody širokého souboru osob, tedy aby se jednalo o reprezentativní výběr.

### **2.2.3 Reliabilita**

Podle Kohoutka (2010) a Řezankové (2007) rozumíme reliabilitou konzistentnost a stabilitu. Jedná se například o stálost výsledků v čase – pokud provedeme šetření jednou a potom provedeme stejné šetření znovu po nějaké delší době, neměly by se výsledky lišit, tedy neměly by být závislé na čase. Kohoutek (2010) a Chráska (2016) uvádějí, že reliabilita se dá zkoumat několika způsoby. Nejjednodušší metodou u klasických dotazníků s výběrovými položkami podle Chrásky (2016) a Kohoutka (2010) je rozdělením přiměřeně velkého reprezentativního výběru ze základního souboru na dva stejně velké výběrové soubory. Výsledky z těchto dvou souborů potom porovnáváme a zjišťujeme jejich sílu souvislosti.

### **2.2.4 Validita**

Podle Chrásky (2016) a Kohoutka (2010) se validitou rozumí to, zda dotazník opravdu a nakolik měří to, co skutečně měřit má (co je výzkumným záměrem). Konstrukce dotazníku v pedagogických výzkumech by měla vycházet z nějaké hypotézy, kterou se snažíme objasnit a jednotlivé otázky v dotazníku by měly být uzpůsobeny tak, aby přinášely data pro verifikaci této hypotézy (Chráska, 2016). Chráska (2016) dodává, že posuzování stupně validity je do určité míry subjektivní, protože vychází hlavně z vlastních znalostí a zkušeností autora

dotazníku. Proto doporučuje, aby autor nechal posoudit dotazník dalšími odborníky.

### **2.2.5 Úspornost**

Podle Kohoutka (2010) by šetření nemělo být, pokud možno časově náročné na administraci a vyhodnocování. Kohoutek (2010) uvádí, že počet osob potřebný pro průzkum se liší podle zkoumaného jevy. Pro některé průzkumy stačí 20-40 osob, pro průzkumy veřejného mínění jich samozřejmě budeme potřebovat mnohem víc, většinou minimálně 1000. Dále by měl být test ekonomický, to podle Kohoutka (2010) znamená nízkou spotřebu materiálu, manipulační jednoduchost a jednoduchost vyhodnocení. Test by měl být srozumitelný, a to hlavně vzhledem ke skupině, která bude testována. Pokud například testujeme žáky na prvním stupni základní školy, je třeba formulovat otázky jinak než pro studenty na vysoké škole. Co se týče počtu otázek, Kohoutek (2010) navrhuje do 50 otázek pro dotazník, který obsahuje hlavně uzavřené otázky.

### **2.2.6 Míra návratnosti**

Podle Chrásky (2016) lze dotazník předat třemi způsoby: rozesláním poštou, osobně nebo prostřednictvím dalších osob. Chráska (2016) uvádí, že nejvýhodnějším způsobem je předání dotazníků osobně. To nám umožňuje okamžitě po vyplnění dotazníků respondenty dotazníky vybrat zpět a zajistit tak prakticky stoprocentní návratnost. Chráska (2016) píše, že pokud dotazník rozesíláme například poštou, je třeba rozeslat alespoň dvojnásobek dotazníků ve srovnání s požadovaným počtem výběru. Dalším problémem neosobního předání a vybrání dotazníků může být podle Chrásky (2016) nereprezentativnost – například bylo prokázáno, že dotazníky vracejí spíše lidé s vyšším vzděláním nebo kladným postojem ke zkoumané problematice (upraveno, Chráska 2016).



Kohoutek (2010) uvádí akceptovatelnou a doporučenou míru návratnosti dotazníků pro různé počty respondentů, zpracovanou podle Mareše (2006):

<b>Počet respondentů</b>	<b>Minimální akceptovatelná návratnost v %</b>	<b>Doporučovaná návratnost v %</b>
5 - 20 osob	nejméně 80 %	více než 80 %
21 - 30 osob	nejméně 75 %	více než 75 %
31 - 50 osob	nejméně 66 %	75 % a více
51 - 100 osob	nejméně 60 %	75 % a více
100 a více osob	nejméně 50 %	75 % a více

**Tabulka č. 1 – Akceptovatelná a doporučená míra návratnosti.**

*Zdroj: Mareš (2006)*

### 3 Vytvoření datového souboru

Chráska (2016) upozorňuje, že po získání vyplněných dotazníků je před jejich analýzou vhodné je zkontrolovat z hlediska korektnosti. Doporučuje tedy vyloučit dotazníky, které nejsou vyplněny správně nebo neúplně.

#### 3.1 Škály měření a typy proměnných

U uzavřených otázek jsou odpovědím podle Řezankové (2007) přiřazeny číselné nebo slovní kódy. Hodnoty znaku jsou pak odpovědi neboli kategorie. U otevřených otázek Chráska (2016) doporučuje provést kategorizaci odpovědí. Příklady statistických znaků charakterizovaných pomocí kategorií jsou podle Řezankové (2007) třeba:

- národnost (česká, slovenská, ...)
- úroveň vzdělání (základní, středoškolské, vysokoškolské)
- počet dětí (0, 1, 2, ...)

V uvedených případech lze pozorovat vztahy mezi kategoriemi, tzn. jaké operace můžeme provádět (národnost nemůžeme porovnávat, ale úroveň vzdělání už ano – vyšší nebo nižší). Podle možnosti těchto operací potom Řezanková (2007) hovoří o škálách měření. Jednotlivé odpovědi čili jim příslušející kódy pak tvoří škálu hodnot. Řezanková (2007) uvedené škály měření rozděluje následovně:

- škála nominální – můžeme jen určit, že jsou hodnoty různé, nemůžeme určit jejich pořadí
- škála ordinální – u této škály můžeme stanovit pořadí, nemůžeme však vyčíslit o kolik
- škála intervalová – můžeme určit o kolik je jedna hodnota vyšší než druhá (jde o číselné hodnoty)

- škála poměrová – můžeme určit o kolikrát je jedna hodnota vyšší než druhá (jde jen o kladné hodnoty)

Řezanková (2007) ještě dodává, že intervalové a poměrové škály spadají pod kvantitativní škály. Chráska (2016) a Řezanková (2007) uvádějí, že každá odpověď respondenta je zaznamenána do proměnné. Každá odpověď odpovídá jedné proměnné. Pokud je možnost vybrat např. dvě odpovědi v jedné otázce, jsou pro ně potřeba dvě proměnné (převzato, Řezanková 2007). Podle Chrásky (2016) a Řezankové (2007) dělíme proměnné podle škál na:

- nominální – např. typ absolvované střední školy
- ordinální – např. dosažený stupeň vzdělání
- intervalové – např. počet dětí v rodině
- poměrové – např. počet členů v domácnosti

Řezanková (2007) dodává, že zvláštním typem proměnné je tzv. proměnná dichotomická (alternativní). Tato proměnná nabývá pouze dvou hodnot (např. kuřák – nekuřák) a jedná se tedy o binární proměnnou s hodnotou 0 nebo 1. Tyto proměnné můžeme podle Řezankové (2007) ještě dělit na:

- symetrické – stejná důležitost kategorií (muž – žena)
- asymetrické – jedna kategorie je důležitější (pacient se uzdravil)

## 4 Analýza jednotlivých proměnných

Řezanková (2007) uvádí jako základní typy analýz zjištění rozdělení četností různých variant hodnot pro každý sledovaný znak a výpočet souhrnných charakteristik.

### 4.1 Rozdělení četností

Rozdělení četností nám podle Řezankové (2007) poskytuje názorný přehled o zjištěných hodnotách a může být prezentováno v tabulce nebo v grafu. Chráska (2016) uvádí jako příklad rozdělení četnosti položky podle pohlaví respondentů. Řezanková (2007) a Chráska (2016) dodávají, že v sociálních vědách se tato analýza označuje jako třídění I. stupně.

V tabulce rozdělení četnosti sledujeme absolutní, relativní a u ordinálních a kvantitativních proměnných kumulativní relativní četnost. Absolutní četností  $n_i$  rozumíme počet výskytů kategorie. Relativní četnost  $p_i$  vyjadřuje podíl počtu výskytů dané kategorie na celkovém rozsahu souboru. Lze ji spočítat vztahem

$$p_i = \frac{n_i}{n}$$

Kumulativní relativní četnost  $P_i$  je sumou relativních četností, spočtená postupem  $P_1 = p_1, P_2 = p_1 + p_2, \dots$  (převzato, Řezanková 2007). Symbolika četnostní tabulky viz Tabulka č. 2:

Znak X	Četnost		
	absolutní	relativní	kumulativní relativní
$x_1$	$n_1$	$p_1$	$P_1$
...	...	...	...
$x_i$	$n_i$	$p_i$	$P_i$
...	...	...	...
$x_K$	$n_K$	$p_K$	1
Celkem	n	1	x

**Tabulka č. 2 – Symbolika pro tabulku rozdělení četností**

*Zdroj: zpracováno podle Řezankové (2007)*

Podle Řezankové (2007) se k zobrazení rozdělení četností kategoriální proměnné používá hlavně graf sloupcový. U nominálních proměnných můžeme použít i graf výsečový. Řezanková (2007) dodává, že u sloupcového grafu výška sloupce představuje počet statistických jednotek dané kategorie. U výsečového grafu jsou výseče v poměru, v jakém se nacházejí četnosti jednotlivých kategorií nominální proměnné.

## **4.2 Popisné charakteristiky**

Rozdělení četností nám umožňuje spočítat různé charakteristické statistiky určitého znaku (upraveno, Řezanková 2007). Autorka pojednává především o míře polohy a míře variability.

### **4.2.1 Popisné charakteristiky nominální proměnné**

Řezanková (2007) uvádí, že u nominální proměnné je míra polohy charakterizována *modální kategorií* (označována také jako *modus*), tedy kategorií s největší četností. Pokud jsou kategorie označeny indexem  $i$  ( $i = 1, 2, \dots, K$ , kde  $K$  je počet kategorií), modální kategorie je označena indexem  $M_o$ . Jsou-li absolutní četnosti značeny  $n_i$  a relativní četnosti  $p_i$ , potom  $\max_i(n_i) = n_{M_o}$  a  $\max_i(p_i) = p_{M_o}$ . Nominální proměnné nelze nijak seřadit podle významnosti, a proto u nich můžeme sledovat pouze četnost kategorií. Řezanková (2007) dodává, že se nemusí vyskytovat jen jedna modální kategorie, ale může jich být i více. Podle počtu modálních kategorií je Řezanková (2007) nazývá:

- Unimodální – jen jedna modální kategorie
- K-modální –  $k$  je počet modálních kategorií
- Bimodální – výskyt dvou modálních kategorií
- Trimodální – výskyt třech modálních kategorií

Řezanková (2007) dodává, že modální kategorie s relativní četností vyšší než 0,5 se může nazývat majoritní kategorií.

Pro určení míry variability je podle Řezankové (2007) základem zjištění koncentrace. Pro zjištění míry koncentrace u nominální proměnné autorka uvádí možnost použití relativní četnosti modální kategorie, vyjádřeno vzorcem

$$p_{M_o} = n_{M_o}/n$$

kde  $n$  je celkový rozsah základního souboru a  $n_{M_o}$  je absolutní četnost modální kategorie. Řezanková (2007) dále pokračuje, že pro určení míry variability pak může posloužit tzv. variační poměr  $v$ , který lze spočítat podle vzorce

$$v = 1 - p_{M_o} = 1 - \frac{n_{M_o}}{n}.$$

Další využívanou metodu určení míry variability nominální proměnné Řezanková (2007) označuje jako nominální rozptyl *nomvar*. Vyjadřuje relativní počet všech dvojic, které nejsou ve stejné kategorii a lze jej spočítat vztahem

$$nomvar = 1 - \sum_{i=1}^K p_i^2 = \sum_{i=1}^K (p_i(1 - p_i))$$

kde  $p_i$  je relativní četnost. Hodnoty  $v$  a *nomvar* nabývají hodnot  $< 0; (K - 1)/K >$ . Čím větší je hodnota charakterizující variabilitu, tím vyšší je heterogenita (různorodost) souboru. Řezanková (2007) dodává, že míra variability může být také vyjádřena hodnotami z intervalu od 0 do 1. Autorka uvádí, že takovou hodnotu můžeme dosáhnout tím, že vypočtenou hodnotu pomocí uvedených vzorců pro variační poměr dělíme maximálně možnou hodnotou, tedy  $(K-1)/K$ . Jedná se o normalizovaný nominální rozptyl.

#### 4.2.2 Popisné charakteristiky ordinální proměnné

Podle Řezankové (2007) patří navíc u ordinální proměnné k mírám polohy mimo modální kategorie i tzv. mediánová kategorie  $x_{M_e}$ . Řezanková (2007) píše, že „je to kategorie, pro kterou kumulativní četnost je 0,5 nebo vyšší, když pro předchozí kategorii byla kumulativní četnost menší než 0,5“. Řezanková (2007) ale dodává, že se však častěji používá *medián*  $\tilde{x}$ , vyžadující číselné označení kategorií. Autorka

zmiňuje, že medián lze spočítat jako průměr z mediánové kategorie  $x_{M_e}$  a kategorie následující, tedy pomocí vzorce

$$\tilde{x} = \frac{(x_{M_e} + x_{M_e+1})}{2}.$$

Řezanková (2007) uvádí, že medián je vlastně střední hodnotou. Zároveň píše, že soubor uspořádaných hodnot lze dalšími charakteristikami rozdělit na dvě části, a to i v jiném poměru, než právě 50 %. Řezanková (2007) tyto charakteristiky obecně nazývá kvantily. Jedny z nejběžnějších kvantilů jsou tzv. kvartily, které soubor rozdělují buď v poměru 25 % a 75 % (dolní kvartil) nebo naopak v poměru 75 % a 25 % (horní kvartil). Do kvartilů zařazuje i minimální a maximální hodnotu  $x_{\min}$  a  $x_{\max}$  (první a poslední hodnota v seznamu hodnot vzestupně seřazeném podle velikosti). Řezanková (2007) k charakteristice mediánu dodává, že se právě jedná o prostřední kvartil, jelikož rozděluje soubor na dvě stejné poloviny.

Pro určení míry variability ordinální proměnné lze podle Řezankové (2007) využít právě výše zmíněné kvantily. Příkladem může být zjištění variačního rozpětí, tedy rozpětí mezi minimální a maximální hodnotou. Jako další příklad autorka uvádí mezikvartilové rozpětí, tedy rozpětí mezi horním a dolním kvartilem.

Řezanková (2007) však uvádí, že větší vypovídací schopnost má ordinální rozptyl *dorvar*. Ten lze určit následujícím vzorcem:

$$dorvar = 2 \sum_{i=1}^{K-1} (P_i(1 - P_i))$$

kde  $P_i$  je kumulativní relativní četnost a *dorvar* nabývá hodnot  $< 0; (K - 1)/2 >$ . „Míra *dorvar* nabývá maxima, právě když u 50 % objektů nabývá sledovaná proměnná hodnoty  $x_1$  a u zbylých 50 % objektů hodnoty  $x_K$ “ (citováno, Řezanková 2007). Normalizovaný ordinální rozptyl (nabývající hodnot od 0 do 1) potom získáme následujícím vztahem

$$norm. dorvar = 2 \cdot dorvar / (K - 1).$$

### 4.2.3 Popisné charakteristiky kvantitativní proměnné

Řezanková (2007) uvádí, že mimo charakteristik míry polohy pro ordinální proměnnou lze u kvantitativní proměnné určit ještě *aritmetický průměr*  $\bar{x}$ , který lze spočítat pomocí tvaru:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Pro určení míry variability se podle Řezankové (2007) nejčastěji používá *rozptyl*  $s^2$ . Ten je vyjádřen následujícím vzorcem:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Od rozptylu je potom odvozená *směrodatná odchylka*:

$$s = \sqrt{s^2}$$

Popisné charakteristiky kvantitativní proměnné jsou popsány stručně, protože se v práci zmíněné proměnné nevyskytují. Kvantitativní proměnné se však v dotaznících obecně často vyskytují. Tyto charakteristiky podrobněji popisuje Řezanková (2007, s. 48).

### 4.2.4 Grafické rozdělení četností

Rozdělení četností můžeme znázornit graficky pomocí různých grafů. U kategoriálních proměnných je jako nejčastější graf používám sloupcový graf. Pro četnost nominálních proměnných lze využít také graf výsečový. Popisné charakteristiky lze také znázornit graficky. Některé z těchto grafů mohou sloužit k oběma účelům (upraveno podle, Řezanková 2007). U sloupcového grafu „*výška sloupce představuje počet statistických jednotek, u nichž se hodnota sledovaného znaku rovná určité kategorii*“ (citováno, Řezanková 2007). Výsečový graf je potom rozdělený v poměru, ve kterém jsou četnosti kategorií nominální proměnné. Autorka se dále zmiňuje o tzv. krabicovém grafu, který poskytuje názornou představu o datech. Tento graf je sestaven z krabičky a „vousů“. Řezanková (2007,



s. 52) dále tento graf podrobněji popisuje. Využívá se však především pro kvantitativní data.

## 5 Analýza závislostí

Pokud chceme provést vícerozměrnou statistickou analýzu, musíme podle Řezankové (2007) daná data uspořádat do tabulky, jejíchž řádky znázorňují statistické jednotky (například osoby) a sloupce znázorňují statistické znaky (proměnné). Dvourozměrné tabulky četností nám potom slouží k analýze kategoriálních dat (Řezanková, 2007).

### 5.1 Statistické hypotézy

Tato část je popsána na základě studijních textů Univerzity Hradce Králové a textu Pražáka (2018) „Poznámky k praktické analýze dat“. Hypotézy jsou v praxi vyslovovány většinou slovně. Jedná se o domněnky o vlastnostech a chování veličin či jevů. Statistická hypotéza je oproti té slovní vyjádřena pomocí matematické symboliky a podle Pražáka (2018) se může týkat úsudků o statistickém rozdělení, parametrech statistického rozdělení a vztazích mezi rozloženými dat několika proměnných. *„Cílem testování statistických hypotéz je na základě vlastností sledovaných znaků ověřit, zda je možné vyslovenou hypotézu považovat za platnou“* (citováno, Testy statistických hypotéz).

Když je testována statistická hypotéza, je vždy třeba proti sobě postavit testovanou hypotézu neboli nulovou hypotézu  $H_0$  a alternativní hypotézu  $H_1$ . Nulová hypotéza parametrického testu vždy předpokládá, že se nevyskytuje žádný rozdíl mezi parametrem testovaného základního souboru a danou hodnotou parametru určitého pravděpodobnostního rozdělení. Alternativní hypotéza se potom formuluje jako výzkumná. Ta předpokládá, že existuje rozdíl mezi předpokládanou hodnotou a základním souborem. Výsledek statistického testu lze interpretovat pouze dvěma způsoby. Buď výsledek odporuje nulové hypotéze, kterou zamítneme a přijímáme alternativní hypotézu nebo není možné nulovou hypotézu odmítnout a považujeme ji tak za platnou (upraveno podle, Pražák 2018).

## 5.2 Hladina významnosti

Při každém testování se vždy vyskytuje možnost chybného rozhodnutí. „Hladina významnosti  $\alpha$  vymezuje maximální pravděpodobnost nesprávného zamítnutí  $H_0$ “ (citováno, Testy Statistických hypotéz). Nejčastější volená hodnota této hladiny je podle textu Testy Statistických hypotéz 0,05 či 0,01. Dosažená hladina významnosti při testování se nazývá *p-hodnota*. Pokud je  $p - hodnota \leq \alpha$ , lze zamítnout hypotézu  $H_0$  na hladině významnosti  $\alpha$ , protože pravděpodobnost její platnosti je příliš nízká. V opačném případě nelze  $H_0$  zamítnout (upraveno podle, Pražák 2018).

## 5.3 Dvourozměrné rozdělení četností

I pro názorný přehled dvou proměnných nám slouží zobrazení rozdělení četností, a to jak v tabulce, tak grafu. Této základní analýze se také říká třídění II. stupně (převzato, Řezanková 2007). Dvourozměrnou tabulku četností (tedy kontingenční tabulku) vytvoříme tak, že „u kategoriálních proměnných jsou četnosti zjišťovány pro všechny takové dvojice kategorií, kdy jedna kategorie z dvojice přísluší první proměnné a druhá kategorie druhé proměnné“ (citováno, Řezanková 2007). Podle Řezankové (2007) lze z hodnot kontingenční tabulky už usuzovat, zda jsou dvě kategoriální proměnné závislé či nezávislé. Řezanková (2007) ještě dodává, že pro testování závislosti je kontingenční tabulka základem.

V políčkách kontingenční tabulky mohou být četnosti uváděny jak v absolutní, tak v relativní podobě. Rozsah souboru Řezanková (2007) označuje písmenem  $n$ . Počet kategorií proměnné  $X$  (řádky) je dále značen písmenem  $R$  a počet kategorií proměnné  $Y$  (sloupce) je značen písmenem  $S$ . Samotné zjištěné četnosti Řezanková (2007) značí jako  $n_{ij}$ , kde  $i = 1, 2, \dots, R$  a  $j = 1, 2, \dots, S$ .

	Znak Y					Celkem
	1. kategorie	...	j-tá kategorie	...	S-tá kategorie	
Znak X 1. kategorie	$n_{11}$	...	$n_{1j}$	...	$n_{1S}$	$n_{1+}$
...	...	...	...	...	...	...
i-tá kategorie	$n_{i1}$	...	$n_{ij}$	...	$n_{iS}$	$n_{i+}$
...	...	...	...	...	...	...
R-tá kategorie	$n_{R1}$	...	$n_{Rj}$	...	$n_{RS}$	$n_{R+}$
Celkem	$n_{+1}$	...	$n_{+j}$	...	$n_{+S}$	$n$

**Tabulka č. 3 – Značení pro kontingenční tabulku absolutních četností**

*Zdroj: zpracováno podle Řezankové (2007)*

Značení pro kontingenční tabulku relativních četností je obdobné – místo značení  $n$  (pro absolutní četnost) je uveden symbol  $p$  (pro relativní četnost).

#### **5.4 Princip zjišťování závislostí dvou proměnných**

Řezanková (2007) rozděluje sledovanou závislost u dvou proměnných na symetrickou (vzájemnou) nebo asymetrickou (jednostrannou). V prvním případě můžeme zkoumat například vzájemnou závislost názorů dvou osob (sourozenci, manželé, ...). V druhém případě se může jednat o závislost názoru respondenta na jeho vzdělání (upraveno, Řezanková 2007). Řezanková (2007) dodává, že většina případů zkoumání vztahu kategoriálních proměnných se týká asymetrické závislosti.

Řezanková (2007) uvádí, že základním testem užívaným k zjištění vzájemné závislosti dvou kategoriálních znaků je chí-kvadrát test o nezávislosti. Při takovém testu předpokládáme, že „pokud jsou dva znaky nezávislé, pak rozdělení četností v kontingenční tabulce je úměrné řádkovým a sloupcovým marginálním četnostem“ (citováno, Řezanková 2007). Marginální (okrajové) četnosti rozumí Chráska (2016) jako součty četností v řádcích a sloupcích tabulky. Při chí-kvadrát testu o nezávislosti podle Řezankové (2007) testujeme, zda se shodují zjištěné a očekávané četnosti.

## 5.5 Tabulka pro dvě nominální proměnné

Test nezávislosti dvou kategoriálních proměnných bude popsán na základě textu Řezanková (2007).

### 5.5.1 Chí-kvadrát test o nezávislosti

Závislost dvou nominálních proměnných se nazývá kontingence. Uvedený chí-kvadrát test o nezávislosti je právě základem pro zjištění této závislosti. Relativní četnost je v základním souboru označována jako  $\pi_{ij}$ . Potom je sestavena nulová hypotéza ve tvaru

$$H_0: \pi_{ij} = \pi_{ij,0}$$

kde  $\pi_{ij,0}$  symbolizuje relativní četnost očekávanou v případě nezávislosti. Tato očekávaná relativní četnost se spočítá pomocí vztahu  $\pi_{ij,0} = p_{i+}p_{+j}$ . Následně tuto hypotézu testujeme proti hypotéze  $H_1$ . Jako testové kritérium pro tuto hypotézu je využívána tzv. Pearsonova statistika chí-kvadrát, vyjádřena vztahem

$$X_P^2 = \sum_{i=1}^R \sum_{j=1}^S \frac{(n_{ij} - n\pi_{ij,0})^2}{n\pi_{ij,0}}$$

Statistika chí-kvadrát nabývá hodnot  $<0; n \cdot (q-1)>$ , kde  $q = \min \{R, S\}$ . Tato veličina má chí-kvadrát rozdělení s  $(R - 1)(S - 1)$  stupni volnosti za předpokladu platnosti nulové hypotézy.

Výslednou hodnotu tohoto testového kritéria na závěr proto porovnáme s kvantilem  $X_{1-\alpha}^2[(R - 1)(S - 1)]$ , přičemž  $\alpha$  je podle potřeby zvolená hladina významnosti. Pokud je výsledná hodnota testového kritéria větší než výše uvedený kvantil, můžeme zamítnout nulovou hypotézu o nezávislosti. V opačném případě nemůžeme tuto nulovou hypotézu zamítnout.

Řezanková (2007) dodává, že předpokladem pro použití tohoto testu o nezávislosti je, že v 80 % buněk kontingenční tabulky nesmí očekávaná četnost klesnout pod 5 a v ostatních buňkách se má vyskytovat četnost alespoň 1.

Chráska (2016) uvádí, že pomocí zmíněného testu můžeme jen určit, zda nějaká závislost existuje, ale nevypovídá o stupni závislosti. Tato statistika chí-kvadrát je podle Řezankové (2007) právě základem pro koeficienty.

### 5.5.2 Pearsonův koeficient kontingence

Pearsonův koeficient kontingence je jedním z koeficientů pro zjištění stupně závislosti mezi dvěma jevy, který je podle Řezankové (2007) značen písmenem  $C_P$  a vypočítán podle vztahu:

$$C_P = \sqrt{\frac{X_P^2}{X_P^2 + n}}$$

Uvedený koeficient je popsán jako „symetrická míra, která vyjadřuje intenzitu vzájemné závislosti dvou proměnných“ (citováno, Řezanková 2007). O výsledné hodnotě dále pojednává Řezanková (2007). Koeficient nabývá hodnot z intervalu  $\langle 0; \sqrt{(q-1)/q} \rangle$ , kde  $q = \min \{R, S\}$ . Pokud je výsledná hodnota 0, jedná se o nezávislost. Naopak, čím je větší výsledná hodnota při stejném  $n$ ,  $R$  a  $S$ , tím silnější je závislost.

### 5.5.3 Cramérovo V

Jako další kontingenční koeficient Řezanková (2007) představuje Cramérovo V, vypočítané podle vztahu:

$$V = \sqrt{\frac{X_P^2}{n(q-1)}}$$

kde  $q = \min \{R, S\}$ . Tento koeficient nabývá hodnot z intervalu od 0 do 1, jelikož ve jmenovateli je maximální hodnota, kterou může statistika chí-kvadrát dosáhnout.

Řezanková (2007) dodává, že pokud máme v tabulce alespoň jednu dichotomickou proměnnou, dostaneme výpočtem tzv. koeficient  $\phi$ , který můžeme vypočítat zjednodušeně tímto vztahem:

$$\varphi = \sqrt{\frac{X_p^2}{n}}$$

Tímto koeficientem se měří míra intenzity závislosti dvou nominálních proměnných.

## 5.6 Tabulka pro dvě ordinální proměnné

U nominálních proměnných nazýváme statistickou závislost kontingencí. U ordinálních proměnných se už této závislosti říká podle Řezankové (2007) korelace. Rozlišuje potom dva typy korelace, a to

- pozitivní – nízkým hodnotám jedné proměnné odpovídají nízké hodnoty druhé proměnné
- negativní – nízkým hodnotám jedné proměnné odpovídají vysoké hodnoty druhé proměnné

Pro zkoumání dvojice objektů Řezanková (2007) využívá a popisuje skupinu měr a Spearmanův koeficient pořadové korelace. Tyto míry lze použít i pro jednu ordinální a jednu kvantitativní proměnnou. Pro jejich výpočet definuje tzv. konkordantní, diskordantní a vázané páry následovně:

- konkordantní pár – jsou-li ve sledované dvojici u jednoho objektu hodnoty u obou proměnných menší (resp. větší) než u druhého objektu
- diskordantní pár – Je-li u jedné proměnné hodnota menší a u druhé proměnné větší
- vázaný pár – hodnota u jedné proměnné nebo hodnoty u obou proměnných jsou shodné

Za účelem zjednodušení zápisů vzorců Řezanková (2007) používá následující symboly:

C – počet konkordantních párů

D – počet diskordantních párů

T<sub>X</sub> – počet párů, které obsahují stejnou hodnotu proměnné X, ale různou hodnotu Y

T<sub>Y</sub> – počet párů, které obsahují stejnou hodnotu proměnné Y, ale různou hodnotu X

### 5.6.1 Spearmanův koeficient pořadové korelace

Následující text popisující Spearmanův koeficient pořadové korelace je zpracován podle Řezankové (2007).

Jako základní míru pro měření korelace můžeme považovat Spearmanův koeficient pořadové korelace. Tento koeficient lze vypočítat pomocí následujícího vzorce:

$$r_s = 1 - \frac{D^2}{\frac{1}{6}(n^3 - n)} = 1 - \frac{6D^2}{n(n^2 - 1)}$$

kde  $n$  je rozsah základního souboru a  $D$  je počet diskordantních párů. Odvození tohoto vzorce popisuje podrobněji Řezanková (2007, s. 93). Výsledná hodnota koeficientu nabývá hodnot z intervalu  $< -1; 1 >$ . Koeficient nabývá hodnoty 1, jsou-li u každého respondenta u obou proměnných stejná pořadí. Jedná se o pozitivní korelaci – přímou závislost. Opačnou hodnotu -1 dostaneme tehdy, když seřadíme hodnoty proměnné X vzestupně a vznikne tím sestupné pořadí proměnné Y. Pak se jedná o negativní korelaci – nepřímou závislost. Při hodnotě 0 jde o lineární nezávislost. Chráska (2016) přibližně interpretuje hodnoty korelačního koeficientu následovně (záporné hodnoty lze interpretovat obdobně):

Koeficient korelace	Interpretace
$r = 1$	naprostá závislost (funkční závislost)
$1,00 > r \geq 0,90$	velmi vysoká závislost
$0,90 > r \geq 0,70$	vysoká závislost
$0,70 > r \geq 0,40$	střední (značná) závislost
$0,40 > r \geq 0,20$	nízká závislost
$0,20 > r \geq 0,00$	velmi slabá závislost
$r = 0$	naprostá nezávislost

**Tabulka č. 4 – Přibližná interpretace hodnot korelačního koeficientu**

*Zdroj: zpracováno podle Chrásky (2016)*



### 5.6.2 Kendallův koeficient pořadové korelace

Jako jednu ze symetrických měř Řezanková (2007) uvádí Kendallovo  $\tau_b$  (tau-b) neboli Kendallův koeficient pořadové korelace. Zjišťuje se pomocí vztahu:

$$\tau_b = \frac{C - D}{\sqrt{(C + D + T_X)(C + D + T_Y)}}$$

kde  $C$  je počet konkordantních párů,  $D$  je počet diskordantních párů,  $T_X$  je počet párů, které obsahují stejnou hodnotu proměnné  $X$ , ale různou hodnotu  $Y$  a  $T_Y$  je počet párů, které obsahují stejnou hodnotu proměnné  $Y$ , ale různou hodnotu  $X$ . Tento koeficient nabývá hodnot  $\langle -1; 1 \rangle$ . Podle Řezankové (2007) může koeficient dosáhnout mezních hodnot pouze v tabulce  $R \times R$ , a to v případě, že se žádná marginální četnost nerovná nule.

## 6 Software SPSS

Kapitola o softwaru IBM SPSS Statistics až po jeho rozhraní je zpracována podle Řeháka (2015, 2017). Samotné postupy, popisující vkládání dat do SPSS, vytváření datového souboru a četnostních a kontingenčních tabulek jsou potom popisovány autorem práce.

SPSS je „*univerzální statistický programový systém, který poskytuje komplex funkcí pro výzkum a zpracování dat*“. Autor uvádí, že se jedná o nejpopulárnější statistický systém v českém, ale i pravděpodobně celosvětovém sociálním výzkumu. Jeho moduly umožňují přípravu dotazníků, archivaci výsledků, pořizování dat a sběr informací. Dále umožňuje provádět u souborů určité úpravy, připravit prezentační tabulky a umožňuje nám provést jednoduchou i pokročilou analýzu dat a k nim i příslušné prezenční i statistické grafy. Řehák (2017) jmenuje řadu funkcí a možností tohoto softwaru. Výhodou je dostupnost pro různé operační systémy i pro pracovní stanice a sálové počítače od všech známých výrobců.

IBM SPSS Statistics je tzv. modulární systém – jádrem aplikace je základní část Base. Tento základní modul je vybaven standardními postupy analýzy dat. Na Base potom navazují další moduly a podle svého charakteru jsou buď analytické nebo obslužné. Důvodem tohoto oddělení je nutnost jen určitých modulů pro různé uživatele, a proto je možné si zakoupit jen některé, pokud ostatní nejsou potřeba.

Jednu z předností softwaru je jeho otevřenost. Dokáže importovat nejen soubory svého typu .sav, ale také .xls, .xlsx, .dbf a další. Systém dokáže soubory i v různých formátech ukládat.

### 6.1 Rozhraní programu

Dvě základní okna programu jsou Data View a Variable View. Data view je tabulka zobrazující data vybraného souboru. Variable View je tabulka proměnných, která specifikuje jejich vlastnosti. Obě okna je možné podle potřeby upravovat.

Program se může ovládat podle jednoduchého nabídkového menu s okny a klávesovými zkratkami. SPSS je však možné ovládat i pomocí mnemotechnického uživatelského zadávacího jazyku (syntaxí).

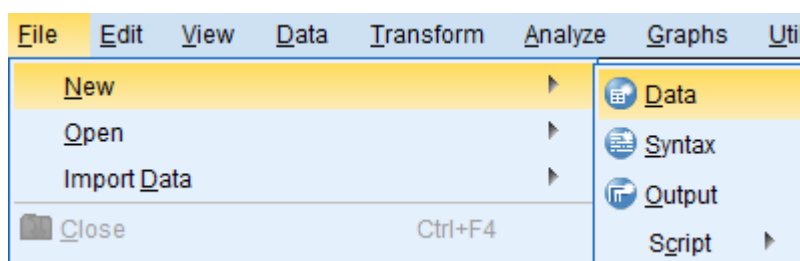
Každý modul poskytuje určité specifické aktivity. Proto se role modulů v procesu datového zpracování liší. Podřizují se obecným funkcím programu:

- Příprava dat na analýzu
- Analytické zpracování dat
- Práce s výstupními tabulkami a grafy

Příprava dat před analýzou se týká souboru jako celku, případů a proměnných. Většina technik pro tuto fázi je zahrnuta v modulu Base. Cílem systému je ale hlavně zpracování informací a výstupu výsledků pro použití v praxi. V systému tedy najdeme všechny často používané statistické metody pro analýzu dat na základní i pokročilé úrovni. Vizualizace výsledků je realizována flexibilně předvolenými nebo vlastními šablonami.

### 6.1.1 Vložení dat do SPSS

Prvním krokem po otevření programu SPSS je vytvoření nového datového souboru přes navigační menu *File -> New -> Data*. Tímto se otevře datový soubor.

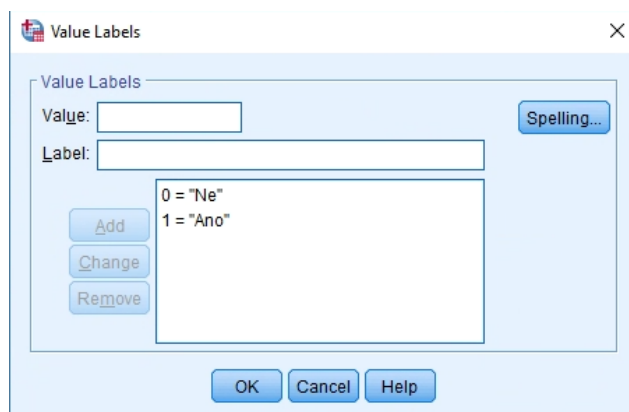


**Obrázek č. 1 - Navigace pro vytvoření datového souboru**

*Zdroj: vlastní zpracování*

V datovém souboru je možné přepínat mezi Data View a Variable View. Pro ruční zadání dat je nutné nejprve definovat proměnné ve Variable View. Zde je třeba vyplnit parametry pro každou proměnnou (otázku). Prvním parametrem je *Name*.

Jedná se o označení jednotlivých otázek. Dalším parametrem je *Type*. Většinou se ponechává *Numeric*, protože se data kódují číselně. Následujícím důležitým parametrem je *Label*. Zatímco v parametru *Name* jsou otázky zakódovány, v *Label* je už rozepsána samotná otázka. V parametru *Value* jsou zakódovány odpovědi pomocí číslic. Při zadávání hodnoty *Value* se pracuje s následujícím oknem:



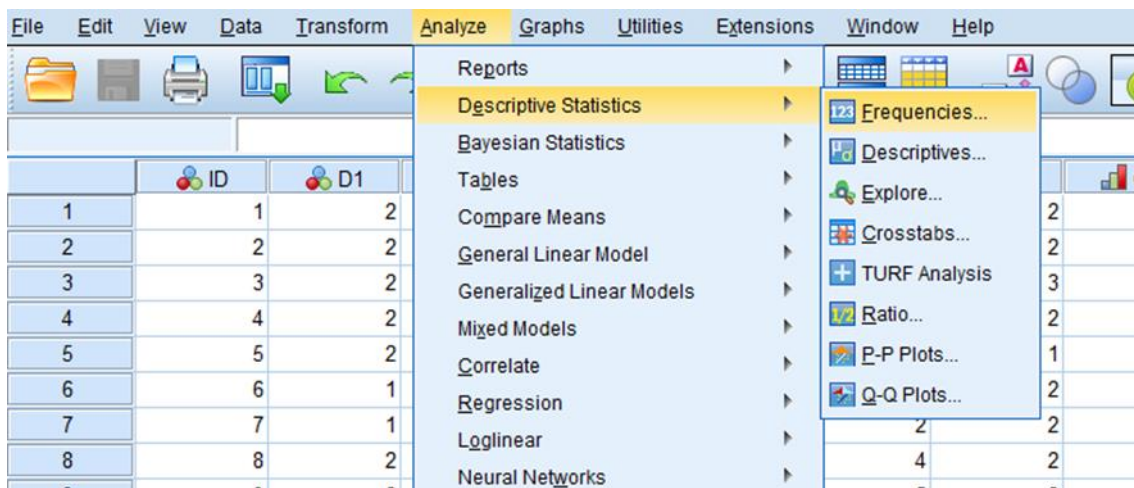
**Obrázek č. 2 - Kódování možností odpovědí**

*Zdroj: vlastní zpracování*

V políčku *Value* uvedeme číslici/hodnotu, pomocí které chceme odpověď kódovat. Do políčka *Label* potom zapíšeme samotnou otázku, ke které chceme číslici přiřadit a tlačítkem *Add* volbu uložíme. Parametrem *Missing* můžeme přiřadit chybějícím hodnotám nějakou hodnotu (například 99). V parametru *Measure* je uvedeno, zda se jedná o nominální, ordinální či kvantitativní proměnnou.

### 6.1.2 Vytvoření četnostní tabulky

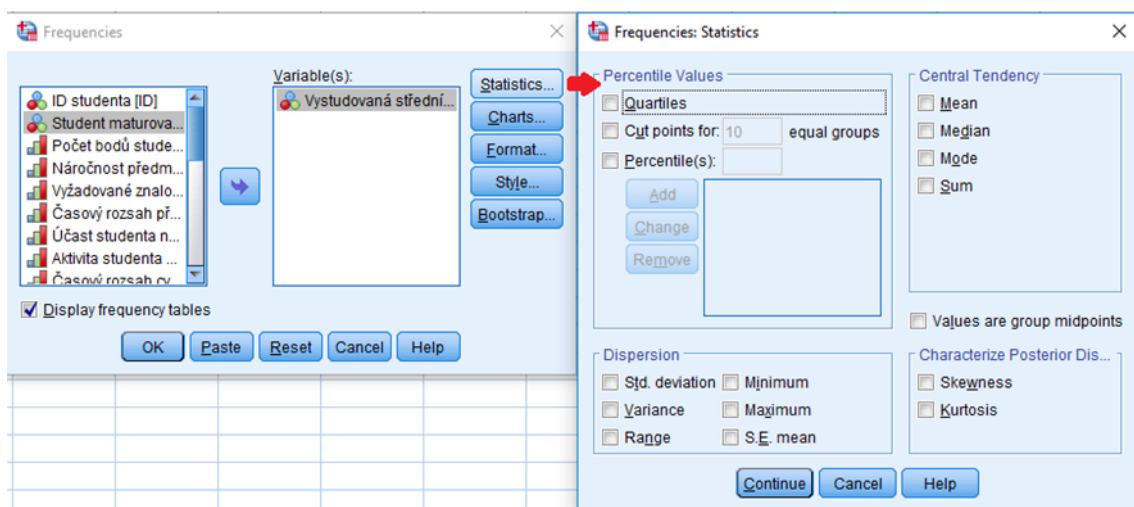
Četnostní tabulka je základem popisné statistiky. Při vytváření četnostní tabulky vybrané proměnné v SPSS se navigujeme přes *Analyze -> Descriptive Statistics -> Frequencies*.



**Obrázek č. 3 - Navigace pro vytvoření četnostní tabulky**

*Zdroj: vlastní zpracování*

V následujícím okně lze vybrat jednotlivé proměnné/položky a tlačítkem *Statistics* otevřít nové okno, ve kterém je možné určit, které popisné statistiky program vypíše. Potvrzením obou oken je vytvořena četnostní tabulka a vypsána popisná statistika pro každou proměnnou.

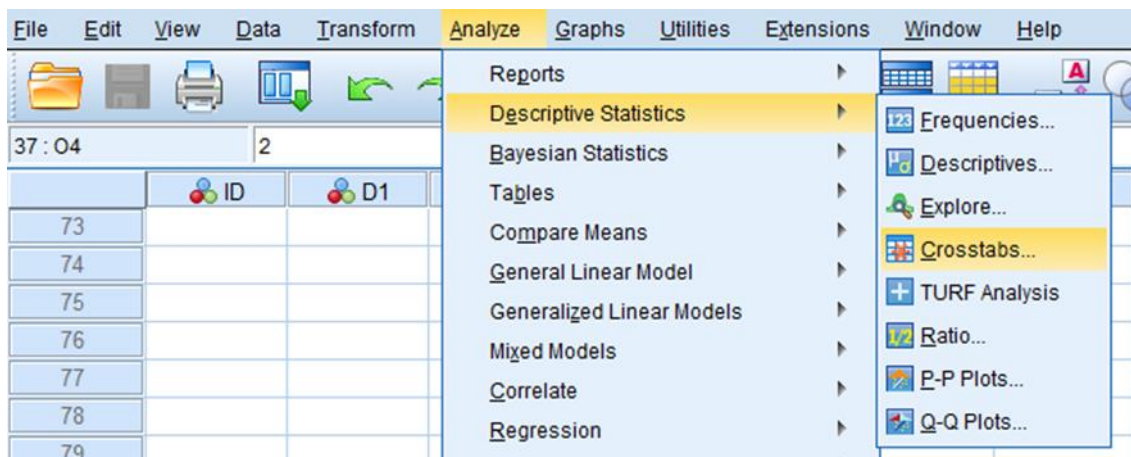


**Obrázek č. 4 - Možnosti popisné statistiky pro četnostní tabulku**

*Zdroj: vlastní zpracování*

### 6.1.3 Vytvoření kontingenční tabulky

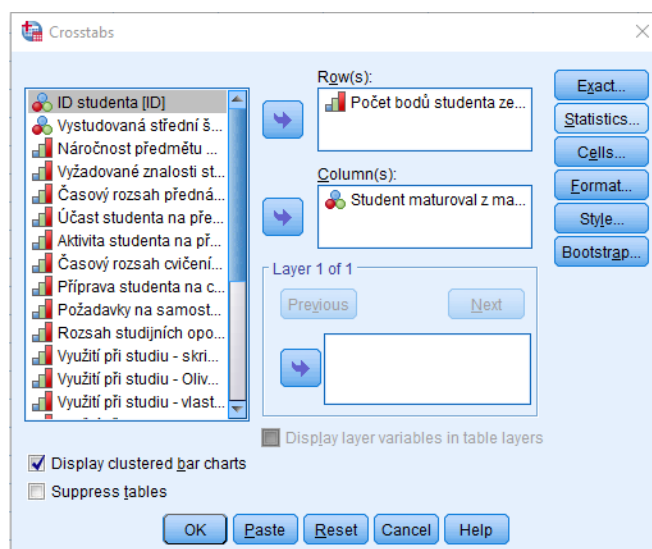
Prvním krokem při analýze závislosti dvou proměnných je vytvoření četností tabulky těchto proměnných, tedy kontingenční tabulky. Tu sestavíme přes navigační menu *Analyze -> Descriptives -> Crosstabs*.



**Obrázek č. 5 – Navigace pro vytvoření kontingenční tabulky**

*Zdroj: vlastní zpracování*

V novém okně lze pomocí šipek přesunout vybrané proměnné do kolonek *Rows* (řádky) a *Columns* (sloupce). Tlačítkem *Cells* jde také v kontingenční tabulce zobrazit očekávané četnosti zaškrtnutím *Expected* v nově otevřeném okně.



**Obrázek č. 6 - Okno výběru možností pro kontingenční tabulku**

*Zdroj: vlastní zpracování*

#### 6.1.4 Překódování proměnných

Pokud budeme chtít překódovat proměnnou například z pěti škálové otázky na tří škálovou, je tomu tak možné přes *Transform -> Recode into Different Variables*. V následujícím okně je možné potom vybrat požadovanou proměnnou,

přejmenovat ji do nové a pomocí tlačítka *Old and New Values* překódovat staré hodnoty na nové.

## 7 Dotazník pro studenty předmětu ZMI1

Aby bylo možné názorně ukázat vybrané metody pro zpracování dat z dotazníku, byl zhotoven dotazník zabývající se spokojeností studentů prvního ročníku Fakulty informatiky a managementu s předmětem Základy matematiky pro informatiky I, viz Příloha č. 1.

Tento dotazník byl inspirován šetřením se stejnou tematikou studentů předmětu Základy matematiky 1 z akademického roku 2011/2012 projektem REFIMAT.

Dotazník se skládá z 25 položek. Studenti ve většině případů hodnotili určený výrok na ordinální škále typu: *souhlas – nesouhlas, spokojen – nespokojen, náročný – nenáročný, hodně – málo* apod. Na konci dotazníku se vyskytují otázky, které studenty dělí podle absolvované střední školy, zda maturovali z matematiky a úspěšnosti v prvním zápočtovém testu.

### 7.1 Dotazníkové šetření

Šetření bylo provedeno osobně 2. 12. 2019 na přednášce předmětu ZMI1 a dále na třech cvičeních předmětu ZMI1 ve dnech 17. 12. a 19. 12. 2019. Dotazník byl vyplňován anonymně.

Jelikož nebyl proveden pilotní průzkum (testovací šetření), objevily se při prvním šetření některé nejasnosti v dotazníku. Díky osobní účasti autora práce byly tyto nejasnosti objasněny a dotazník byl pro následující šetření náležitě opraven.

Ze všech šetření bylo celkem vyplněno 69 dotazníků. Díky osobnímu předávání dotazníků a následnému čekání na výsledky byla návratnost 100 %. Po inspekci byly vyškrtnuty 3 dotazníky z důvodu chybějících hodnot. Pro zpracování bylo tedy využito 66 dotazníků.



## 7.2 Vložení dat do souboru SPSS

Výsledky dotazníkového šetření byly nejprve převedeny do datového souboru v SPSS. V záložce Variable View byly všechny otázky převedeny jako proměnné a následně byly vyplněny patřičné parametry. Odpovědi jsou v datovém souboru zakódovány numericky a parametr *Missing* zůstal nevyplněn, protože dotazníky s chybějícími daty byly předem vyškrtnuty.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	ID	Numeric	8	0	ID studenta	None	None	8	Right	Nominal	Input
2	D1	Numeric	8	0	Vystudovaná st...	{1, Gymnázi...	None	8	Right	Nominal	Input
3	D2	Numeric	8	0	Student maturo...	{0, Ne}...	None	8	Right	Nominal	Input
4	D3	Numeric	8	0	Počet bodů stu...	{1, 0-20%}...	None	8	Right	Ordinal	Input
5	O1	Numeric	8	0	Náročnost před...	{1, Velmi ná...	None	8	Right	Ordinal	Input
6	O2	Numeric	8	0	Vyžadované zn...	{1, Nedostat...	None	8	Right	Ordinal	Input
7	O3	Numeric	8	0	Časový rozsah ...	{1, Nedostat...	None	8	Right	Ordinal	Input
8	O4	Numeric	8	0	Účast studenta...	{1, Pravideln...	None	8	Right	Ordinal	Input
9	O5	Numeric	8	0	Aktivita student...	{1, Dávám p...	None	8	Right	Ordinal	Input
10	O6	Numeric	8	0	Časový rozsah ...	{1, Nedostat...	None	8	Right	Ordinal	Input
11	O7	Numeric	8	0	Příprava studen...	{1, Průběžn...	None	8	Right	Ordinal	Input
12	O8	Numeric	8	0	Požadavky na ...	{1, Velmi vy...	None	8	Right	Ordinal	Input
13	O9	Numeric	8	0	Rozsah studijní...	{1, Neposta...	None	8	Right	Ordinal	Input
14	O10A	Numeric	8	0	Využití při studí...	{1, Studijní ...	None	8	Right	Ordinal	Input
15	O10B	Numeric	8	0	Využití při studí...	{1, Studijní ...	None	8	Right	Ordinal	Input
16	O10C	Numeric	8	0	Využití při studí...	{1, Studijní ...	None	8	Right	Ordinal	Input
17	O10D	Numeric	8	0	Využití při studí...	{1, Studijní ...	None	8	Right	Ordinal	Input
18	O10E	Numeric	8	0	Využití při studí...	{1, Studijní ...	None	8	Right	Ordinal	Input
19	O10F	Numeric	8	0	Využití při studí...	{1, Studijní ...	None	8	Right	Ordinal	Input
20	O10G	Numeric	8	0	Využití při studí...	{1, Studijní ...	None	8	Right	Ordinal	Input
21	O10H	Numeric	8	0	Využití při studí...	{1, Studijní ...	None	8	Right	Ordinal	Input
22	O11A	Numeric	8	0	Využití - rozšíře...	{1, Rozhodn...	None	8	Right	Ordinal	Input
23	O11B	Numeric	8	0	Využití - placen...	{1, Rozhodn...	None	8	Right	Ordinal	Input
24	O11C	Numeric	8	0	Využití - audiovi...	{1, Rozhodn...	None	8	Right	Ordinal	Input
25	O11D	Numeric	8	0	Využití - audiovi...	{1, Rozhodn...	None	8	Right	Ordinal	Input
26	O11E	Numeric	8	0	Využití - audiovi...	{1, Rozhodn...	None	8	Right	Ordinal	Input

**Obrázek č. 7 – Variable View**

*Zdroj: vlastní zpracování*

Otázky jsou v parametru Name označeny podle několika skupin pomocí velkého tiskacího písmena a číslice, která určuje pořadí. První skupinou jsou otázky D1-D3. Jejich účelem je rozřídění respondentů. Druhou skupinou jsou otázky O1-O9. Každá z nich má jedinečné škálu možných odpovědí. Poslední dvě skupiny jsou značeny O10A-O10H a O11A-O11E. Tyto dvě sloučené skupiny položek používají pro každou skupinu stejnou škálu odpovědí.

Označení	Otázka
D1	Vystudovaná střední škola
D2	Maturita z matematiky
D3	Počet bodů ze zápočtového testu
O1	Náročnost předmětu ZMI1
O2	Vyžadované znalosti studenta ze střední školy
O3	Časový rozsah přednášek
O4	Účast studenta na přednáškách
O5	Aktivita studenta na přednáškách
O6	Časový rozsah cvičení
O7	Příprava studenta na cvičení
O8	Požadavky na samostudium studenta
O9	Rozsah studijních opor
O10A	Využití studijních opor - skripta
O10B	Využití studijních opor - Oliva
O10C	Využití studijních opor - vlastní poznámky z přednášek
O10D	Využití studijních opor - vlastní poznámky ze cvičení
O10E	Využití studijních opor - doporučená literatura
O10F	Využití studijních opor - internetové zdroje
O10G	Využití studijních opor - konzultace s vyučujícím
O10H	Využití studijních opor - doučování
O11A	Využití pomoci - rozšíření studijních materiálů v olivě
O11B	Využití pomoci - placený kurz ze základů matematiky
O11C	Využití pomoci - audiovizuální záznamy přednášek
O11D	Využití pomoci - audiovizuální záznamy výkladu vybraných témat
O11E	Využití pomoci - audiovizuální řešení vzorových příkladů

**Tabulka č. 5 – Kódování otázek**

*Zdroj: vlastní zpracování*

Vytvořené proměnné se v Data View zobrazují ve sloupcích. Jednotliví respondenti jsou potom uvedeni v řádcích. V buňkách jsou zakódovány odpovědi pro každou otázku pomocí zmíněného parametru *Value*. Odpovědi jsou zakódovány stejnými číslicemi, které jsou uvedeny u odpovědí každé otázky v dotazníku (vyjma otázky na maturitu z matematiky, kde odpověď „Ano“ je kódována číslicí 1 a odpověď „Ne“ číslicí 0), viz Příloha č. 1.

	ID	D1	D2	D3	O1	O2	O3	O4	O5	O6
1	1	2	1	5	2	2	2	1	1	2
2	2	2	0	4	2	4	2	3	2	2
3	3	2	1	3	2	3	3	2	2	1
4	4	2	0	3	1	2	2	2	3	2
5	5	2	0	4	1	3	1	1	2	1
6	6	1	1	4	2	2	2	4	2	2
7	7	1	0	4	2	2	2	4	3	2
8	8	2	0	4	2	4	2	3	2	2
9	9	2	0	3	2	5	2	3	1	2
10	10	2	1	4	2	3	2	4	2	2
11	11	2	1	3	2	5	2	3	2	1
12	12	2	0	4	2	5	2	4	1	2
13	13	2	1	4	2	2	2	2	2	2
14	14	2	0	3	1	1	1	2	1	2
15	15	2	1	5	1	2	1	2	2	2
16	16	2	0	3	1	2	2	1	1	2
17	17	3	0	2	1	2	2	3	2	1
18	18	2	1	3	1	2	2	5	4	2
19	19	2	1	3	2	3	2	5	4	1
20	20	3	0	3	1	2	2	3	2	3
21	21	1	1	5	2	3	2	2	2	1
22	22	1	1	4	3	5	2	5	4	1
23	23	1	1	4	3	5	2	4	1	2
24	24	1	1	4	3	4	2	2	2	2
25	25	2	1	3	2	2	1	2	3	2
26	26	2	0	3	1	1	2	2	2	2
27	27	7	1	5	2	3	3	1	3	2
28	28	1	1	4	3	5	2	1	2	2
29	29	2	0	2	2	2	2	2	2	2
30	30	1	1	4	4	5	2	2	2	2

**Obrázek č. 8 – Data View**

*Zdroj: vlastní zpracování*

## 8 Popisná statistika a analýza závislostí v SPSS

Pro ukázkou popsání jednotlivých proměnných a analýzu závislostí je pro každý případ vybráno několik proměnných. Na nich je následně provedena ukázkou vybraných statistických metod pomocí programu SPSS a vysvětlena interpretace výsledků.

Pro analýzu závislostí s nominální proměnnou jsou vybrány proměnné D2 (maturita z matematiky) a D3 (Počet bodů ze zápočtového testu). Předpokladem je, že tyto proměnné na sobě budou silně závislé, tedy úspěšnost studentů u zápočtového testu bude záviset na tom, zda student maturoval z matematiky. Pro analýzu závislostí s nominální proměnnou jsou vybrány proměnné O10C (Využití studijních opor - vlastní poznámky z přednášek) a O10D (Využití studijních opor - vlastní poznámky ze cvičení). U těchto proměnných se předpokládá, že je mezi nimi korelace, tedy že studenti používající poznámky ze cvičení používají i poznámky z přednášek a naopak.

### 8.1 Popisná statistika nominálních proměnných

Pro účel ukázky analýzy jednotlivých nominálních proměnných jsou vybrány dvě proměnné D1 (Vystudovaná střední škola), D2 (Maturita z matematiky), viz Tabulka č. 5. Pro každou proměnnou je vytvořena četnostní tabulka v programu SPSS pomocí postupů pro vytvoření četnostní tabulky a následně určeny míry polohy a variability.

### 8.1.1 Proměnná D1 – Vystudovaná střední škola

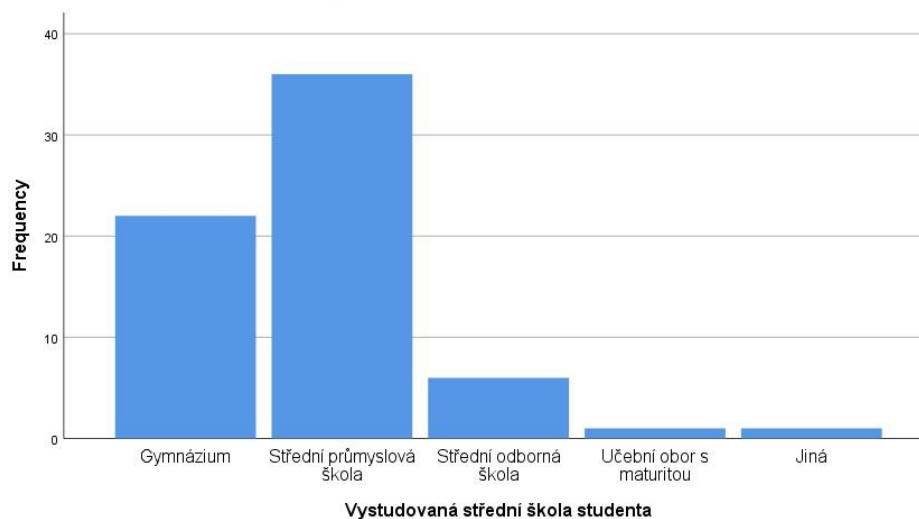
	Frequency	Percent	Valid Percent	Cumulative Percent
Gymnázium	22	33,3	33,3	33,3
Střední průmyslová škola	36	54,5	54,5	87,9
Střední odborná škola	6	9,1	9,1	97,0
Učební obor s maturitou	1	1,5	1,5	98,5
Jiná	1	1,5	1,5	100,0
Total	66	100,0	100,0	

**Tabulka č. 6 – Četnostní tabulka proměnné D1**

*Zdroj: vlastní zpracování – výpis z SPSS*

Ve sloupcích četnostní tabulky je zobrazena četnost (*Frequency*), relativní četnost (*Percent* a *Valid Percent* – tyto hodnoty by se lišily, pokud by v datech byly chybějící údaje) a kumulativní relativní četnost (*Cumulative Percent*). V řádcích jsou potom jednotlivé kategorie.

Z tabulky je možné vidět, že modální kategorií, tedy nejčastější hodnotou, je druhá odpověď „Střední průmyslová škola“. Tuto kategorii vybralo 54,5 %



**Graf č. 1 - Četnosti proměnné D1**

*Zdroj: vlastní zpracování – výpis z SPSS*

respondentů, proto ji můžeme nazvat i majoritní kategorií. Četnosti lze pro lepší vizualizaci také vyjádřit např. sloupcovým grafem.

Z četnostní tabulky lze vypočítat variační poměr  $v = 1 - 0,545 = 0,455$ . Jedná se o středně silný stupeň variability, lze tedy říct, že proměnná je vcelku heterogenní. Dále lze vypočítat nominální rozptyl *nomvar* s využitím pomocné tabulky.

$i$	$p_i$	$p_i(1 - p_i)$
1	0.333	0.222
2	0.545	0.248
3	0.091	0.083
4	0.015	0.015
5	0.015	0.015
<b>Součet</b>		<b>0.582</b>

**Tabulka č. 7 – pomocná tabulka k proměnné D1**

*Zdroj: vlastní zpracování*

Nominální rozptyl: *nomvar* = 0,582, převedeno na normovaný nominální rozptyl *norm. nomvar* = 0,729. Jedná se tedy o vyšší stupeň variability a lze tedy potvrdit, že proměnná je heterogenní.

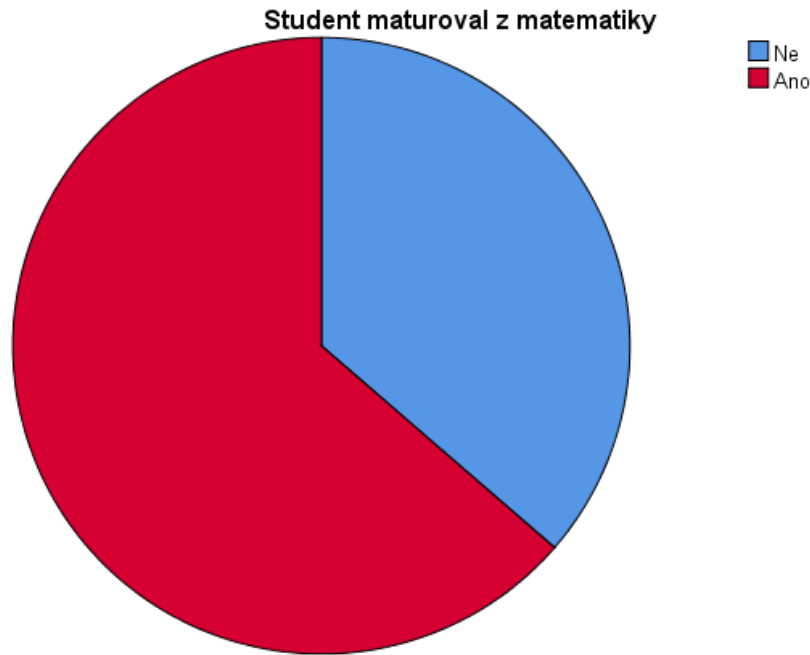
### 8.1.2 Proměnná D2 – Maturita z matematiky

	Frequency	Percent	Valid Percent	Cumulative Percent
<b>Ne</b>	24	36,4	36,4	36,4
<b>Ano</b>	42	63,6	63,6	100,0
<b>Total</b>	66	100,0	100,0	

**Tabulka č. 8 – Četnostní tabulka proměnné D2**

*Zdroj: vlastní zpracování – výpis z SPSS*

V četnostní tabulce vidíme, že modální kategorií je odpověď „ano“, kterou zaškrtno 63,6 % respondentů. Jedná se tedy podruhé o majoritní kategorii. K lepší vizualizaci lze znovu použít některý z grafů, tentokrát je vhodný graf výsečový.



**Graf č. 2 – Četnosti proměnné D2**  
 Zdroj: vlastní zpracování – výpis z SPSS

Následně vypočítáme variační poměr  $v = 1 - 0,636 = 0,364$ . Jelikož v tomto případě může variační poměr nabývat hodnot  $\langle 0; 0,5 \rangle$ , jedná se silnější stupeň variability. Sestavíme pomocnou tabulku:

$i$	$p_i$	$p_i(1 - p_i)$
1	0.364	0.232
2	0.636	0.232
<b>Součet</b>		<b>0.463</b>

**Tabulka č. 9 – Pomocná tabulka k proměnné D2**  
 Zdroj: vlastní zpracování

Tímto získáme  $nomvar = 0,463$ . Nominální normovaný rozptyl  $norm. nomvar = 0,926$ . Hodnota se blíží k nejvyšší možné hodnotě 1, takže můžeme říct, že se jedná o velmi vysoký stupeň variability a proměnná je tedy heterogenní.

## **8.2 Popisná statistika ordinálních proměnných**

Pro ukázkou analýzy jednotlivých ordinálních proměnných jsou k analýze vybrány proměnné O1 (Náročnost předmětu ZMI1) a O4 (Účast studenta na přednáškách), viz Tabulka č. 5. Znovu je ke každé zkonstruována četnostní tabulka

v SPSS pomocí postupů pro vytvoření četnostní tabulky a následně určeny míry polohy a variability.

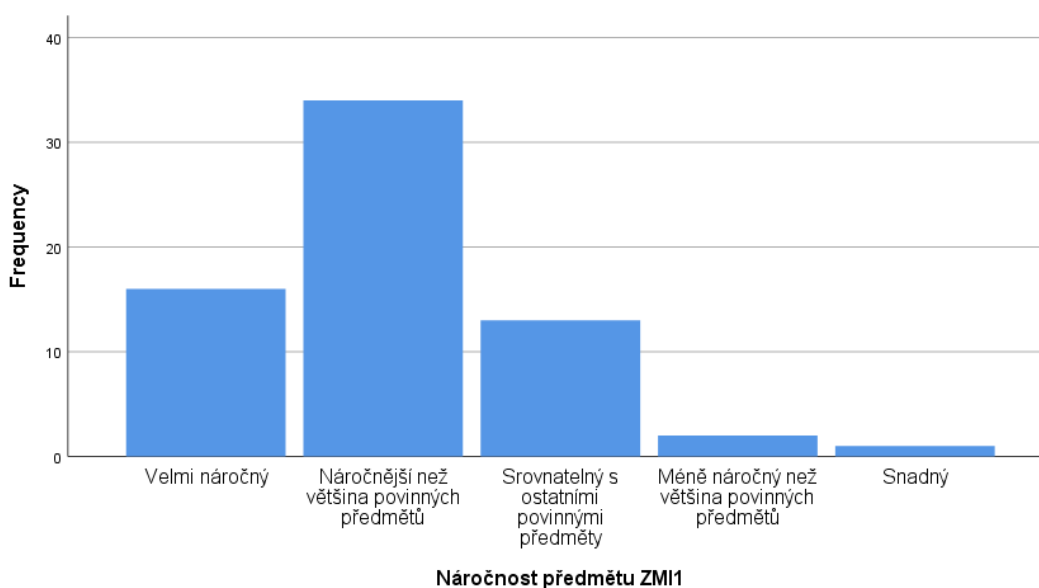
### 8.2.1 Proměnná 01 – Náročnost předmětu ZMI1

	Frequency	Percent	Valid Percent	Cumulative Percent
<b>Velmi náročný</b>	16	24,2	24,2	24,2
<b>Náročnější než většina povinných předmětů</b>	34	51,5	51,5	75,8
<b>Srovnatelný s ostatními povinnými předměty</b>	13	19,7	19,7	95,5
<b>Méně náročný než většina povinných předmětů</b>	2	3,0	3,0	98,5
<b>Snadný</b>	1	1,5	1,5	100,0
<b>Total</b>	66	100,0	100,0	

**Tabulka č. 10 – Četnostní tabulka proměnné 01**

*Zdroj: vlastní zpracování – výpis z SPSS*

Z četnostní tabulky můžeme vidět, že modální kategorií je druhá kategorie. 51,5 % respondentů odpovědělo právě možností „Náročnější než většina povinných předmětů. Zároveň se tedy jedná i o majoritní kategorii. Mediánovou kategorií je též kategorie druhá. První a druhou odpověď zvolilo 75,8 % respondentů (první zvolilo 24,2 %). Pro lepší vizualizaci jednotlivých četností lze využít znovu sloupcový graf.



**Graf č. 3 – Četnosti proměnné 01**

*Zdroj: vlastní zpracování – výpis z SPSS*



Míru variability lze následně určit pomocí ordinálního rozptylu *dorvar*. Ten zjistíme sestavením pomocné tabulky.

$i$	$P_i$	$P_i(1 - P_i)$
1	0,242	0,183
2	0,758	0,183
3	0,955	0,043
4	0,985	0,015
<b>Součet</b>		<b>0,425</b>

**Tabulka č. 11 – Pomocná tabulka k proměnné O1**

*Zdroj: vlastní zpracování*

Tímto získáme  $dorvar = 2 \cdot (0,425) = 0,850$ . Normalizovaný ordinální rozptyl je potom  $norm. dorvar = 0,425$ . Můžeme tedy říct, že se jedná o středně silnou míru variability.

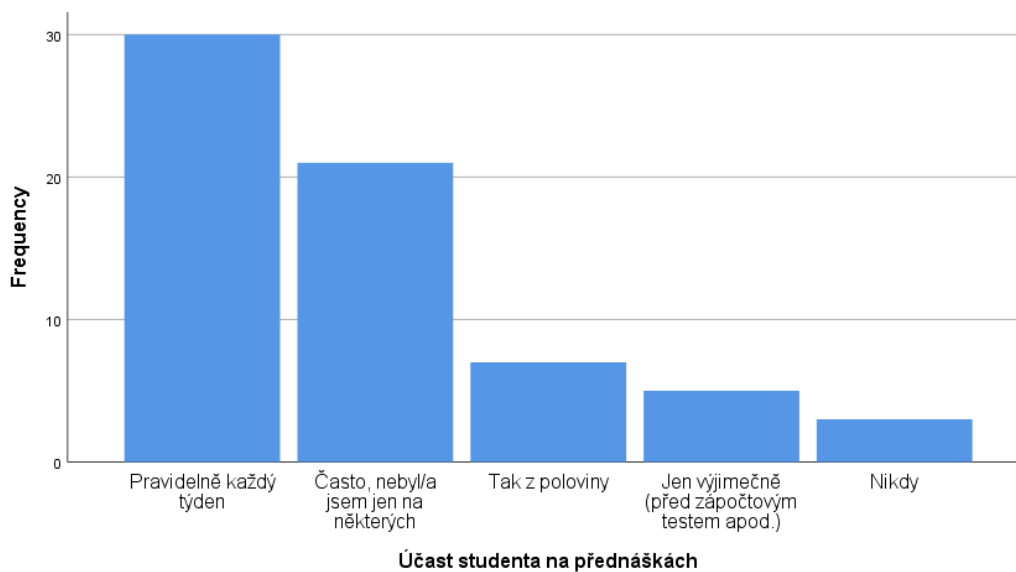
### 8.2.2 Proměnná O4 – Účast studenta na přednáškách

	Frequency	Percent	Valid Percent	Cumulative Percent
<b>Pravidelně každý týden</b>	30	45,5	45,5	45,5
<b>Často, nebyl/a jsem jen na některých</b>	21	31,8	31,8	77,3
<b>Tak z poloviny</b>	7	10,6	10,6	87,9
<b>Jen výjimečně (před zápočtovým testem apod.)</b>	5	7,6	7,6	95,5
<b>Nikdy</b>	3	4,5	4,5	100,0
<b>Total</b>	66	100,0	100,0	

**Tabulka č. 12 – Četnostní tabulka proměnné O4**

*Zdroj: vlastní zpracování – výpis z SPSS*

V tomto případě lze z četnostní tabulky vyčíst, že první kategorie s odpovědí „Pravidelně každý týden“ je modální kategorií. Počet těchto odpovědí však nepřesahuje 50 % a proto se nejedná o majoritní kategorii. Druhá odpověď je mediánovou kategorií a vidíme, že první a druhou odpověď vybralo 77,3 % respondentů (první vybralo 45,5 %). Četnost lze znovu ukázat na grafu pro lepší vizualizaci.



**Graf č. 4 – Četnosti proměnné O4**

*Zdroj: vlastní zpracování – výpis z SPSS*

Míru variability znovu zjistíme pomocí ordinálního rozptylu *dorvar*. Nejprve si vytvoříme pomocnou tabulku, ze které budeme schopni určit *dorvar*.

$i$	$P_i$	$P_i(1 - P_i)$
1	0,455	0,248
2	0,773	0,175
3	0,879	0,106
4	0,955	0,043
<b>Součet</b>		<b>0,573</b>

**Tabulka č. 13 – Pomocná tabulka k proměnné O4**

*Zdroj: vlastní zpracování*

Díky pomocné tabulce určíme ordinální rozptyl  $dorvar = 2 \cdot (0,573) = 1,146$ . Normovaný ordinální rozptyl je následně  $norm. dorvar = 0,573$ . Znovu můžeme uvést, že se jedná o středně silnou variabilitu, avšak o něco větší než u otázky O1.

### 8.3 Analýza závislostí s nominální proměnnou – závislost maturity z matematiky a úspěšnosti v zápočtovém testu

Pro ukázkou analýzy závislosti s nominální proměnnou byly vybrány proměnné D2 (Maturita z matematiky) a D3 (Počet bodů ze zápočtového testu), viz Tabulka č. 5. Pro účely ukázkou je zvolena hladina významnosti 0,05.

		Student maturoval z matematiky		Total
		Ne	Ano	
Počet bodů studenta ze zápočtového testu	0-20%	1	1	2
	21-40%	4	2	6
	41-60%	10	13	23
	61-80%	8	12	20
	81-100%	1	14	15
<b>Total</b>		24	42	66

**Tabulka č. 14 - Kontingenční tabulka proměnné D2 a D3**

*Zdroj: výpis z SPSS*

		Student maturoval z matematiky		Total
		Ne	Ano	
Počet bodů studenta ze zápočtového testu	0-20%	,7	1,3	2,0
	21-40%	2,2	3,8	6,0
	41-60%	8,4	14,6	23,0
	61-80%	7,3	12,7	20,0
	81-100%	5,5	9,5	15,0
<b>Total</b>		24,0	42,0	66,0

**Tabulka č. 15 - Kontingenční tabulka proměnné D2 a D3 - očekávaná četnost**

*Zdroj: vlastní zpracování – výstup z SPSS*

Při porovnání kontingenčních tabulek četností a očekávaných četností je vidět, že se tabulky v některých případech liší. Z tabulky očekávaných četností je možné si všimnout, že více než 20 % buněk (přesně 4 buňky, tedy 40 %) má četnost menší než 5. Tímto je porušen jeden z předpokladů pro použití chí-kvadrát testu o nezávislosti. Jednou z možností, jak tento problém vyřešit, je překódovat některé kategorie a sloučit je tak dohromady. Pomocí okna *Recode into Different Variables* byly proměnné překódovány z pěti do čtyř kategorií.

		Student maturoval z matematiky		Total
		Ne	Ano	
Počet bodů studenta ze zápočtového testu - překódováno	0-40%	5	3	8
	41-60%	10	13	23
	61-80%	8	12	20
	81-100%	1	14	15
Total		24	42	66

**Tabulka č. 16 - Kontingenční tabulka proměnné D2 a D3 (překódováno)**

*Zdroj: vlastní zpracování – výpis z SPSS*

		Student maturoval z matematiky		Total
		Ne	Ano	
Počet bodů studenta ze zápočtového testu - překódováno	0-40%	2,9	5,1	8,0
	41-60%	8,4	14,6	23,0
	61-80%	7,3	12,7	20,0
	81-100%	5,5	9,5	15,0
Total		24,0	42,0	66,0

**Tabulka č. 17 - Kontingenční tabulka proměnné D2 a D3 (překódováno) – očekávaná četnost**

*Zdroj: vlastní zpracování – výpis z SPSS*

Jak je z Tabulky č. 17 vidět, jen jedna buňka má menší očekávanou četnost než 5 a je tedy možné pokračovat s testováním. Pro provedení chí-kvadrát testu o nezávislosti a následné vypočítání koeficientů síly závislosti, pokud se závislost prokáže, bylo zvoleno *Statistics* při vytváření kontingenční tabulky a v novém okně zaškrtnouto *Chi-Square, Contingency coefficient a Phi and Cramer's V*.

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	8,696 <sup>a</sup>	3	,034
Linear-by-Linear Association	7,330	1	,007
N of Valid Cases	66		

a. 1 cells (12,5%) have expected count less than 5. The minimum expected count is 2,91.

**Tabulka č. 18 - Výsledek testu nezávislosti**

*Zdroj: vlastní zpracování – výpis z SPSS*

Tato tabulka zobrazuje chí-kvadrát test o nezávislosti (Pearson Chi-Square). Ve sloupcích tabulky je nejdřív samotná hodnota (Value), počet stupňů volnosti (df) a minimální hladina významnosti (Asymptotic Significance) od které je zamítána hypotéza  $H_0$ . Z tabulky je vidět, že minimální hladina významnosti 0,034 je menší, než stanovená 0,05, takže zamítáme hypotézu  $H_0$  o nezávislosti a lze vypočítat sílu závislosti.

		Value	Approximate Significance
Nominal by Nominal	Phi	,363	,034
	Cramer's V	,363	,034
	Contingency Coefficient	,341	,034
N of Valid Cases		66	

**Tabulka č. 19 – Výpočty kontingenčních koeficientů**

*Zdroj: vlastní zpracování – výpis z SPSS*

V řádcích tabulky je vidět koeficient  $\phi$ , Cramerovo V a Pearsonův kontingenční koeficient. V sloupcích je hodnota každého koeficientu a minimální hladina významnosti (popsány stejně jako v tabulce testů nezávislosti). Cramerovo V a Pearsonův kontingenční koeficient nabývají velmi podobných hodnot 0,363 a 0,341 a můžeme tak říct, že se jedná o nižší stupeň závislosti. Lze tedy potvrdit hypotézu, že úspěšnost studentů souvisí s tím, zda maturovali z matematiky, avšak ne se silným stupněm závislosti.

#### **8.4 Analýza závislosti ordinálních proměnných – korelace mezi studenty využívající poznámky z přednášek a ze cvičení**

Za účelem demonstrace analýzy závislosti dvou ordinálních proměnných byly vybrány proměnné O10C (Využití studijních opor - vlastní poznámky z přednášek) a O10D (Využití studijních opor - vlastní poznámky ze cvičení), viz Tabulka č. 5. Kontingenční tabulka je v následujícím výstupu.

Využití při studiu - vlastní poznámky z přednášek	Využití při studiu - vlastní poznámky ze cvičení					Total
	Studijní oporu jsem zatím nepoužil/a	Studijní oporu využívám výjimečně	Studijní oporu využívám občas	Studijní oporu využívám často, ale ne systematicky	Studijní oporu využívám pravidelně a systematicky	
Studijní oporu jsem zatím nepoužil/a	0	3	0	5	5	13
Studijní oporu využívám výjimečně	1	0	6	4	1	12
Studijní oporu využívám občas	0	2	1	6	8	17
Studijní oporu využívám často, ale ne systematicky	0	1	0	5	6	12
Studijní oporu využívám pravidelně a systematicky	0	0	0	0	12	12
Total	1	6	7	20	32	66

**Tabulka č. 20 - Kontingenční tabulka proměnných O10C a O10D**

*Zdroj: vlastní zpracování – výpis z SPSS*

Abychom mohli určit, zda se jedná o závislost a zároveň stanovit její sílu, je třeba se v okně při vytváření kontingenční tabulky navigovat přes tlačítko *Statistics* a v novém okně tentokrát zaškrtnout možnosti *Correlations* a *Kendall's tau-b*. Alternativní možností je navigace přes *Analyze -> Correlate -> Bivariate*. Po potvrzení se zobrazí okno, ve kterém vybereme dané dvě proměnné a zaškrtneme *Kendall's tau-b* a *Spearman*.

	Value	Asymptotic Standard Error <sup>a</sup>	Approximate T <sup>b</sup>	Approximate Significance
<b>Ordinal by Ordinal</b> Kendall's tau-b	,386	,089	4,288	,000
Spearman Correlation	,451	,102	4,044	,000
<b>N of Valid Cases</b>	66			

**Tabulka č. 21 - Výsledky korelačních koeficientů**

*Zdroj: vlastní zpracování – výstup z SPSS*

Ve výsledné tabulce vzájemné závislosti jsou v řádcích vybrané metody měření síly závislosti. Ve sloupcích je potom důležitá hodnota koeficientů (*Value*) a minimální hladina významnosti, od které je zamítána hypotéza  $H_0$ . Hodnoty koeficientů se pohybují od 0,386 (Kendalovo tau-b) do 0,451 (Spearmanův koeficient pořadové korelace). Je tedy možné říct, že na hladině významnosti 1 % zamítáme nulovou hypotézu o nezávislosti a přijímáme alternativní hypotézu, že proměnné jsou závislé. Jedná se o středně silnou závislost, korelace je pozitivní.

Můžeme tedy říct studenti používající poznámky ze cvičení využívají i poznámky z přednášek a naopak.

## 9 Závěr

Bakalářská práce si klade za cíl sloužit jako návod k tomu, jak nejprve vytvořit vhodné dotazníkové šetření ke statistickému zpracování dat, jak následně zakódovat data z dotazníkového šetření do programu SPSS a jakým způsobem provést vybranými základními statistickými metodami popsání zvolených proměnných a analýzu závislostí mezi vybranými proměnnými.

Čtenář získá poznatky o principech správné konstrukce dotazníku, typech proměnných a správném provádění dotazníkového šetření. Dále mu je objasněna problematika analýzy jednotlivých proměnných. Je vysvětleno, jakými způsoby lze popsat různé proměnné, jejich míry polohy a míry variability, a to zejména proměnných nominálních a ordinálních. Následně je čtenář obeznámen se základními metodami analýzy závislostí mezi proměnnými. Jsou popsány koeficienty pro výpočet síly závislostí nominálních a ordinálních proměnných. Nakonec je popsán software SPSS a jeho rozhraní. Čtenář je poučen o tom, jak překódovat data z dotazníkového šetření do datového souboru a naučí se vytvářet proměnné. Dále je ukázáno, jak lze v programu SPSS vytvořit četnostní tabulky pro jednotlivé proměnné a kontingenční tabulky pro dvojice proměnných. Jsou vysvětleny výstupy četnostních tabulek a čtenář se naučí, jak v programu SPSS použít dané statistické metody pro analýzu závislostí a interpretovat výstupy programu.

Při práci nastalo několik problémů. Neprovedení pilotního testu způsobilo nejasnosti při prvním šetření. Díky osobnímu předání dotazníků však bylo možné vysvětlit nejasnosti a dotazníky byly korektně vyplněny. Autor chyby opravil a u následujících šetření se už žádné problémy nevyskytly. Dalším problémem bylo nesplnění předpokladů pro provedení chí-kvadrát testu o nezávislosti, a to hned v několika případech. Tento test je velmi citlivý na rozsah souboru a šetření o 66 respondentech se ukázalo být rozsahově nedostatečné, vzhledem k počtu odpovědí v otázkách. Problém byl však vyřešen překódováním odpovědí z pěti škálové otázky na čtyř škálovou a byly tak sjednoceny odpovědi respondentů do větších



četností. Tímto způsobem bylo možné test aplikovat. Další možností by bylo použít jiné metody, než které byly rozebírány, tzv. exaktní testy.

Práce může posloužit jako návod pro základní použití statistických metod při zpracování dotazníku, inspirace pro další práce a jako ukázka několika problémů, které mohou při šetření nastat.

Práci by bylo zejména možné rozšířit o kvantitativní data. V provedeném dotazníkovém šetření se taková data přímo nevyskytovala, a proto jim v této nebyla projevována velká pozornost. Nejprve by bylo vhodné rozsáhleji popsat možnosti popisné statistiky kvantitativních dat a provést grafické zobrazení těchto dat pomocí krabicového grafu. Dalším krokem by bylo popsání a následná ukázka analýzy závislostí dvou kvantitativních proměnných. Zároveň by bylo možné práci obohatit a některé další metody analýzy závislostí, které nebyly rozebrány.

## 10 Seznam použité literatury

- [1] CHRÁSKA, Miroslav. *Metody pedagogického výzkumu: základy kvantitativního výzkumu*. 2., aktualizované vydání. Praha: Grada, 2016, 256 s. Pedagogika (Grada). ISBN 978-80-247-5326-3.
- [2] DRAESSLER, Jan. Testy statistických hypotéz. In: *Univerzita Hradec Králové* [online]. c2020 [cit. 2020-07-06]. Dostupné z: [https://oliva.uhk.cz/bbcswebdav/pid-102345-dt-content-rid-186928\\_1/courses/KIKM-PSTA/PSTA\\_6\\_1.pdf](https://oliva.uhk.cz/bbcswebdav/pid-102345-dt-content-rid-186928_1/courses/KIKM-PSTA/PSTA_6_1.pdf)
- [3] KOHOUTEK, Rudolf. DOTAZNÍK jako průzkumná a výzkumná metoda [online]. 2010 [cit. 2020-01-31]. Dostupné z: <http://rudolfkohoutek.blog.cz/1002/dotaznik-jako-pruzkumna-metoda>
- [4] PECÁKOVÁ, Iva. *Statistika v terénních průzkumech*. Praha: Professional Publishing, 2008, 231 s. ISBN 978-80-86946-74-0.
- [5] PRAŽÁK, Pavel. *Poznámky k praktické analýze dat*. 2018.
- [6] ŘEHÁK, Jan a Ondřej BROM. *SPSS – Praktická analýza dat*. Brno: Computer Press, 2015. ISBN 978-80-251-4609-5.
- [7] ŘEZANKOVÁ, Hana. *Analýza dat z dotazníkových šetření*. Praha: Professional Publishing, 2007, 212 s. ISBN 978-80-86946-49-8.
- [8] Test shody četnosti - skript softwaru IBM SPSS Statistics | ACREA. *Analytická kreativita v business analytics | ACREA CR spol. s r.o.* [online]. c2020 [cit. 2020-04-28]. Dostupné z: <https://acrea.cz/skripty/bezplatne-skripty/test-shody-četnosti/>

## 11 Přílohy

1) Příloha č. 1

### ZÁKLADY MATEMATIKY PRO INFORMATIKY - ZPĚTNÁ VAZBA

AKADEMICKÝ ROK 2019/2020

Vážený studente/studentko, chtěl bych Tě požádat o vyplnění mého dotazníku za účelem zpětné vazby na předmět ZMI1. Odpovídej prosím zakroužkováním číslice vlevo u příslušné odpovědi. U baterie otázek pak zakroužkuj příslušnou číslici na pravé straně.

1) Předmět ZMI1 je pro mne v porovnání s ostatními předměty:

1	Velmi náročný
2	Náročnější než většina povinných předmětů
3	Srovnatelný s ostatními povinnými předměty
4	Méně náročný než většina povinných předmětů
5	Snadný

2) Obsah předmětu ZMI1 vyžaduje znalosti ze střední školy. Moje znalosti jsou:

1	Nedostatečné, musím si je výrazně rozšířit
2	Nepostačující, ale není problém si je doplnit
3	Odpovídající, výuka plynule navazuje na učivo SŠ
4	Některé tematické celky již znám
5	Většinu učiva už znám

3) Časový rozsah přednášek předmětu ZMI1 je podle mého názoru:

1	Nedostatečný, bylo by vhodné časovou dotaci rozšířit
2	Odpovídající rozsahu učiva
3	Přednášky jsou zbytečně dlouhé

4) Přednášky navštěvuji:

1	Pravidelně každý týden
2	Často, nebyl/a jsem jen na některých
3	Tak z poloviny
4	Jen výjimečně (před zápočtovým testem apod.)
5	Nikdy

5) Na přednáškách:

1	Dávám pozor a píšu si poznámky
2	Snažím se dávat pozor, ale mám problémy se celou dobu soustředit
3	Nejsem schopen/na se soustředit, spíše se nudím
4	Na přednášky nechodím

6) Časový rozsah cvičení předmětu ZMI1 je podle mého názoru:

1	Nedostatečný, bylo by vhodné časovou dotaci rozšířit
2	Odpovídající rozsahu učiva
3	Cvičení jsou zbytečně dlouhé

7) Na cvičení se připravuji:

1	Průběžně, systematicky
2	Jen před zápočtovými testy
3	Nikdy

8) Požadavky na samostudium jsou:

1	Velmi vysoké, přípravě jsem věnoval/a více času než ostatním povinným předmětům
2	Srovnatelné s ostatními povinnými předměty
3	Menší než u některých povinných předmětů

9) Rozsah a obsah studijních opor (Oliva, skripta, „Nko“ apod.) je:

1	Nepostačující, bylo by vhodné je rozšířit
2	Odpovídající k požadavkům splnění předmětu
3	Zbytečně rozsáhlý

V otázkách číslo 10) až 17) vyjádři míru využívání studijních opor podle následující škály:

5	Studijní oporu využívám pravidelně a systematicky
4	Studijní oporu využívám často, ale ne systematicky
3	Studijní oporu využívám občas
2	Studijní oporu využívám jen výjimečně
1	Studijní oporu jsem zatím nepoužil/a

10)	Skripta	1	2	3	4	5
11)	Studijní materiály z Olivy	1	2	3	4	5
12)	Vlastní poznámky z přednášek	1	2	3	4	5
13)	Vlastní poznámky ze cvičení	1	2	3	4	5
14)	Doporučená literatura	1	2	3	4	5
15)	Zdroje z internetu	1	2	3	4	5
16)	Konzultace s vyučujícím	1	2	3	4	5
17)	Doučování	1	2	3	4	5

V otázkách číslo 18) až 22) pomocí následující škály uveď, které položky by Ti pomohly ke studiu:

5	Rozhodně ano
4	Spíše ano
3	Nedokážu posoudit
2	Spíše ne
1	Rozhodně ne

18)	Rozšíření studijních materiálů v Olivě	1	2	3	4	5
19)	Placený kurz ze základů matematiky	1	2	3	4	5
20)	Audiovizuální záznamy přednášek	1	2	3	4	5
21)	Audiovizuální záznamy výkladu vybraných témat	1	2	3	4	5
22)	Audiovizuální řešení vzorových příkladů	1	2	3	4	5

23) Studoval/a jsem:

1	Gymnázium
2	Střední průmyslová škola
3	Střední odborná škola
4	Obchodní akademie
5	Lyceum
6	Učební obor s maturitou
7	Jiná

24) Maturoval/a jsem z matematiky:

1	Ano
2	Ne

25) Z prvního zápočtového testu jsem získal/a:

1	0-20 % bodů
2	21-40 % bodů
3	41-60 % bodů
4	61-80 % bodů
5	81-100 % bodů

## Podklad pro zadání BAKALÁŘSKÉ práce studenta

Jméno a příjmení: **Jakub Horáček**  
Osobní číslo: **I1700514**  
Adresa: **Škroupova 352, Chrudim – Chrudim III, 53701 Chrudim 1, Česká republika**  
Téma práce: **Statistické metody pro zpracování dotazníku.**  
Téma práce anglicky: **Statistical methods in the questionnaire analysis**  
Vedoucí práce: **doc. RNDr. Pavel Pražák, Ph.D.**  
**Katedra informatiky a kvantitativních metod**

### Zásady pro vypracování:

Cílem práce je popsat a na praktickém problému ukázat použití vybraných statistických metod při zpracování dotazníku.

Osnova:

1. Úvod
2. Teoretická část - vybrané statistické metody pro zpracování dotazníku a jejich popis
3. Statistický software SPSS
4. Analýza dat z dotazníku podle vybraných statistických metod
5. Závěr a doporučení
6. Literatura

### Seznam doporučené literatury:

Řezanková, H.: Analýza dat z dotazníkových šetření, Professional Publishing, 2010.  
Chrástka, M.: Metody pedagogického výzkumu, Grada Publishing, 2016.  
Řehák, J., Brom, O.: SPSS – praktická analýza dat, Computer Press, 2015  
Pecáková I.: Statistika v terénních průzkumech, Professional Publishing, 2008  
Mareš, P., Rabušic, L., Soukup, P.: Analýza sociálněvědních dat (nejen) v SPSS, Masarykova univerzita, 2015

Podpis studenta:

Datum:

Podpis vedoucího práce:

Datum: